Funding of Clinical Trials and Reported Drug Efficacy*

Tamar Oostrom†

February 4, 2021

Abstract

Clinical trials are a key component of drug approvals and prescription decisions, and are often funded by pharmaceutical firms. This paper estimates the effect of financial sponsorship of clinical trials on reported drug efficacy, leveraging the insight that the exact same sets of drugs are often compared in different randomized control trials conducted by parties with different financial interests. In principle, randomized control trials comparing the same drugs should yield comparable estimates, regardless of the interests of the trial's funders. In practice, I use newly assembled data on hundreds of psychiatric clinical trials to estimate that a drug appears 0.15 standard deviations more effective when the trial is sponsored by that drug's manufacturer, compared with the same drug in the same trial without the drug manufacturer's involvement. Observable characteristics of trial design and patient enrollment explain little of this effect. In contrast, publication bias – in which the publication decision is more responsive to the manufacturer's drug's efficacy than the efficacy of other drugs in the trial – can account for nearly half of the sponsorship effect. The sponsorship effect decreases over time as pre-registration requirements were implemented, suggesting that pre-registration can be effective in overcoming sponsorship bias.

This paper is updated frequently. The latest version can be found here.

*I am very grateful to Amy Finkelstein, Heidi Williams, and Jim Poterba for their enthusiasm and guidance. I would like to extend a special thanks to Pierre Azoulay, Jonathan Gruber, Frank Schilbach, and Scott Stern for helpful comments and support. This paper also benefited greatly from discussions with Sarah Abraham, David Autor, Ivan Badinski, Jane Choi, Joe Doyle, Colin Gray, Ryan Hill, Allan Hsiao, Simon Jaeger, Madeline Mckelway, Parinitha Sastry, Cory Smith, Carolyn Stein, Sean Wang, Michael Wong, seminar participants at Columbia Business School, Harvard Business School, the University of Illinois at Chicago, UCLA Anderson, Reed College, and Ohio State, and several anonymous clinical trial managers. Audrey Pettigrew provided excellent research assistance. This material is based upon work supported by the National Institute on Aging under Grant Number T32-AG000186 and the National Science Foundation Graduate Fellowship Program under Grant Number 1122374. First draft April 2019.

†The Ohio State University, Department of Economics. Email: oostrom.1@osu.edu

1 Introduction

In 1993, Wyeth Pharmaceuticals introduced a new antidepressant drug venlafaxine, also known by the brand name Effexor. Over the next decade and a half, Wyeth funded numerous randomized control trials (RCTs) comparing the effectiveness of its new drug with a main competitor—Eli Lilly's blockbuster drug fluoxetine, also known as Prozac. In twelve out of fourteen trials funded solely by Wyeth, their drug venlafaxine had a higher efficacy point estimate compared to fluoxetine.[1] In contrast, only one of the three clinical trials with alternate funding found that venlafaxine had a higher efficacy point compared to fluoxetine.[2] Motivated by such examples—which might be due to idiosyncratic differences across these trials—I construct a data set of hundreds of psychiatric clinical trials and systematically investigate the effect of an RCT's funder on the reported efficacy of the tested drugs.

Clinical trials are a key component of pharmaceutical research and development. These trials are an expensive investment[3] and the results are the basis of regulatory, prescribing, and medical treatment decisions (Davidoff et al., 2001) in market worth $480 billion in the United States (Yu et al., 2018) and over a trillion worldwide (Mikulic, 2020). In addition to financial implications, the results of clinical trials also have direct consequences for the health of the population, as seen by the recent Pfizer and Moderna clinical trials for a COVID-19 vaccine. Over the past decade the number of industry funded trials has increased by 43%, while trials funded by the National Institutes of Health have decreased by 24% (Ehrhardt et al., 2015). More than half of life science researchers have some financial relationship with industry (Zinner et al., 2013), and one fourth have a direct industry affiliation (Bekelman et al., 2003). If clinical trials with alternate funding sources reach different conclusions about the efficacy of the same drugs, then approval and prescription decisions may be based on inaccurate efficacy assessments. Obtaining accurate efficacy information for drugs is an important input in determining optimal health care utilization.

This paper examines how financial incentives can affect the results of randomized control trials. Specifically, I examine how much the funder of a clinical trial can change the reported efficacy of the drugs tested. Of course, privately funded trials are often very different from publicly funded trials in terms of the diseases studied, the drug and comparator treatments tested, and the outcomes examined. If trials with different funding sources simply tested different drugs or comparators, any efficacy differences observed may be due to differences in the drugs themselves, not the causal effect of changing funding sources. My paper accounts for these concerns by directly comparing RCTs with identical drugs and comparators. The key insight is that the exact same sets of drugs are often compared in different RCTs conducted by parties

[1] As discussed in Section 2.3.3, I focus on a consistent efficacy outcome across trials. For antidepressant medications, this standard outcome is the share of patients who respond to treatment within eight weeks.

[2] One trial was funded by the Department of Health of Taiwan. One was funded by Wyeth, but the authors were also consultants for Eli Lilly. The authors of the final trial received grant funding from Wyeth, were consultants for both Eli Lilly and Wyeth, and received speaking fees from Eli Lilly.

[3] Estimates for the average cost of late stage clinical trials range from $20–35 million per trial, and tens of thousands of clinical trials are conducted annually (Sertkaya et al., 2016; Moore et al., 2018). These statistics are from 2004-2012 and 2015-2016 data, respectively, and have not been adjusted for inflation.

with different financial interests. Therefore, I isolate the effect of changing financial sponsorship from the different objectives of pharmaceutical firms, non-profit sponsors, and other agencies.[4] I also explore mechanisms to mitigate this conflict of interest. One such policy is pre-registration of clinical trials, which has debatable efficacy in economics (Abrams et al., 2020).

I construct a data set of clinical trials where the exact same sets of drugs are studied numerous times in trials with different sponsorship interests. Specifically, my analysis focuses on psychiatric disorders because of data availability and their large economic costs. There were two large meta-analyses of the efficacy of antidepressant and antipsychotics published recently (Cipriani et al., 2018; Leucht et al., 2013) and 12.7% of the U.S. adult population takes antidepressant medication monthly (Pratt et al., 2017). Each of these trials is a double-blind RCT, enrolls adults with a primary diagnosis of major depressive disorder or schizophrenia according to standard diagnostic criteria, and examines the same standard outcomes.[5] The majority of these trials are post-market and were published after the drugs gained approval from the Food and Drug Administration (FDA). Some of these trials are sponsored by the manufacturer of one of the drugs; others have alternate funding sources, such as governments, alternate private firms, or unacknowledged funders.[6]

Utilizing dozens of drug combinations across hundreds of clinical trials, I estimate that a drug appears 36 percent more effective (0.15 standard deviations off of a base of 0.42) when the trial is sponsored by that drug's manufacturing or marketing firm, compared with the same drug, evaluated against the same comparators, but without the drug manufacturer's involvement. As in the medical literature, I measure efficacy, in the case of antidepressants, as the share of patients that respond to medication or, in the case of schizophrenia, as the average decline in symptoms. Sponsored drugs are also 47 percent more likely to report statistically significant improvements over other arms (0.10 off of a base of 0.22), and 42 percent more likely to be the most effective drug in a clinical trial (0.16 off of a base of 0.39), again, compared with the same drug tested against the same set of drugs, but without funding from the drug's manufacturer. Consistent with the effect of sponsorship being driven by financial incentives of sponsors, the sponsorship effect is greater for drugs with a larger market size.

There are two classes of potential mechanisms that could be driving this sponsorship effect. Trials could either be planned or conducted differently ex-ante or presented or published differently ex-post. I refer to the first class of mechanisms as differential trial design. Sponsored arms might differentially select patients that are more likely to respond to a given drug, or might set trial characteristics that are advantageous for the sponsored drug. Differential trial design also includes modifications to the sample of patients analyzed

[4]The focus of my research is to investigate whether the funding source affects reported efficacy. The pharmaceutical industry is not the only type of funder that might have an interest in augmenting the efficacy of a particular arm. Governmental funded trials might be conducted by investigators with strong priors about the efficacy of a particular drug; patient organizations might want what they perceive to be the newest and best medications made available.

[5]As a separate point, the outcomes that clinical trial results chose to report and highlight may often be endogenously selected. Among pre-registered trials, 31% showed disparities between the outcomes registered and the outcomes published (Mathieu et al., 2009). In this analysis, I focus on a consistent set of outcomes to focus on differences in apparent efficacy, not reporting. The choice of which outcomes to report is an interesting topic, but not the focus of this work.

[6]If no funding is acknowledged, the authors are almost always academic researchers at a university or medical center.

based on their responsiveness to treatment. I incorporate data on trial characteristics such as the length of the trial, the drug's dosage, total enrollment, recruitment area, and treatment setting and patient characteristics such as the mean age, gender, and baseline severity. I find limited support for differential trial design or patient selection as a mechanism for the sponsorship effect. This analysis is constrained by characteristics which are observable, and of course trials could be differentially selected on a number of important but unobserved characteristics. For each of the observed trial characteristics, I estimate drug-specific predicted efficacy and find that sponsored arms, conditional on the drug and set of drugs examined, do not have higher predicted efficacy.

In contrast, I classify any mechanisms that occur after the completion of the clinical trial as publication bias. Publication bias might involve the decision to differentially publish results based on their favorability to the sponsor. My analysis focuses on a consistent set of outcomes, so I don't consider endogenous outcome selection as a mechanism and focus on changes to actual reported efficacy. A variety of evidence suggests that publication bias can partially explain this sponsorship effect. Incorporating data on unpublished clinical trials, I find sponsored trials are less likely to publish non-positive results for their drugs. In addition, the sponsorship effect decreases over time as scientific norms increasingly encouraged pre-registration of clinical trials and expanded access to clinical trial results. The International Committee of Medical Journal Editors (ICMJE) required pre-registration as a condition for publication in their journals starting in 2005, and the effect of sponsorship on reported drug efficacy is statistically significantly lower after that year and no longer statistically significantly different from zero. In addition, there is no evidence of a sponsorship effect among the set of trials pre-registered in ClinicalTrials.gov. However, my estimates are underpowered to distinguish between a decrease after the enforcement of pre-registration requirements and a general decline in the sponsorship effect over time.

As a final component, I estimate how much of this sponsorship effect can be explained by publication bias by incorporating data on both unpublished trials and all pre-registered antidepressant trials in recent years. Under the assumptions that the unpublished trials I observe are a random subset of all unpublished trials and that all initiated clinical trials were pre-registered, I estimate that 40–50% of this sponsorship effect can be explained by publication bias. This 40–50% estimate is likely a lower bound. To the extent that some recent clinical trials were neither published nor pre-registered and the unpublished trials I observe are more favorable to the sponsors than all unpublished trials, my estimate would underestimate the share explained by publication bias. The remaining unexplained share of the sponsorship effect could be due to underestimating the publication channels described above, differential selection on trial characteristics unobserved in my data, or data manipulation or reconciliation errors.

This paper provides evidence that the funder of a trial affects the reported efficacy of tested drugs, which has consequences for drug approval and prescription decisions. However, welfare consequences of different funding sources depend on several factors. For example, if pharmaceutical firms were restricted from researching their own products, the total amount of innovative research would likely decrease. In addition, if physicians, patients, and regulators already appropriately incorporate trial funding when evaluating the

results of clinical research, then changes to sponsorship would have limited consequences for downstream financial outcomes such as prescriptions. The consequences of alternate clinical trial funding also depend on whether the sponsor of a trial affects either the availability of the knowledge produced or the external validity of the research. My findings suggest that sponsors affect the publication decision, and thus the availability of knowledge produced. I find no evidence that sponsors affect the external validity of estimates to observably different populations or settings. In aggregate, my findings show that the sponsor of a clinical trial has a substantial and significant effect on both the reported efficacy of the drugs and the availability of trial results.

My paper is the first to examine the effect of financial sponsorship on RCT outcomes by directly comparing a large set of trials in which the exact same arms are tested with differing financial interests. This paper builds on a large medical literature documenting the association between clinical trial outcomes and funding sources. Clinical trials funded by industry are more likely to report positive outcomes than those funded by the government or non-profits (Bourgeois et al., 2010; Perlis et al., 2005), more likely to report outcomes that favor the sponsor (Lexchin et al., 2003), and less likely to report unfavorable cost-effectiveness assessments (Friedberg et al., 1999). This positive association has been robustly corroborated in large meta-analyses (Lundh et al., 2017; Bekelman et al., 2003; Perlis et al., 2005). However, this association could be because pharmaceutical companies selectively fund trials on drugs they consider to be more effective (Lexchin et al., 2003), or due to selection of the comparative treatment (Bourgeois et al., 2010). For example, pharmaceutical companies could fund newer drugs, which on average are more effective than previous versions (Lathyris et al., 2010). Alternately, they could test their drugs against differentially effective drugs in that class (Psaty et al., 2006). In these cases, a correlation might exist between industry funded trials and more positive outcomes, but it would not measure the causal effect of changing sponsorship *for a given drug and trial* as in this paper.

In addition to the medical literature, this paper contributes to the literature on sources of bias and external validity in RCTs. While RCTs are an effective tool for evaluating the effectiveness of interventions, recent literature in medicine and economics has found reasons to interpret their results with caution. Trials with inadequate concealed treatment, or partial unblinding are associated with larger estimates of treatment effects (Schulz et al., 1995), and many of the most cited RCTs worldwide suffer from issues with blinding and randomization among trial groups (Krauss, 2018). This paper considers only double-blind RCTs and finds evidence of an additional form of bias.

In the economics literature, Allcott (2015) estimates site selection bias in RCTs in the evaluation of an energy conservation program. Because environmentally friendly areas are more likely to both adopt the program first and respond well to treatment, earlier RCTs of a given program produced much larger efficacy estimates compared to subsequent trials. Given the cost of RCTs, it is unusual for the same intervention to be rigorously evaluated at more than a small handful of sites in economics (Allcott, 2015), so the pharmaceutical industry provides a unique setting for testing bias in evaluations. As in the literature above, I focus on RCT's as a consistent type of experiment, but my analysis applies to any bias due to the funder's

interest. Unlike Allcott (2015), I focus on documenting and explaining bias in this particular context, rather than highlighting a mechanism. This work is also related to a literature on the economics of clinical trials, such as identifying placebo effects (Malani, 2006) and the distortion of innovation away from long-term private research investment (Budish et al., 2015).

Section 2 presents institutional background on clinical trials and explains the setting. I outline my empirical strategy in Section 3, which also discusses my data and provides some initial descriptive analysis. I present my main results on the effect of sponsorship on reported drug efficacy in Section 4. Section 5 decomposes mechanisms, focusing on differential trial design and publication bias, and Section 6 concludes and discusses implications for the funding of clinical trials.

2 Clinical Trials and Psychiatric Drugs

2.1 Clinical Trials

The clinical trial development process involves huge financial stakes. There are the direct costs of conducting clinical trials, high failure rates, and the large opportunity cost of capital during the average of eight to twelve years of development (Danzon and Keuffel, 2014). Estimates of the research and development spending per drug approved range from $600 million to $2.6 billion (DiMasi et al., 2016; Prasad and Mailankody, 2017).[7] On the benefit side, the financial returns from bringing a new drug to market can be enormous. Among all cancer drugs approved during 1989–2017, half had cumulative sales of more than $5 billion and the upper 5% of these drugs had sales of more than $50 billion (Tay-Teo et al., 2019). Therefore, pharmaceutical firms balance large financial risks and strong incentives throughout the drug development and post-approval process.

Some of the clinical trials in my data are pre-market trials, which were conducted to assess the efficacy of a new drug. Drug development usually begins with pre-clinical testing of new molecules in non-human subjects. Subsequent clinical trials in humans are organized into several phases, with increasing scale and costs. Phase I clinical trials are conducted to assess the safety of new molecules in human subjects and often enroll only a few dozen patients. Drugs that demonstrate safety are then assessed for efficacy in Phase II clinical trials. Promising candidates proceed to Phase III clinical trials where the efficacy of the new drug is tested in a larger sample of hundreds or thousands of patients. Manufacturers submit these clinical trial reports for regulatory review. In the United States, the FDA is the regulatory body that approves new drugs. There is substantial leeway in interpreting the FDA's guidelines for pharmaceutical companies—the guidelines are "not intended to be immutable, nor are they to be used to stifle innovative approaches." For example, separate analysis of efficacy in demographic subsets is not required in most cases (US Food and Drug Administration, Center for Drug Evaluation and Research, 1977). For antidepressants, the FDA

[7] Some of these estimates been criticized due to the high assessment for capital costs and the confidential underlying data provided by drug makers (Avorn, 2015). However, alternate estimates without these assumptions are similar in magnitude (Adams and Brantner, 2006).

recommends three to five adequate and well-controlled clinical trials demonstrating substantial evidence of efficacy in order to support approval. The FDA also recommends testing new antidepressants both in trials against a placebo and against the current standard of treatment; the guidelines vary in other classes of drugs.

After a drug is approved, post-market clinical trials, also known as Phase IV trials, are continually conducted to assess the drug's efficacy for use in promotional material. The majority of clinical trials in my paper are post-market trials. After a new drug has been approved, clinical trials might be conducted for marketing by the original drug manufacturer, to demonstrate efficacy or a favorable side effect profile against a new competitor. Scientific publications are "the ultimate basis for most treatment decisions" (Davidoff et al., 2001) and their content affects physician's prescription choices (Azoulay, 2004; McKibbin, 2020). Publications of clinical trial results provide material for pharmaceutical sales representatives to cite in the promotion of drugs to physicians, also known as detailing.

Traditionally, pharmaceutical firms both financed and managed clinical trials. In the past three decades, an increasing share of clinical trial management has been contracted out to contract research organizations (CROs) and site management organizations (SMOs). CROs provide project management support for all components of trials, while SMOs find investigative sites, negotiate site contracts, train investigators, and recruit patients (Rettig, 2000). Typically, pharmaceutical firms make most high-level decisions and determine the approach and strategy of the clinical trial, while the CROs and SMOs help implement the day-to-day logistics.[8] Once the trial is completed, the results are often published in peer-reviewed journals, submitted as FDA Statistical and Medical Reviews[9], from individual pharmaceutical firms directly, or reported on clinical trial registries. This paper aggregates trials from all of these sources.

2.2 Why Antidepressant and Antipsychotic Drugs?

My analysis focuses on psychiatric medications for major depressive disorder and schizophrenia. These categories were chosen because of their prevalence, large economic costs, and the robust debate regarding their efficacy (Carroll, 2018). Potentially as a consequence, large meta-analyses for antidepressant and antipsychotics medications were published recently, which provide data on the near-universe of clinical trials in these categories.

Antidepressants and antipsychotics treat common diseases: 8.1% of American adults have depression in a given two week period and approximately 0.5% are currently diagnosed with schizophrenia (Wu et al., 2006; Brody et al., 2018). An even larger share of Americans take psychiatric medications, either to treat major depressive disorder, prophylactically or for maintenance once symptoms subside. 12.7% of the U.S. population over age 12 takes antidepressant medication in each month, a 64% increase from 1999–2014, and 1.6% take antipsychotics (Pratt et al., 2017; Moore and Mattison, 2017). In 2006, five out of the 35 drugs with the largest sales in the United States were antidepressants, and each of these drugs had annual

[8]This statement is based on interviews with clinical research scientists and managers at Boston-area pharmaceutical firms.

[9]The FDA Statistical and Medical Reviews are hosted on the FDA's website for all drugs approved after 1997; earlier reports can be made available through Freedom of Information Act requests.

sales of more than a billion dollars (Ioannidis, 2008).[10] The economic burden of depressive disorders in the United States is estimated to be $210 billion annually, which includes direct health costs, suicide-related costs, and workplace costs (Greenberg et al., 2015).

Psychiatric medications are not only prevalent, but particularly amenable for this analysis because of the vibrant debate regarding their efficacy, both in general (Ioannidis, 2008; Carroll, 2018; Kirsch, 2010), and for specific drugs within this class (Gartlehner et al., 2011). Many potentially substitutable drugs are used to treat major depressive disorder and, separately, schizophrenia; my paper considers 21 antidepressants and 15 antipsychotic drugs. The active efficacy debate among this large drug class has resulted in both numerous clinical trials and cases in which the same sets of drugs are tested in clinical trials conducted by different sponsors. This variation is essential to identifying the effect of sponsorship on drug efficacy.

One additional reason my work focuses on antidepressants and antipsychotics is the availability of comprehensive recent meta-analyses in both of these drug classes. These meta-analyses provided a listing of hundreds of clinical trials in these categories, and, in some cases, efficacy information and trial characteristics. Most clinical trials for psychiatric medications were published in the 1980s and 1990s before the existence of centralized clinical trial registries. Therefore, the process of identifying all relevant clinical trials is highly labor intensive. The availability of meta-analyses on all clinical trials in these classes allowed my analysis to contain a plausible universe of clinical trials.

Both antidepressant and antipsychotic drugs have distinct subclasses. Antidepressants were developed in several waves, beginning with the monoamine oxidase inhibitors in 1958 (Hillhouse and Porter, 2015). The earliest drugs included in my analysis are two tricyclic antidepressants: amitriptyline, which was approved by the FDA in 1961, and clomipramine, which was approved in Europe in 1970. Both are on the World Health Organization's Model List of Essential Medications. My analysis includes all second-generation antidepressants approved either in the United States, Europe, or Japan, as well as trazodone and nefazodone. Second-generation antidepressants include selective serotonin reuptake inhibitors (SSRIs) such as escitalopram (brand name Lexapro). It also includes atypical antidepressants such as bupropion (brand name Wellbutrin) and serotonin-norepinphrine reuptake inhibitors (SNRIs) such as duloxetine (brand name Cymbalta). This list of included antidepressants is based on the prior literature (Cipriani et al., 2018).[11] For antipsychotics, this analysis includes the first generation antipsychotics chlorpromazine (approved in 1957) and haloperidol (approved in 1967) along with thirteen second generations antipsychotics such as aripiprazole (brand name Abilify) and risperidone (band name Risperdal).[12] Similarly, the included drugs are based on prior literature (Leucht et al., 2013).

[10]These blockbuster drugs include venlafaxine (brand name Effexor), escitalopram (Lexapro), sertraline (Zoloft), bupropion (Wellbutrin), and duloxetine (Cymbalta).

[11]The full list of included antidepressants is agomelatine, amitriptyline, bupropion, clomipramine, citalopram, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, nefazodone, paroxetine, reboxetine, sertraline, trazodone, venlafaxine, vilazodone, and vortioxetine.

[12]The full list of included antipsychotics is amisulpride, aripiprazole, asenapine, chlorpromazine, clozapine, haloperidol, iloperidon, lurasidone, olanzapine, paliperidone, quetiapine, risperidone, sertindole, zipraidone, and zotepine.

2.3 Psychiatric Clinical Trial Data

The clinical trial data contain all available double-blind RCTs for either antidepressants or antipsychotics. The antidepressant clinical trial data is based on Cipriani et al. (2018), and includes placebo-controlled and head-to-head trials of 21 antidepressants used for the acute treatment of adults with major depressive disorder. This comprehensive meta-analysis searched the Cochrane Central Register of Controlled Trials, Embase, Medline, other databases, the websites of regulatory agencies, and international registers for all published and unpublished, double-blind RCTs. The included papers span from 1979 through January 8, 2016. This sample excludes non-controlled clinical trials, non-double blinded analysis, trials with pediatric populations, and trials for indications other than major depressive disorder. Leucht et al. (2013) conducted a similar large meta-analysis of antipsychotic clinical trials. Their analysis incorporated data from Cochrane Schizophrenia Group's Registrar, Medline, Embase, the Cochrane Central Register of Controlled Trials, and ClinicalTrials.gov. Their papers span from 1959 through September 1, 2018. Both meta-analyses also incorporated data from FDA reports, Freedom of Information Act requests and data requested from pharmaceutical companies. Both were multi-year projects of over a dozen authors and effectively contain the universe of all available clinical trials on these drugs.

For all possible trials, I obtained the original publications or clinical trial reports. In a few cases the original publications or reports were available in non-English language journals or have since been removed from company archives. For the antidepressant data, the full original reports provide more detailed funding data and helpful case studies. For the antipsychotics, these primary sources are used to obtain efficacy, funding data, and additional trial characteristics. Occasionally, the original clinical trial reports contain additional arms that are not included in the meta-analyses. In order to correctly define the full set of drugs in a trial, I include these additional treatment arms as well. The final dataset contains efficacy and sponsorship information, as well as the length of the trial, the drug's dosage, total enrollment, recruitment area, treatment setting and patient characteristics such as the mean age, gender, dropout rate and baseline severity. In my final analysis sample, I exclude trials and treatment arms with missing efficacy information.

My clinical trial data contain both published and unpublished trials; unpublished trials are available from FDA Statistical and Medical Reviews for approved drugs, from clinical trial registries, or directly from pharmaceutical firms.[13]

2.3.1 Defining Terms

The subsequent exposition relies on defining a few key terminologies. First, a *drug set* refers to the unique combination of drugs in a clinical trial. For example, paroxetine versus placebo is one drug set; paroxetine versus venlafaxine is another; paroxetine versus venlafaxine versus placebo is yet another. A *drug pair* refers to two drugs compared in the same trial. For example, a trial comparing paroxetine versus venlafaxine and a trial comparing paroxetine versus venlafaxine versus placebo both have the same drug pair of paroxe-

[13] I thank Dr. Erick Turner for sharing FDA Statistical and Medical Reviews he obtained via the Freedom of Information Act.

tine versus venlafaxine, though they test different drug combinations. A *trial* is a published or unpublished RCT. Each trial contains at least two treatment *arms*. A treatment arm is the unit at which randomization occurs. Arms are often unique drugs but occasionally refer to unique drug and dosage combinations.

2.3.2 Defining Sponsorship

I define a treatment arm as sponsored if any of the following cases hold: the text indicates that the trial was funded by the company that manufactured or marketed the drug, one of the authors was affiliated with the company, or the data came from documents provided on the company website, the authors listed the names of the drug manufacturers in their declaration of conflicts of interest. Any of a drug's manufacturers or marketers in any country are considered sponsors.[14] For example, consider a trial that compares escitalopram to venlafaxine and to a placebo. Suppose one author of that trial was affiliated with Forest Labs, the firm that markets escitalopram in the United States. In this case, the citalopram arm in that trial would be considered sponsored. If there were no other funding sources, the venlafaxine and placebo arms would be considered unsponsored.

Sponsorship was defined for each treatment arm in the antidepressant meta-analysis; I applied the same definition to the antipsychotic trials.[15] For each antipsychotic drug, I constructed a list of that drug's global manufacturers and marketers each year. Equivalent to the definition for antidepressants, a treatment arm was considered sponsored if any of the drug's current manufacturer or marketers funded, authored, or were acknowledged in the trial's report.

2.3.3 Defining Efficacy

Efficacy for psychiatric drugs is measured on an observer-rated scale. A psychiatrist or psychologist will observe a patient and map their current or past behavior to a numeric score. The most common scale for antidepressants is the Hamilton Score for Depression (HAMD) (Naudet et al., 2011; Taylor et al., 2014); this is available for 85% of the antidepressant sample in my analysis. Another 5% of trials use the Montgomery–Åsberg Depression Rating Scale (MADRS) and the remaining 10% of trials do not specify their scale. The efficacy outcome for antidepressants is the share of patients that responded to treatment, as defined by a reduction of greater than or equal to 50% of the total depression score. Response is measured at eight weeks; if this length is not reported, the authors use the closest length of time available. This outcome is the

[14] This is the same as Cipriani et al. (2018)'s definition of sponsorship, except they consider cases where the authors list the drug manufacturers in their conflict of interest statements as unclear sponsorship, but at high risk of bias. I report summary statistics on sponsorship with and without conflict of interest sponsorship in Table 1. I also consider robustness to the definition of sponsorship in Table 6.

[15] In three cases, I revised the Cipriani et al. (2018) sponsorship definitions based on likely errors after reviewing the initial publications. Using exclusively the original coding for antidepressants increases most point estimates and makes no significant difference in my results. Specifically, Åberg-Wistedt et al. (2000) and Lydiard et al. (1997) acknowledged funding from Pfizer, so I consider the sertraline arm in both trials as sponsored. Amsterdam et al. (1986) was sponsored by AstraZenca, so amitriptyline is considered sponsored.

standard outcome for measuring efficacy for antidepressants (Cipriani et al., 2018); in robustness checks, I also consider the percent decline in the total depression score.

Observer-rated scales for antipsychotics include the Positive and Negative Syndrome Scale (PANSS), the Brief Psychiatric Rating Scale (BPRS), and the Clinical Global Impressions–Schizophrenia Scale (CGI-S). The standard outcome used to measure efficacy for antipsychotics is the mean change in the total PANSS score or, if the PANSS score is not available, the BPRS or the CGI-S, in that order (Leucht et al., 2013). For both drug classes, outcomes are normalized so that higher values represent greater efficacy (e.g. a larger share of patients respond to treatment, a greater decline in the PANSS score). To combine the antidepressant and antipsychotic outcomes in a single framework, I standardize each score to have a mean of zero and a standard deviation of one. In robustness checks, I also consider the percent decline in these antipsychotic scales, rather than the absolute change.

2.3.4 Defining Trial Characteristics

I incorporate data on trial characteristics to assess the comparability of sponsored versus unsponsored arms and evaluate potential mechanisms. Year refers to the year the trial was either published or made publically available. I also compute the year relative to approval, which is the year the trial was made available relative to the year the FDA first granted approval for the tested drug for either of two categories: major depressive disorder or schizophrenia. Most trials for antidepressants use the HAMD score and most trials for schizophrenia use one of the PANSS, BPRS or CGI-S score; I include an indicator for whether the trial arm uses one of these standard outcomes. I also code an indicator for whether the trial was registered and has a National Clinical Trial (NCT) number.

In terms of trial design, I observe the number of patients randomized to each arm of the trial, the length of time the trial was conducted in weeks, and the baseline severity of the enrolled patients at the start of the trial. As several different scales are used to measure the severity of major depressive disorder and schizophrenia (see Section 2.3.3), baseline severity is standardized to have mean zero and standard deviation one within each scale. I also record the drug dosage (in milligrams) for each trial arm. Some trials have a lower initial dosage and scale up this amount over the course of the trial; for comparability I consider the minimum dosage. Lastly, I observe the share of patients that dropout the average age, and the share female within each trial arm.

2.4 Other Data Sets

Supplemental data include state drug utilization data from the Medicaid Drug Rebate Program from 1991-2017. This data reports total prescriptions and dollars reimbursed for covered outpatient drugs paid by state Medicaid agencies. In the Medicaid utilization data, drugs are identified by their National Drug Code (NDC). I use the FDA's Approved Drug Products with Therapeutic Equivalence Evaluations publication (commonly known as the Orange Book) to link the NDC codes to the generic drug names.

This work also incorporates clinical trial data from the ClinicalTrials.gov registry. This registry contains the conditions, drugs, interventions, authors, and funders for over 300,000 clinical trials. The first clinical trials in this registry were submitted in 1999; initially just over a thousand clinical trials were added annually. The registry grew substantially with the International Committee of Medical Journal Editors' (ICMJE) requirement that clinical trials published in any of their affiliated journals had to be pre-registered starting in 2005. In recent years, ClinicalTrials.gov has been growing by ten to twenty thousand clinical trials annually.

2.5 Sample Construction and Summary Statistics

The antidepressant and antipsychotic meta-analyses contain 732 total clinical trials. For 656 of the 732 trials, I obtained the original publications or clinical trial reports. After dropping observations with missing efficacy or sponsorship information, my final clinical trial data contain 586 trials and 1,412 treatment arms.

Table 1 presents summary statistics on trial characteristics. The average trial in my sample was published in 2001, and occurred just over ten years after the drug gained FDA approval. Just under half of all treatment arms are considered sponsored, and six percent are considered sponsored due to conflicts of interest alone.[16] Approximately three-quarters of the data are from antidepressant trials and the remaining quarter are from antipsychotic trials. The average treatment arm enrolled 100 patients and the average trial length was just over eight weeks. On average 29% of patients dropped out of each arm before the trial completed. These arms enrolled 52% women on average, and the mean patient was 42 years old. Since my identification strategy uses variation in sponsorship within a drug set or drug pair, I present summary statistics for the subset of trials with variation in sponsorship separately; these subsets are comparable to the full sample in terms of trial characteristics.

In total, my analysis contains 36 unique drugs, 25 of which are included in at least one drug set with variation in sponsorship. Appendix Figure B1 shows the share of trials in which a drug is sponsored in relation to the drug's FDA approval year.[17] Most antidepressant and antipsychotic drugs were approved in the 1980s, 1990s, and early 2000s. Older drugs are sponsored the least often in my analysis sample. This is likely because these drugs no longer have patent protection during the years covered in my sample, thus the original manufacturer has weak financial incentives to fund clinical trials with these drugs. These older drugs might also no longer be the comparison standard of treatment against which new drugs are tested. In contrast, the very newest drugs always have the drug's manufacturer or marketer involved.

[16] A total of 319 arms (23% of the total) are placebo arms, and would never be considered sponsored.

[17] There are six drugs in my analysis that are not yet approved in the United States. These are excluded from this figure, given that the x-axis is the United States FDA approval date. These are agomelatine, amisulpride, milnacipran, reboxetine, sertindole, and zotepine.

3 The Effect of Sponsorship: Empirical Strategy and Results

3.1 Description of Sponsorship Variation

In total, the final clinical trial data contain 215 drug sets. As shown in Table 2, 43 of these drug sets have variation in sponsorship. In the other drug sets, each drug is either always sponsored or always unsponsored. These sets with variation in sponsorship contain 230 trials and 499 treatment arms. Turning to the subsample with variation within drug pairs, there are a total of 203 drug pairs. Of these 60 drug pairs have variation in sponsorship, which corresponds to 400 unique trials and 900 unique treatment arms.

The main types of drug combinations are presented graphically in Table 2. The first category ("Active vs. Placebo") compares a given psychiatric drug ("drug A") to a placebo directly. Among the sample with variation in sponsorship, this category contains 5 drug sets, 25% of the trials, and 25% of the treatment arms. Within these drug sets, some trials are sponsored by the company that manufactures drug A ("company A"). These trials could have additional funders; it is sufficient that company A is affiliated with the trial in some capacity. Other trials have alternate funding not provided by company A. In this case drug A would be considered unsponsored.

The second category in Table 2 ("Active vs. Active") contains drug sets that compare an active drug to another active drug. This contains 35 drug sets, 70% of the trials, and 67% of the treatment arms. There are three main subgroups considered; in each, a given psychiatric drug ("drug A") varies in sponsorship. First, the company that manufacturers the other active drug ("company B") could never be involved in the trial. Secondly, company B could always be involved. In the third subgroup, the sponsorship interests of both active arms vary.[18] The last category ("Three or More Drugs") contains three drug sets, 6% of the trials, and 9% of the treatment arms. In each of these drug sets, coincidentally, only one drug has varying sponsorship interests.[19]

Drugs are also included in clinical trials as a control group in an alternate pharmaceutical firm's analysis, as in the example cited in my introduction. Some clinical trials are funded by government organizations. For example, my paper includes trials from the National Institutes of Health and the National Institute of Mental Health, the São Paulo Research Foundation and the Deutsche Forschungsgemeinschaft.

The focus of some trials ranged from neuropsychological test performance, to saliva concentrations in patients taking antidepressants, to genetic predictors of drug-specific responses. Most trials in my sample were conducted to study the efficacy of a drug for either major depressive disorder or schizophrenia. All clinical trials included in my analysis reported a consistent set of primary efficacy outcomes, regardless of

[18] More than one type of variation could occur within a drug set. For example, there are four trials in the drug set that directly compares olanzapine and risperidone, two antipsychotics. In the first trial, olanzapine is sponsored and risperidone is not; in the second and third trials risperidone is sponsored and olanzapine is not; in the fourth trial, neither drug is sponsored. The first and fourth trials could be compared in the first subgroup, as could the second, third, and fourth. In addition, the first, second, and third trials could be compared in the third subgroup. The regression framework aggregates all of these comparisons so that each arm in the sample has the same weight.

[19] For example, in the drug set that includes fluoxetine, venlafaxine, and a placebo arm, only fluoxetine has varying sponsorship. Pfizer (the manufacturer of venlafaxine) is associated with each of the five trials in this drug set.

the trial's primary purpose.

Figure 1(a) plots the average share of treatment arms that are sponsored by the number of years since the tested drug gained FDA approval. Placebo arms are not included. On average, just over 60% of all active treatment arms are sponsored. Prior to FDA approval, most treatment arms are tested in trials that are conducted by that drug's manufacturer. For the two decades after FDA approval, approximately half of a drug's arms are sponsored. Thirty or more years after FDA approval, almost none of the arms are still sponsored. Figure 1(b) plots the share of arms by the number of year relative to the FDA approval year. The majority of the trials occur just before and in the ten years immediately after FDA approval.

3.2 Estimating Equations

In my main analysis, I estimate the following specification:

$$y_{ij} = \alpha + \beta Sponsor_{ij} + X_{ij}\gamma + G_{d(i),s(j)} + \varepsilon_{ij} \tag{1}$$

where y_{ij} is the efficacy for arm i in trial j. The coefficient of interest is on $Sponsor_{ij}$, which is a dummy for whether arm i was sponsored in trial j. I control for X_{ij} which denotes the type of measurement scale for arm i and the year published for trial j. As described in Section 2.3.3, some trials report efficacy using alternate depression or schizophrenia scales; I include fixed effects for each type of measurement scale to control for any mean differences in outcomes across these scales. I control for the trial's publication year in ten year bins and include a separate fixed effect for unpublished trials. Standard errors are robust to heteroscedasticity and clustered at the trial level, since most unobserved shocks would occur for all arms in a clinical trial.

The outcome y_{ij} is computed *relative* to the placebo arm in trial j, if available, or least effective arm, otherwise. For example, suppose the standardized efficacy for an arm in a given trial is 0.4, while the standardized efficacy of the placebo arm is 0.3. Then the *relative* standardized efficacy for the arm y_{ij} is 0.1. A given arm can be the least effective arm in its own trial; in that case its relative efficacy is zero.[20] Conceptually, this is similar to adding trial fixed effects.

Most importantly, $G_{d(i),s(j)}$ is a dummy for each unique drug $d(i)$ in each separate drug set $s(j)$. Each arm i can be mapped to a unique drug $d(i)$. In most cases, each arm in a trial is a unique drug; in a few cases, a trial may contain multiple arms with the same drug and different dosages. As described in Section 2.3.1, a drug set is the unique combination of drugs in a clinical trial; each trial j can be mapped to a single drug set $s(j)$. Therefore, paroxetine has a separate fixed effect in a trial comparing paroxetine to citalopram, in a trial comparing paroxetine to placebo, and in a trial comparing paroxetine to citalopram and a placebo, since these are separate drug combinations. This is key to my analysis, because it ensures that the sponsorship effect is estimated using differences in funding sources among trial comparing the exact

[20] Appendix Table B4, panel A, includes results for non-relative outcomes as well.

same set of drugs. In this example, β reflects the effect of sponsoring e.g. paroxetine, within the set of trials that directly compare e.g. paroxetine to citalopram. In my first set of specifications, the sponsorship effect is conservatively identified using only the drug sets that have variation in sponsorship. Appendix Table B1, column (1), provides a more detailed example of this fixed effects structure.

In an alternate empirical strategy, I include a dummy for each drug in each separate drug pair. In this case, I estimate the following specification:

$$y_{ij} = \alpha + \beta Sponsor_{ij} + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{2}$$

where each term is identical to equation 1 above, except $G_{d(i),p(j)}$ is a separate fixed effect for each unique drug $d(i)$ when compared in each separate drug pair $p(j)$. Each trial j examines potentially multiple drug pairs $p(j)$. For example, paroxetine has the same fixed effect in a trial comparing paroxetine to citalopram as in a trial comparing paroxetine to both citalopram and a placebo, since both trials contain the same drug pair: paroxetine and citalopram. Conceptually, this specification assumes that the presence of an additional arm should not affect the comparison between an existing drug pair.[21] One technical point regarding this fixed effect structure is that a trial with three unique drugs will contain three drug pairs. Therefore, each arm in that trial will be counted in two separate drug pairs.[22] Therefore, I re-weight the observations so that each treatment arm receives the same weight. Appendix Table B1, column (2), provides a more detailed example.

3.3 Difference in Difference Framework

The empirical framework in this paper can be succinctly summarized in Table 3. This contains all antidepressant drug sets that compare one active drug to a placebo and have variation in sponsorship (the "Active vs. Placebo" row in Table 2).[23] Each row is a unique drug set and, in my initial empirical specification, each drug in each row would receive its own fixed effect.

As mentioned in Section 2.3.3, the efficacy of antidepressants is measured as the share of patients that respond to treatment. In the first row, I consider trials that directly compare paroxetine to a placebo. There are 33 such trials; 32 in which paroxetine is sponsored and one trial in which paroxetine is not sponsored. In the trials where paroxetine is sponsored, an average of 47% of patients receiving paroxetine respond to treatment, while an average of 32% of patients respond to the placebo. Therefore, on average, paroxetine is 15 percentage points more effective than the placebo. Turning to trial in which paroxetine is

[21] There are some practical considerations for including additional arms in trials. If a trial comparing an active drug (i.e. drug A) to a placebo fails to show efficacy for the active drug, then it is considered evidence of a lack of efficacy for drug A. However, if the trial included a drug that was known to be effective (i.e. drug B) and drug B failed to show efficacy against the placebo as well, then the trial would be considered a failed trial that does not speak to the efficacy of drug A. However, the drug-pair specification would still be valid unless the existence of drug B changed the efficacy of drug A relative to the placebo.

[22] In the trials with n treatment arms, each drug will be counted in $n-1$ drug pairs. Thus each treatment arm is weighted by $\frac{1}{n-1}$, where n is the number of treatment arms in the trial.

[23] There are no antipsychotic drug sets that compare one active drug to a placebo that have variation in sponsorship.

not sponsored, 25% of patients receiving paroxetine respond to treatment, while 23% of patients responded to the placebo, so, paroxetine is 2 percentage points more effective than the placebo. As shown in the last column, the difference in difference estimate of the sponsorship effect for paroxetine versus a placebo is 13 percentage points. Averaging across all antidepressant drug sets that compare an active antidepressant drug to a placebo, and weighting by the number of trials, the mean sponsorship effect is 4.8 percentage points (Table 3, row 1).

Table 4 presents the analogous estimates for the "Active vs. Active" category in Table 2. The left-hand column now lists both drugs in the drug set. The first drug listed varies in sponsorship interests across trials in that drug set. The second drug's sponsorship interests remain constant.[24] The first row considers the drug set comparing amitriptyline and paroxetine. If amitriptyline is sponsored and paroxetine is not, an average of 66% of patients respond to amitriptyline, while 65% of patients respond to paroxetine, an average of one percentage point higher efficacy for amitriptyline. In the fourteen trials where neither amitriptyline nor paroxetine were sponsored, an average of 46.5% of patients responded to amitriptyline, while 47.4% responded to paroxetine, an average of 0.8 percentage points lower efficacy for amitriptyline. Thus the sponsorship effect, i.e. the "difference in difference," is 1.8 percentage points. Averaging across all antidepressant drug sets in this category, and weighting by the number of trials, the average sponsorship effect is 6.4 percentage points.

Table 4 contains antidepressant trials; the antipsychotic trials are shown in Appendix Table B2. For antipsychotics, efficacy is measured as the average decline in the schizophrenia score, as described in Section 2.3.3, and the average sponsorship effect is 0.10 points.

3.4 The Effect of Sponsorship on Reported Efficacy

3.4.1 Efficacy and Statistical Significance

The difference-in-difference estimates from Table 3, Table 4, and Appendix Table B2 can be combined in a regression framework. The coefficient on sponsorship in equations 1 and 2 is analogous to the average of the difference-in-difference values in Tables 3, 4, and Appendix Table B2, weighted by the number of arms in each estimate.

As shown in Table 5, column (1a), I find that a sponsored drug is 0.15 standard deviations more effective than the same drug in the same drug set without sponsorship. Controlling for the publication year and the type of psychiatric score in column (2a) makes no different in this result. The sponsorship effect is 36% of the average relative efficacy difference of 0.42 standard deviations between a given arm and either the placebo or the least effective arm in that trial. Therefore, the funding interests of a given drug can explain a third of the relative efficacy of that drug.

[24] If both drugs vary in sponsorship, they are included as two separate entries. Therefore, the total trial count is slightly inflated to include the variation in sponsorship for both drugs. In the regression specifications, these trials are not over-counted. For antidepressants, both drugs vary in sponsorship in the paroxetine vs. fluvoxamine and reboxetine vs. citalopram drug sets. For antipsychotics, this occurs in the olanzapine vs. aripiprazole and ziprasidone vs. olanzapine drug sets.

In column (3a), the outcome y_{ij} is an indicator for whether the arm was statistically significantly more effective, relative to the placebo arm or least effective arm in that trial. Appendix A provides details on the construction of this variable. On average, sponsored arms are 10 percentage points more likely to be statistically significant at the 5% level. This represents a 47% increase over the baseline of 22% statistical significance. As described in Section 2.1, the FDA suggests that pharmaceutical companies present at least three statistically significant clinical trials to gain FDA approval for antidepressants, so this increase in significance may be pivotal for gaining regulatory approval. While the statistical significance threshold in most clinical trial publications is 5%, trials alternately report results at the 10% level. In column (4a), the outcome is an indicator for whether the arm was statistically significant at the 10% level. This coefficient is positive, but not significant. In column (5a), the outcome is an indicator for whether the given arm was the most effective arm in that trial. Sponsored arms are 0.16 percentage points more likely to be the most effective arm, compared with that same drug evaluated in the drug set, but without sponsorship. This is a 41% increase over a baseline of 0.39.[25]

In Panel B, I show that including drug by drug pair fixed effects, rather than drug by drug set fixed effects, yields very similar estimates in magnitude, with more statistical precision. As described in Section 3.2, these estimates are weighted to adjust for the mechanical over-counting of larger trials. These results are all similar in magnitude but more precisely estimated than the results in panel (a). In panel (a), with drug by drug pair fixed effects, the sponsorship effect is identified using only trials with variation in sponsorship within a specific drug set. In the drug by drug pair fixed effect specification, trials with differences in sponsorship in any of the occurrences of a drug pair can be used to identify the sponsorship effect. My preferred specifications is columns (2a) and (2b), which use the relative standardized outcome and control for the measurement scale and calendar year. The effect of sponsorship on reported drug efficacy is 0.15 standard deviations, or 36% of the average relative efficacy.[26]

3.4.2 Comparability of Sponsored and Not Sponsored Arms

Identification of the causal effect of sponsorship on drug efficacy requires that, *within the same drug and drug pair*, sponsored and unsponsored arms are equivalent tests of a drug's efficacy. This could be violated if sponsored and unsponsored arms within a drug and drug pair are systematically different and those differences relate to the measured efficacy of those drugs. To assess this issue, Figure 2 presents differences in general characteristics and trial design for sponsored relative to unsponsored arms. The left panel presents the overall, unconditional differences between sponsored and unsponsored arms. For each characteristic c_{ij} for arm i in trial j, I estimate

$$c_{ij} = \alpha + \beta Sponsor_{ij} + \varepsilon_{ij} \tag{3}$$

[25] Some trials have more than two arms, so the mean of this variable is below 0.50.

[26] In panel (b), the standardized efficacy is computed relative to the other drug pair in that trial. Estimates are highly similar if I use the standardized efficacy computed relative to all arms in that trial, as in panel (a).

and plot the coefficient on $Sponsor_{ij}$, a dummy for whether the arm i is sponsored in trial j, along with 95% confidence intervals clustered at the trial level. The right panel presents the differences between sponsored and unsponsored arms within a drug pair. In this case, I estimate

$$c_{ij} = \alpha + \beta Sponsor_{ij} + G_{d(i),p(j)} + \varepsilon_{ij} \tag{4}$$

and again plot the coefficient on $Sponsor_{ij}$. Here, $G_{d(i),p(j)}$ is a fixed effect for each drug in each drug pair, as defined in Section 3.2. As shown in the left panel of Figure 2, sponsored and unsponsored arms are very similar in terms of registration status, length of trial, whether the outcome was a standard metric, the baseline severity of patients, the dosage, and the share of female patients. However, there are a few notable exceptions.

Overall, sponsored arms occur one standard deviation, or approximately ten years, earlier relative to the drug's approval year. This reiterates the findings from Figure 1; drugs are more likely to be sponsored earlier in their life cycle. Within drug pairs however (right panel), the difference between sponsored arms falls to 0.4 standard deviations or about four years. Similarly, overall sponsored arms enroll 0.2 standard deviation or 15 more patients per arm, while within a drug pair, sponsored arms enroll only a statistically insignificant 0.1 standard deviations more patients. This pattern is also seen with the dropout rate; sponsored arms have a 0.15 standard deviation smaller dropout rate, while within drug pairs, the difference in dropout rates is an statistically insignificant -0.06 standard deviations. Within a drug pair, the only statistically significant differences in trial observables are the mean age of enrollees (which is considered and rejected as a mechanism in Section 4.1) and the aforementioned trial timing.

Trial timing could be a concern if sponsored arms occur at different points in a drug's life cycle *and* those different points represent different tests of a drug's efficacy. Appendix Figure B2 plots the average efficacy of drugs relative to their approval year. There is a slight decrease in relative drug efficacy around the time of approval. This decrease might be explained by mean reversion – by construction, this figure only includes drugs that have made it through the FDA approval process. Potentially some drugs got unexpectedly high efficacy draws and therefore were able to gain FDA approval. After approval, their mean efficacy would decrease to match their true efficacy. After approval, both the standardized drug efficacy and the standardized drug efficacy relative to the placebo or least effective arm are fairly stable over time. To assess the importance of mean reversion, Section 3.4.3 presents results that control for trial order, year relative to drug approval, and restricts to only post-approval drug trials. In all cases, the sponsorship effect is similar, suggesting that mean reversion cannot explain most of the sponsorship effect.

3.4.3 Robustness

As describe in the previous section, one potential concern with my analysis is that a drug's manufacturer or marketer is more likely to be involved in earlier trials for that drug. I can account for any systematic changes in efficacy over the drug's life cycle by controlling for the publication order of the trial within the

drug set as shown in Table 6, column (2a). Estimates using variation with drug pairs is presented in column (2b). In either case, controlling for the publication order slightly decreased the sponsorship effect estimate by approximately 10%, though the difference is not statistically significant. In a similar check, columns (3a) and (3b) control for the year relative to the drug's approval year; these estimates are very similar to the previous column As an additional test of whether the FDA approval benchmark is distortionary, I restrict my sample to only post-approval trials as shown in Table 6, column (4a) and (4b). Again, the point estimates decrease slightly but the estimate of the sponsorship effect remains statistically significant.

As described in Section 2.3.2, some trials are considered sponsored because the authors listed the names of the drug manufacturers in their declaration of conflicts of interest, rather than because the trial was directly sponsored by the company, one of the authors was affiliated with the company, or the documents were solely provided by the company. I examine robustness to excluding sponsorship definitions based on only conflict of interest statements (column (5a) and (5b)). This change decreases the sponsorship effect, suggesting that conflict of interest sponsorship affects reported drug efficacy.

Finally, I consider robustness to alternate weighting schemes. Conceptually, the counterfactual for sponsorship would involve altering the funding for a given drug in a particular clinical trial. Therefore, my analysis weights each treatment arm equally. However, a hypothetical conceptual experiment could involve randomizing sponsorship of drugs at the patient level. This approach might be accurate if a drug's manufacturer's involvement could vary at the patient level. Alternately, this weighting may be preferable if physicians interpret the results for each patient in a trial individually, instead of considering each trial as an observation. In either case, I also present estimates that are weighted by the total trial enrollment (column (6a)). The drug by drug set fixed effect estimates are very similar to the baseline estimates, though the drug by drug pair estimates in Panel B are smaller. Since my analysis considers the effect of changing trial funding, not patient-level funding, my preferred estimates are not weighted by trial enrollment.

Prior literature has used either only drug fixed effects or no fixed effects to estimate the sponsorship effect. For completeness, Appendix Table B4 presents results for even less restrictive fixed effects, such as only drug controls (column 3), or no controls (column 4).[27] Each of the estimates of the sponsorship effect with the relative standardized outcome is significantly positive and robust - though this does not necessarily reflect a causal sponsorship effect. For example, in column (4), this merely reflects that active drugs are both more effective and more likely to be sponsored than a placebo. The estimates that use the standardized outcome y_{ij} are less robust and, in column (3) with just drug controls the estimates are statistically indistinguishable from zero. The absolute efficacy across trials conducted at different times and with different patient populations can be very different and is difficult to compare across trials. In particular, the placebo efficacy in trials run by pharmaceutical firms is often lower than the placebo efficacy in trials with alternate funding. Drugs with lower absolute efficacy in those funded trials might have high efficacy

[27]The appendix of Cipriani et al. (2018) reports whether the absolute efficacy of a drug varies depending on its sponsorship status. The authors find that sponsorship status does not affect drug efficacy, as presented in the first two columns of Table B4, panel C.

relative to the placebo group. Within a trial, all outcomes are usually interpreted relative to the placebo effect or relative to the other drug arm. This is accomplished by considering the relative standardized outcome, as in Panel A.

3.5 Heterogeneity by Drug Type and Financial Incentives

The effect of funding could differ by drug class. To explore this distinction, Table 7 decomposes the analysis by the class of drug—antidepressant or antipsychotic. The first column reproduces the baseline estimate from Table 5, panel (a), column (2a) and panel (b), column (2b). The next set of columns restrict the sample to only antidepressants. The outcome in columns (2a) and (2b) is the same relative standardized efficacy as in the baseline specification. The sponsorship effect for only antidepressants is similar in magnitude to the combined baseline specification. I also present estimates using the original, non-standardized outcomes. For antidepressants, the original outcome is the share of patients that respond to treatment, as described in Section 2.3.3. As shown in column (3a), sponsored arms have a 3 percentage point higher response rate than non-sponsored arms for the same drug and drug set. This is a 50% increase, compared to the average share of patients that respond to treatment relative to the placebo or least effective arm of 6%. Columns (2a) and (3b) are identical in statistical significance. Column (4a) presents results with the percent decline in the observer-rated depression score as the outcome. This is not the standard antidepressant outcome in the medical literature, but the estimate is similarly positive and statistically significant.

The second set of columns considers only antipsychotics. The outcome in columns (5a) and (5b) is the relative standardized efficacy, as in the baseline specification. The sponsorship effect is smaller in the antipsychotic subsample. In addition, antipsychotics are a small subsample of the analysis sample, so none of the results are statistically significant. There are several common scales for measuring schizophrenia. For brevity and clarity, this table only reports the mean decline in the PANSS (see Section 2.3.3) in column (6a). Approximately two-thirds of the antipsychotic trials recorded the mean decline in the PANSS, thus the sample size falls from columns (5a) to (6a). The baseline standardized outcome allows for the full antipsychotic sample to be included. Columns (7a) and (7b) present results using the percent decline in the observer-rated schizophrenia score. This is not the standard antipsychotic outcome; as with the other antipsychotic outcomes, the sponsorship effect is positive but not significant.

The sponsorship effect could also be heterogeneous with respect to the financial incentives of pharmaceutical firms. If the potential market for a given drug is larger due to more prescriptions or fewer competitors within a subclass, there might be additional incentives to obtain higher reported efficacy for a given drug. To assess this correlation, I compute the sponsorship effect separately for each drug by estimating:

$$y_{ij} = \alpha + Sponsor_{ij} + \sum_{d} \eta_d Sponsor_{ij} * d(i) + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{5}$$

where $d(i)$ is an indicator for each drug. Recall that i indexes arms and j indexes trials. As in equation 2, the

$G_{d(i),p(j)}$ are drug by drug pair fixed effects. Figure 3 plots the coefficients for each drug η_d against a proxy for market size: the total Medicaid prescriptions in the five years after FDA approval for that drug. This relationship could be driven either by high projected sales incentivizing a high sponsorship effect or by a high sponsorship effect driving higher sales. In either case, the positive correlation between the sponsorship effect and prescriptions does show that this sponsorship effect is related to real market conditions.

3.6 External Validity

3.6.1 Which Drug Trials Have Variation in Sponsorship?

The identification is driven by the subset of drug sets or pairs that have variation in sponsorship. If this sample is highly selected, these estimates might not be applicable to the full sample of psychiatric drugs in particular, or clinical trials in general. To assess which types of trials have variation in sponsorship, Figure 4 presents the network of comparisons between drugs. Each circle represents an antidepressant drug. The drugs are arranged counterclockwise in the order their first generic entered the United States market.[28] Each line represents a clinical trial that directly compares these two drugs, weighted by the number of trials. The solid maroon lines refer to comparisons that have variation in sponsorship; the gray dashed lines refer to comparisons that do not have variation in sponsorship.[29]

As seen in Figure 4, one of the best predictors of variation in sponsorship is the generic entry year.[30] Among the drugs in the top section of the network plot, which have earlier generic entrants, most drug pairs have variation in sponsorship as the connections are marked by solid maroon lines. Among the drugs in the bottom section of the plot, which do not yet have generic entrants, none of the drug pairs have variation in sponsorship. Therefore, these connections all marked by dashed gray lines. Once a drug has generic competitors, that drug might be included more often as a control in other trials. This pattern also holds in Appendix Figure B3, which plots the network for schizophrenia drugs. In terms of external validity, most drug pairs have variation in sponsorship, and those without will potentially acquire variation in future years. Therefore, my analysis sample is representative among the established market. The effect of sponsorship is also consistent across different types of variation, as shown in Appendix Table B3.

3.6.2 Predictors of Variation in Sponsorship

Table 8 plots the share of trials with drug combinations that have variation in funding by various characteristics. Among antidepressants, the drug classes of tricyclics and SSRIs are most likely to be a part of drug sets or pairs with variation in funding. The former are older drugs that are often included as control arms in other trials, and the latter are the most commonly prescribed type of antidepressant, which might be more

[28] In my sample, generic entry occurs an average of twelve years after FDA approval.

[29] In this figure, I define variation based on drug pair combinations. If I define variation based on drug sets that have variation in sponsorship, teh figure looks very similar.

[30] The network graph looks very similar if the drugs are plotted in the order of FDA approval year, but the pattern of which drug pairs have variation in sponsorship is slightly less striking.

likely to be included in trials conducted for marketing. trials with placebo arms are mechanically less likely to be part of drug sets or pairs with variation in funding, since the placebo arm is always unsponsored. A later drug approval year also predicts less variation in sponsorship. Drugs that were approved in later years have less time to be included in trials with different funding sources. The number of trials in a given drug set is also strongly predictive of variation in sponsorship within that drug set, as variation in sponsorship requires multiple trials with different funders.

The subsample of trials with variation in sponsorship might also depend on the results of previous trials. In particular, firms might be more likely to test drugs that had higher efficacy in previous trials. These drugs could have larger potential financial returns. Figure 5 presents the relationship between a given trial's efficacy results and the probability that a future trial assessing that drug pair will be conducted. I find that arms with higher reported efficacy are more likely to be tested in future trials, and that those trials are more likely to be sponsored. Therefore, the set of trials with variation in sponsorship probably represent more effective drugs than the full sample.

4 Mechanisms

The sponsorship effect could be driven by two classes of mechanisms: trial design or publication bias. The first class covers all cases that occur before or during data collection (i.e. ex-ante mechanisms). Interviews with clinical trial managers suggests one main potential mechanism for conflicts of interest to manifest before or during data collection is in the choice of inferior comparator (Østengaard et al., 2020). As my research design accounts for the choice of comparator, I test for differences in other aspects of trial design such as observable trial characteristics or patient selection. The second class of mechanisms covers all cases that occur after data collection (i.e. ex-post mechanisms). Examples of ex-post mechanisms given by clinical trial managers include fabrication of data and blocking access to data. I cannot test for fabrication of the data, but I test for blocking access to the data in terms of publication bias.

4.1 Trial Design

Trial design and patient selection can substantially affect reported efficacy for psychiatric medications. As an example, in 1996, an unsponsored meta-analysis concluded that St. John's wort, an herbal supplement, was "more effective than placebo for the treatment of mild to moderately severe depression" (Linde et al., 1996). Subsequently, Pfizer, with their own lucrative antidepressant drug sertraline (brand name Zoloft) on the market, conducted a clinical trial and concluded that "St. John's wort was not effective for the treatment of major depression" (Shelton et al., 2001). Shelton et al. (2001) criticized the earlier work for "inadequate doses of the antidepressant" and stated the "blind may have been transparent." Shelton et al. (2001) was subsequently criticized for differential patient selection: "patients in the Pfizer-backed [trial] were also seriously depressed. Even the staunchest advocates [of St. John's wort] don't believe it works for serious

depression" (Parker-Pope, 2001).

To test whether anecdotal examples such as these systematically explain the sponsorship effect. First, I test for manipulation of the randomization process by running balance tests on the treatment arms. Secondly, I test whether sponsored arms are more likely to prematurely stop trials. Finally, I consider whether sponsored arms are selected in terms of observable trial characteristics or patient selection, and whether any selection is correlated with predicted drug efficacy.

My clinical trial data contain information on the number of patients enrolled in each arm, the length of the trial in weeks, baseline severity, dropout rate, dosage, mean patient age, and share of female patients. To test whether sponsored arms are different on observable dimensions, I re-estimate equation 2 with observable trial characteristics as the dependent variable. Table 9 presents the results. There are no broad patterns in terms of observable characteristics between sponsored and un-sponsored arms of the same drug in the same drug pair, except age. Sponsored arms enroll slightly older patients, but only this one covariate is statistically significant.

While I find sponsored arms are not different in terms of observable characteristics, these characteristics could still be differentially predictive of efficacy *within* specific drugs. As an example, suppose drug A is more favorable in female patients than drug B. A sponsor of drug A might enroll more women in a clinical trial comparing these two drugs, while a sponsor of drug B might enroll more male patients. The opposite could occur for a different set of drugs, so on average sponsored arms wouldn't enroll more women, though the gender enrollment was still important for explaining the sponsorship effect.

To test whether sponsored arms have higher drug-specific predicted efficacy, I estimate:

$$y_{ij} = \alpha + \beta_{Z_k} Z_k * I_i + X_{ij}\gamma + \varepsilon_{ij} \tag{6}$$

where y_{ij} is the outcome for arm i in trial j, Z_k is the characteristic k (e.g. baseline severity, share female), and X_{ij} controls for the type of measurement scale and the year published as in Section 3.2. This specification aggregates information across all trials and thus does not have drug by drug set or drug by drug pair fixed effects.

I use the estimates from equation 6 to compute $\hat{y}_{ij}$, the predicted efficacy for arm i in trial j for every characteristic. Then, I re-estimate my main regression with predicted efficacy, relative to the predicted efficacy for the placebo or least effective arm, on the left hand side:

$$\hat{y}_{ij} = \alpha + \beta Sponsor_{ij} + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{7}$$

The coefficient on $Sponsor_{ij}$ can be interpreted as "how large would we expect the sponsorship effect to be, simply due to the fact that sponsored arms are more or less likely to enroll characteristic k?" Table 10 shows that sponsored arms do not have higher predicted efficacy in general. The largest coefficient is on the dropout share. Trials with lower dropout rates generally have higher efficacy, and sponsored arms are

more likely to have lower dropout rates. However, even the predicted difference due to the dropout rate is not statistically significant.

Finally, I combine all of these covariates in one prediction, using LASSO to select the most predictive characteristics.[31] As shown in Table 11, sponsored arms are not predicted to have higher relative efficacy based on the full set of observable characteristics. In fact, sponsored arms are predicted to be slightly less effective based on just trial characteristics, and excluding patient characteristics. I conclude that the observable characteristics of trial design and patient enrollment do not explain the sponsorship effect. An important caveat of my analysis is there are many characteristics of trial design not included in these observable characteristics, such as the patient's willingness to adhere to treatment, their underlying health conditions, or the level of monitoring during treatment. These might be notable components of the sponsorship effect.

Differential trial design might be less prevalent in this setting because identifying characteristics that are favorable for particular psychiatric medications is difficult. A meta-analysis of antidepressant clinical trials found that, "no differences in efficacy were seen in patients with accompanying symptoms or in subgroups based on age, sex, ethnicity, or comorbid conditions" (Gartlehner et al., 2011). If trial funders are not able to identify characteristics that determine differences in efficacy, then differential trial design and patient selection is difficult to implement.

4.2 Publication Bias

4.2.1 General Tests for Publication Bias

The results of clinical trials often affects their publication outcomes. In antidepressant clinical trials submitted to the FDA, thirty-six out of thirty-seven trials viewed as having positive results by the FDA were published, while only fifteen out of thirty-six trials viewed as negative were published (Turner et al., 2008). My contribution is to test whether this is publication bias differs for trials by funding source.

First, I test whether sponsored arms are differentially likely to be published based on their reported efficacy, compared to unsponsored arms of the same drug in the same drug pair. I estimate:

$$1\{Published_j\} = \alpha + Sponsor_{ij} + Sponsor_{ij} * y_{ij} + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{8}$$

where the outcome is an indicator for whether trial j was published. I interact sponsorship with y_{ij}, the standardized efficacy of a given arm i in trial j. The rest of the terms are the same as in equation 2, and the controls are as in my preferred specification in Table 5, column (2b). As shown in Table 12, while sponsored arms are statistically significantly less likely to be published, sponsored arms with higher efficacy are *more* likely to be published.[32]

[31] LASSO refers to the least absolute shrinkage and selection operator; see Tibshirani (1996).

[32] Another standard test for publication bias is to measure the level of bunching around z-score cutoffs (Brodeur et al., 2016). Appendix Figure B4 plots the z-score distribution for published trials. There is weak evidence of bunching at the 5% and 10% cutoffs. However, this bunching occurs for both sponsored and unsponsored arms and is underpowered.

I also assess whether the sponsorship effect is smaller among trials linked to ClinicalTrials.gov. As described in Section 2.3.4, a subset of the trials have a NCT number and are linked to their record on Clinicaltrials.gov. As shown in Figure 6b, these trials were disproportionately published in later years. To assess whether pre-registered trials have lower sponsorship effects, I estimate the following specification:

$$y_{ij} = \alpha + Sponsor_{ij} + Sponsor_{ij} * Link_j + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{9}$$

where $Link_j$ is an indicator for whether trial j was pre-registered in ClinicalTrials.gov. As shown in Table 13, among this linked sample, there is no sponsorship effect. The first column presents my baseline sponsorship effect, while the second column presents the interacted coefficients on sponsorship. I find that the sponsorship effect among trials pre-registered on ClinicalTrials.gov is -0.094 standard deviations, relative to sponsored but not pre-registered trials. Relative to the baseline sponsorship effect of 0.135, the average sponsorship effect among linked trials is a statistically insignificant 0.041 standard deviations.

4.3 Decomposition of Sponsorship Effect by Mechanisms

The unpublished trials in my data are a subset of the universe of all unpublished trials ever conducted. This latter set of trials have never been publicly available. With an approximation of how many clinical trials were conducted but never available, I can estimate the total share of the sponsorship effect that is explained by publication bias. I approximate the set of additional unpublished trials with the full set of relevant trials in ClinicalTrials.gov during recent years, when the pre-registration requirements were enforced.

Within the ClinicalTrials.gov data, I restrict the sample to trials for "major depressive disorder" that tested antidepressant drugs included in my sample; see Section 2.2 for a full list. I require the trials to be initially registered between 2005-2010, which allows six years for the results to be available, either in published or obtainable unpublished form. Out of the 147 pre-registered trials that fit this criteria, my clinical trial data contains results for just 22% of these trials. Therefore, I estimate that there exist five times more unpublished trials in the universe of all clinical trials than I observe in my analysis sample.

This estimate relies on two strong assumptions: that the observed unpublished trials are a random sample of all initiated trials and that the clinical trial registry contains the full universe of trials conducted. To the extent the observed unpublished trials are more favorable to funders than the unobserved unpublished trials, or that the clinical trials registry undercounts trials, I would underestimate the true share of the sponsorship effect explained by publication bias.

Figure 7 presents the share of the sponsorship effect that is explained under alternate assumptions about the magnitude of publication bias. With only published trials, my estimate of the sponsorship effect is 0.155; in my analysis sample of both published and unpublished trials the sponsorship effect is 0.124. Since I estimate that I observe only 22% of all pre-registered trials in my sample, the "Add 5x" bar is my preferred estimate for the size of the sponsorship effect with the full universe trials. With five times more unpublished trials, the sponsorship effect would fall to 0.073 and would only be significant at the

10% level. This estimate represents a decrease of 41% from the baseline sponsorship effect of 0.124 and a decrease of 53% from the effect with only published trials of 0.155. Therefore, I estimate that 40-50% of the sponsorship effect can be explained by publication bias. As shown in the last bar of Figure 7, with our assumptions on the selection of unpublished trials, in order for the sponsorship effect to be fully explained by publication bias, the universe of unpublished trials would need to be approximately thirty times larger than the currently observed set of unpublished trials.

4.4 Mitigation and Pre-Registration

As a second test, if this sponsorship effect is due to publication bias, then pre-registration requirements might mitigate these effects. As of July 1, 2005, the ICMJE agreed to only publish clinical trials in affiliated journals that had been registered before patient enrollment. Trials that had begun enrollment before 2005 were required to pre-register before submission to ICMJE journals. To test whether pre-registration changed the sponsorship effect, I estimate the following specification:

$$y_{ij} = \alpha + Sponsor_{ij} + \sum_{y} \beta_y Sponsor_{ij} * y(j) + \sum_{y} y(j) + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{10}$$

where the sponsorship effect is interacted with publication year bins $y(j)$. Here, X_{ij} includes only controls for the measurement scale. Figure 6a plots the coefficients β_y. The introduction of pre-registration requirements is marked by the vertical dashed line. The pre-registration requirement affected trials according to the trial enrollment date. Therefore the treatment intensity, as measured by the share of published trials pre-registered on ClinicalTrials.gov, increases gradually over time (Figure 6b). Similarly, the sponsorship effect in Figure 6a decreases in magnitude gradually after the 2005 pre-registration requirements. The effect of sponsorship on reported drug efficacy is statistically significantly lower after required pre-registration than before required pre-registration. However, this change is statistically indistinguishable from a linear decline in the sponsorship effect over time.

5 Conclusion

This paper provides empirical evidence that financial incentives affect the results of clinical trials. I find that a sponsored drug appears substantially more effective than that same drug in the same drug set but without the drug manufacturer's involvement. Across a variety of specifications and outcomes, this effect is large and consistently represents approximately a third of the average difference in efficacy between trial arms. Publication bias can conservatively explain about half of this effect, while I find no evidence that trial design or patient enrollment play a large role. The share of the sponsorship effect explained by publication bias could be larger than I estimate due to either a lack of compliance with pre-registration requirements or selection of the observed unpublished trials. The remaining unexplained share of the sponsorship effect

may also be due to characteristics of trial design that are unobservable in my clinical trial data, or data falsification.

The magnitude of the effect of funding on drug efficacy is large enough to have substantial implications for drug approvals and prescriptions. In terms of drug approvals, my sample includes 28 drugs approved by the FDA and six drugs that were not approved. The average relative efficacy of a given drug in pre-approval trials is strongly predictive of gaining approval from the FDA. If this relationship were causal and if drug efficacy decreased by the average sponsorship effect of 0.135 standard deviations, then the approval rate would fall from 80% to 69%. This decrease would correspond to three fewer approved psychiatric drugs. In terms of prescriptions, if the relationship between a drug's effectiveness and prescriptions in Figure 3 were causal, then reducing the sponsorship effect to zero would result in 0.77 million fewer Medicaid prescriptions per drug, or a 32% decrease.

While psychiatric medications are a large and economically substantial drug class, there are various reasons why financial incentives might be more or less relevant in this setting. Sponsorship could be less salient for psychiatric medications because of the difficulty in predicting treatment responses to particular drugs. On the other hand, efficacy for these medications is measured on a subjective scale, which provides more leeway than laboratory tests. Future work could examine alternate drug classes. Classes which also have numerous substitutable drugs and variation in sponsorship could be viable candidates.[33] Another important distinction is that my paper intentionally focuses on a consistent set of outcomes to measure drug efficacy. Thereby, I address how financial incentives affect reported efficacy itself, rather than the choice of which efficacy measure to report. However, outcome selection is a key component of clinical trial design and is potentially also affected by financial incentives.

My results are agnostic about the welfare consequences of different funding sources for clinical trials. Whether it would be socially beneficial for pharmaceutical research to be conducted by parties with more limited financial stakes in the results depends on several factors. Potentially these restrictions would limit the total amount of innovative research. Alternate funding schemes should also consider additional factors such as how sponsored clinical research is interpreted by physicians and patients, the availability of subsequent publications, and the external validity of clinical research. The evidence in this paper informs this debate by documenting that the funding source of a clinical trial affects the reported drug efficacy and that publication bias is an important mechanism.

[33] Potential candidates include anti-inflammatory drugs for osteoarthritis and stimulants for attention deficit hyperactivity disorder.

References

Åberg-Wistedt, Anna, Hans Ågren, Lisa Ekselius, Finn Bengtson, and Ann-Charlotte Åkerblad, "Sertraline Versus Paroxetine in Major Depression: Clinical Outcome After Six Months of Continuous Therapy," *Journal of Clinical Psychopharmacology*, 2000, *20* (6), 645–652.

Abrams, Eliot, Jonathan Libgober, and John A. List, "Research Registries: Facts, Myths, and Possible Improvements," *National Bureau of Economic Reserach Working Paper 27250*, 2020.

Adams, Christopher and Van V. Brantner, "Estimating the Cost of New Drug Development - Is it Really $802 Million?," *Health Affairs*, 2006, *25* (2), 420–428.

Allcott, Hunt, "Site Selection Bias in Program Evaluation," *The Quarterly Journal of Economics*, 2015, *130* (3), 1117–1165.

Amsterdam, JD, WG Case, E Csanalosi, M Singer, and K Rickels, "A Double-Blind Comparative Trial of Zimelidine, Amitriptyline, and Placebo in Patients with Mixed Anxiety and Depression," *Pharmacopsychiatry*, 1986, *19* (3), 115–119.

Avorn, Jerry, "The $2.6 Billion Pill - Methodologic and Policy Considerations," *New England Journal of Medicine*, 2015, *372*, 1877–1879.

Azoulay, Pierre, "Do Pharmaceutical Sales Respond to Scientific Evidence?," *Journal of Economics & Management Strategy*, 2004, *11* (4), 551–594.

Bekelman, Justin E., Yan Li, and Cary P. Gross, "Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review," *Journal of the American Medical Association*, 2003, *289* (4), 454–465.

Bourgeois, Florence, Srinivas Murthy, and Kenneth Mandl, "Outcome Reporting Among Drug Trials Registered in ClinicalTrials.gov," *Annals of Internal Medicine*, 2010, *153* (3), 158–166.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg, "Star Wars: The Empirics Strike Back," 2016, *8* (1), 1–32.

Brody, Debra P., Laura A. Pratt, and Jeffery P. Hughes, "Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013 to 2016, NCHS Data Brief No. 303," 2018.

Budish, Eric, Benjamin N. Roin, and Heidi Williams, "Do Firms Underinvest in Long-Term Research? Evidence from Cancer Clinical Trials," *American Economic Review*, 2015, *105* (7), 2044–2085.

Carroll, Aaron, "Do Antidepressants Work?," *The New York Times*, 2018.

Cipriani, Andrea, Toshi A Furukawa, Georgia Salanti, Anna Chaimani, Lauren Z. Atkinson, Yusuke Ogawa, Stefan Leucht, Henricus G Ruhe, Erick H Turner, Julian P T Higgins, Matthias Egger, Nozomi, Takeshima, Yu Hayasaka, Hissei Imai, Kiyomi Shinohara, Aran Tajika, John P A Ioannidis, and John R Geddes, "Comparative Efficacy and Acceptability of 21 Antidepressant Drugs for the Acute Treatment of Adults with Major Depressive Disorder: A Systematic Review and Network Meta-Analysis," *The Lancet*, 2018, *391*, 1357–1366.

Danzon, Patricia and Eric Keuffel, "Regulation of the Pharmaceutical-Biotechnology Industry," in Nancy Rose, ed., *Economic Regulation and Its Reform: What Have We Learned?*, 2014, pp. 407–484.

Davidoff, Frank, Catherine D. DeAngelis, Jeffrey M. Drazen, John Hoey, Liselotte Højgaard, Richard Horton, Sheldon Kotzin, M. Gary Nicholls, Magne Nylenna, A. John P.M. Overbeke, Harold C. Sox, and Martin B. Van Der Weyden, "Sponsorship, Authorship, and Accountability," 2001, *345*, 825–827.

DiMasi, Joseph, Henry Grabowski, and Ronald Hansen, "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs," *Journal of Health Economics*, 2016, *47*, 20–33.

Ehrhardt, Stephan, Lawrence Appel, and Curtis Meinert, "Trends in National Institutes of Health Funding for Clinical Trials Registered in ClinicalTrials.gov," *Journal of the American Medical Association*, 2015, *314* (23), 2566–2567.

Friedberg, Mark, Bernard Saffran, Tammy J. Stinson, Wendy Nelson, and Charles L. Bennett, "Evaluation of Conflict of Interest in Economic Analysis of New Drugs Used in Oncology," *Journal of the American Medical Association*, 1999, *282* (15), 1453–1457.

Gartlehner, Gerald, Richard A. Hansen, Laura C. Morgan, Kylie Thaler, Linda Lux, Megan Van Noord, Ursula Mager, Patricia Thieda, Bradley N. Gaynes, Tania Wilkins, Michaela Strobelberger, Stacey Lloyd, Ursula Reichenpfader, and Kathleen N. Lohr, "Comparative Benefits and Harms of Second-Generation Antidepressants for Treating Major Depressive Disorder: An Updated Meta-Analysis," *Annals of Internal Medicine*, 2011, *155* (11), 772–785.

Greenberg, Paul E., Andree-Anne Fournier, Tammy Sisitsky, Crystal T. Pike, and Ronald C. Kessler, "The Economic Burden of Adults with Major Depressive Disorder in the United States (2005 and 2010)," *Journal of Clinical Psychiatry*, 2015, *76* (2), 155–162.

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

Ioannidis, John PA, "Effectiveness of Antidepressants: An Evidence Myth Constructed from a Thousand Randomized Trials?," *Philosophy, Ethics, and Humanities in Medicine*, 2008, *3* (14).

Kirsch, Irving, "The Emperor's New Drugs: Exploding the Antidepressant Myth," 2010.

Krauss, Alexander, "Why All Randomised Controlled Trials Produce Biased Results," *Annals of Medicine*, 2018, *50* (4), 312–322.

Lathyris, D. N., N. A. Patsopoulos, G. Salanti, and J. P. A. Ioannidis, "Industry Sponsorship and Selection of Comparators in Randomized Clinical Trials," *European Journal of Clinical Investigation*, 2010, *40* (2), 172–182.

Leucht, Stefan, Andrea Cipriani, Loukia Spineli, Dimitris Mavridis, Deniz Örey, Franziska Richter, Myto Samara, Corrado Barbui, Rolf R Engel, John R Geddes, Werner Kissling, Marko P Stapf, Bettina Lässig, Georgia Salanti, and John M Davis, "Comparative Efficacy and Tolerability of 15 Antipsychotic Drugs in Schizophrenia: A Multiple-Treatments Meta-Analysis," *The Lancet*, 2013, *382* (9896), 951–962.

Lexchin, Joel, Lisa A. Bero, Benjamin Djulbegovic, and Otavio Clark, "Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systemic Review," *BMJ*, 2003, *326*, 1167–1170.

Linde, Klaus, Gilbert Ramirez, Cynthia D Mulrow, Andrej Pauls, Wolfgang Weidenhammer, and Dieter Melchart, "St John's Wort for Depression: An Overview and Meta-Analysis of Randomised Clinical Trials," *BMJ*, 1996, *313*, 253–258.

Lundh, Andreas, Joel Lexchin, Barbara Mintzes, Jeppe B Schroll, and Lisa Bero, "Industry Sponsorship and Research Outcomes," *Cochrane Datatabase of Systemic Reviews*, 2017, (MR000033).

Lydiard, R. Bruce, Stephan M. Stahl, Marc Hertzman, and Wilma M. Harrision, "A Double-Blind, Placebo-Controlled Study Comparing the Effects of Sertraline Versus Amitriptyline in the Treatment of Major Depression," *Journal of Clinical Psychiatry*, 1997, *58* (11), 484–491.

Malani, Anup, "Identifying Placebo Effects with Data from Clinical Trials," *Journal of Political Economy*, 2006, *114* (2), 236–256.

Mathieu, Sylvain, Isabelle Boutron, David Moher, Douglas G. Altman, and Philippe Ravaud, "Comparison of Registered and Published Primary Outcomes in Randomized Controlled Trials," *Journal of the American Medical Association*, 2009, *302* (9), 977–984.

McKibbin, Rebecca, "The Effect of RCTs on Demand for Off-Label Cancer Drugs," *Working Paper*, 2020.

Mikulic, Matej, "Revenue of the worldwide pharmaceutical market from 2001 to 2019," 2020. https://www.statista.com/statistics/263102/pharmaceutical-market-worldwide-revenue-since-2001/.

Moore, Thomas and Donald Mattison, "Adult Utilization of Psychiatric Drugs and Differences by Sex, Age, and Race," *JAMA Internal Medicine*, 2017, *177* (2), 274–275.

Moore, Thomas J., Hanzhe Zhang, Gerard Anderson, and G. Caleb Alexander, "Estimated Costs of Pivotal Trials for Novel Therapeutic Agents Approved by the US Food and Drug Administration, 2015-2016," *JAMA Internal Medicine*, 2018, *1787*, 1451–1457.

Naudet, Florian, Anne Solène Maria, and Bruno Falissard, "Antidepressant Response in Major Depressive Disorder: A Meta-Regression Comparison of Randomized Controlled Trials and Observational Studies," 2011, *6* (6).

Østengaard, Lasse, Andreas Lundh, Tine Tjørnhøj-Thomsen and Suhayb Abdi, Mustafe H A Gelle, Lesley A Stewart, Isabelle Boutron, and Asbjørn Hróbjartsson, "Influence and management of conflicts of interest in randomised clincial trials: qualitative interview study," *BMJ*, 2020, *371* (m3764).

Parker-Pope, Tara, "Study That Discredits SJW Draws Dubious Conclusions," *The Wall Street Journal*, 2001.

Perlis, Roy H., Clifford S. Perlis, Yelena Wu, Cindy Hwang, Megan Joseph, and Andrew Nierenberg, "Industry Sponsorship and Financial Conflict of Interest in the Reporting of Clinical Trials in Psychiatry," *American Journal of Psychiatry*, 2005, *162* (10), 1957–1960.

Prasad, Vinay and Sham Mailankody, "Research and Development Spending to Bring a Single Cancer Drug to Market and Revenues After Approval," *JAMA Internal Medicine*, 2017, *177* (11), 1569–1575.

Pratt, Laura A., Debra J. Brody, and Qiuping Gu, "Antidepressant Use Among Persons Aged 12 and Over: United States, 2011 - 14, NCHS Data Brief No. 283," 2017.

Psaty, Bruce M., Noel S. Weiss, and Curt D. Furberg, "Recent Trials in Hypertension: Compelling Science or Commercial Speech?," *Journal of the American Medical Association*, 2006, *295* (14), 1704–1706.

Rettig, Richard A., "The Industrialization of Clinical Research," *Health Affairs*, 2000, *19* (2).

Schulz, Kenneth F., Iain Chalmers, Richard J. Hayes, and Douglas G. Altman, "Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials," *Journal of the American Medical Association*, 1995, *273* (5), 408–412.

Sertkaya, Aylin, Hui-Hsing Wong, Amber Jessup, and Trinidad Beleche, "Key Cost Drivers of Pharmaceutical Clinical Trials in the United States," *Clinical Trials*, 2016, *13* (2), 117–126.

Shelton, Richard C., Martin B. Keller, Alan Gelenberg, David L. Dunner, Robert Hirschfeld, Michael E. Thase, James Russell, R. Bruce Lydiard, Paul Crits-Christoph, Robert Gallop, Linda Todd, David Hellerstein, Paul Goodnick, Gabor Keitner, Stephen M. Stahl, and Uriel Halbreich, "Effectiveness of St John's Wort in Major Depression: A Randomized Controlled Trial," *Journal of the American Medical Association*, 2001, *285* (15), 1978–1986.

Tay-Teo, Kiu, André Ilbawi, and Suzanne R. Hill, "Comparison of Sales Income and Research and Development Costs for FDA-Approved Cancer Drugs Sold by Originator Drug Companies," *JAMA Network Open*, 2019, *2* (1).

Taylor, David, Anna Sparshatt, Seema Varma, and Olubanke Olofinjana, "Antidepressant Efficacy of Agomelatine: Meta-Analysis of Published and Unpublished Studies," *BMJ*, 2014, *348*, 1888.

Tibshirani, Robert, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, 1996, *58* (1), 267–288.

Turner, Erick H., Annette M. Matthews, Eftihia Linardatos, Robert A. Tell, and Robert Rosenthal, "Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy," *New England Journal of Medicine*, 2008, *358*, 252–260.

US Food and Drug Administration, Center for Drug Evaluation and Research, "Guidance for Industry: Guidelines for the Clinical Evaluation of Antidepressant Drugs.," 1977.

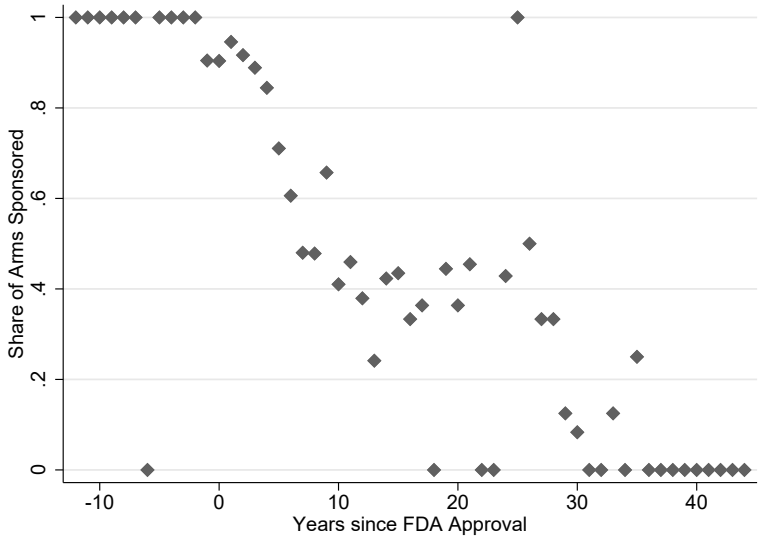
Wu, Eric Q., Lizheng Shi, Howard Birnbaum, Teresa Hudson, and Ronald Kessler, "Annual Prevalence of Diagnosed Schizophrenia in the USA: A Claims Data Analysis Approach," 2006, *36* (11), 1535–1540.

Yu, Nancy L., Preston Atteberry, and Peter B. Bach, "Spending On Prescription Drugs In The US: Where Does All The Money Go?," *Health Affairs Blog*, 2018.

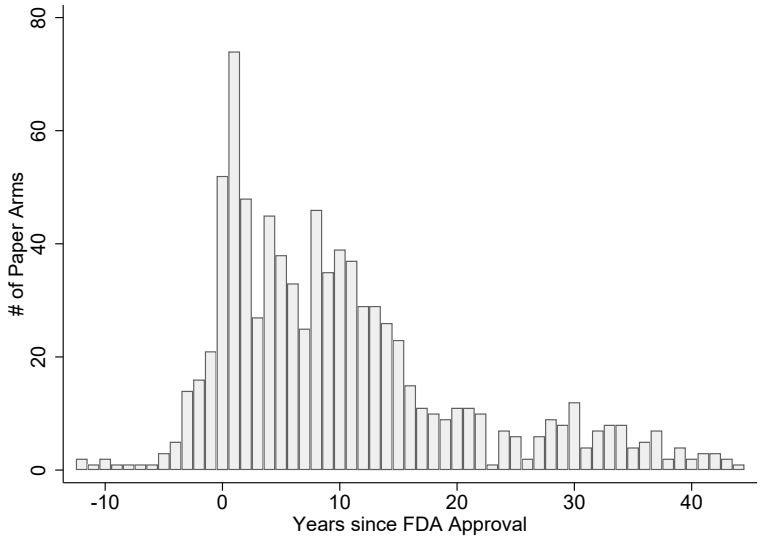
Zinner, Darren E., Dragana Bolcic-Jankovic, Brian Clarridge, David Blumenthal, and Eric G. Campbell, "Participation of Academic Scientists in Relationships with Industry," *Health Affairs*, 2013, *28* (6), 1814–1825.

Figure 1: Variation in Sponsorship Over Time

(a) Sponsorship by Relative Year

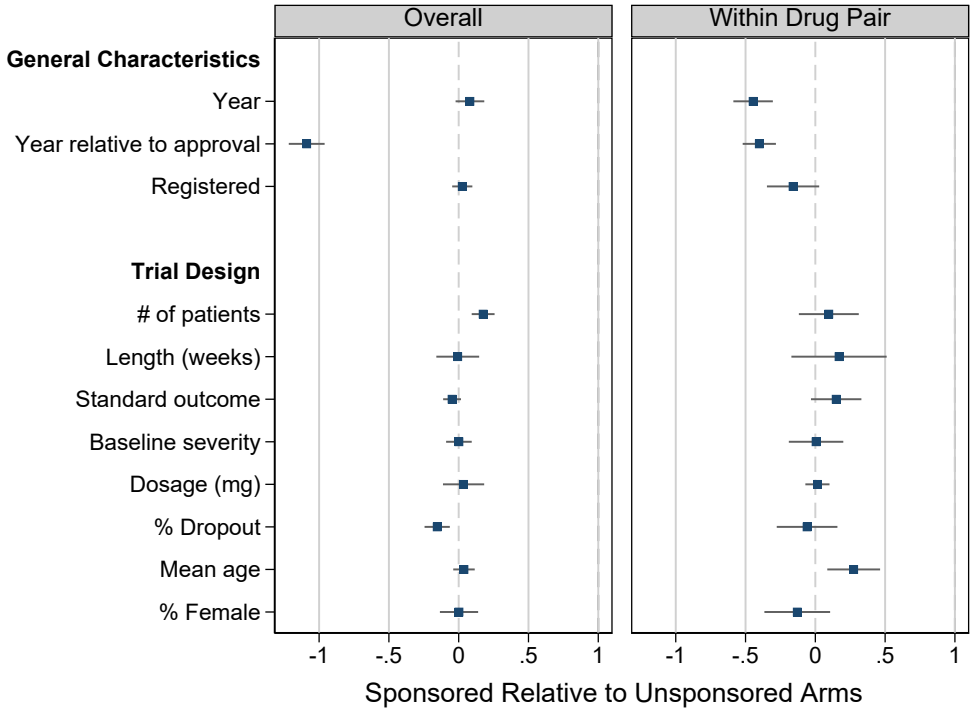


(b) Sample Size



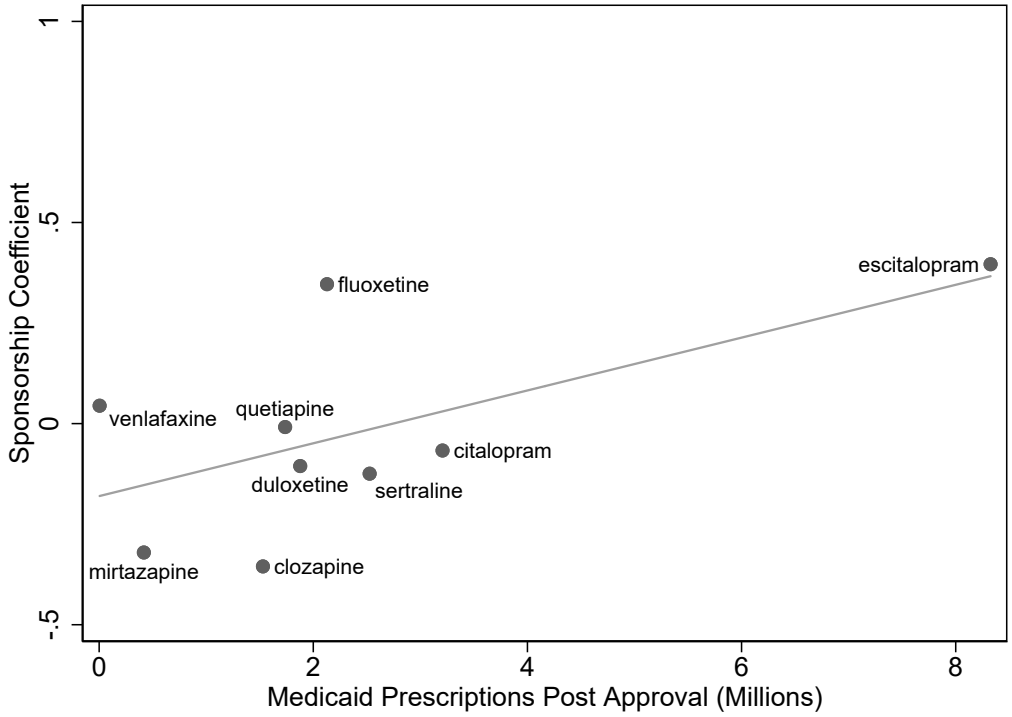
Notes: This figure (Panel A) presents the average share of sponsored arms over time. The x-axis plots the number of years since FDA approval for a given drug. The y-axis plots the share of those arms that are sponsored. This figure excludes placebo arms and drugs that are not approved by the FDA (agomelatine, amisulpride, milnacipran, reboxetine, sertindole, and zotepine). Panel B presents the number of trial arms in my sample by the number of years since FDA approval.

Figure 2: Characteristics of Sponsored Relative to Unsponsored Arms



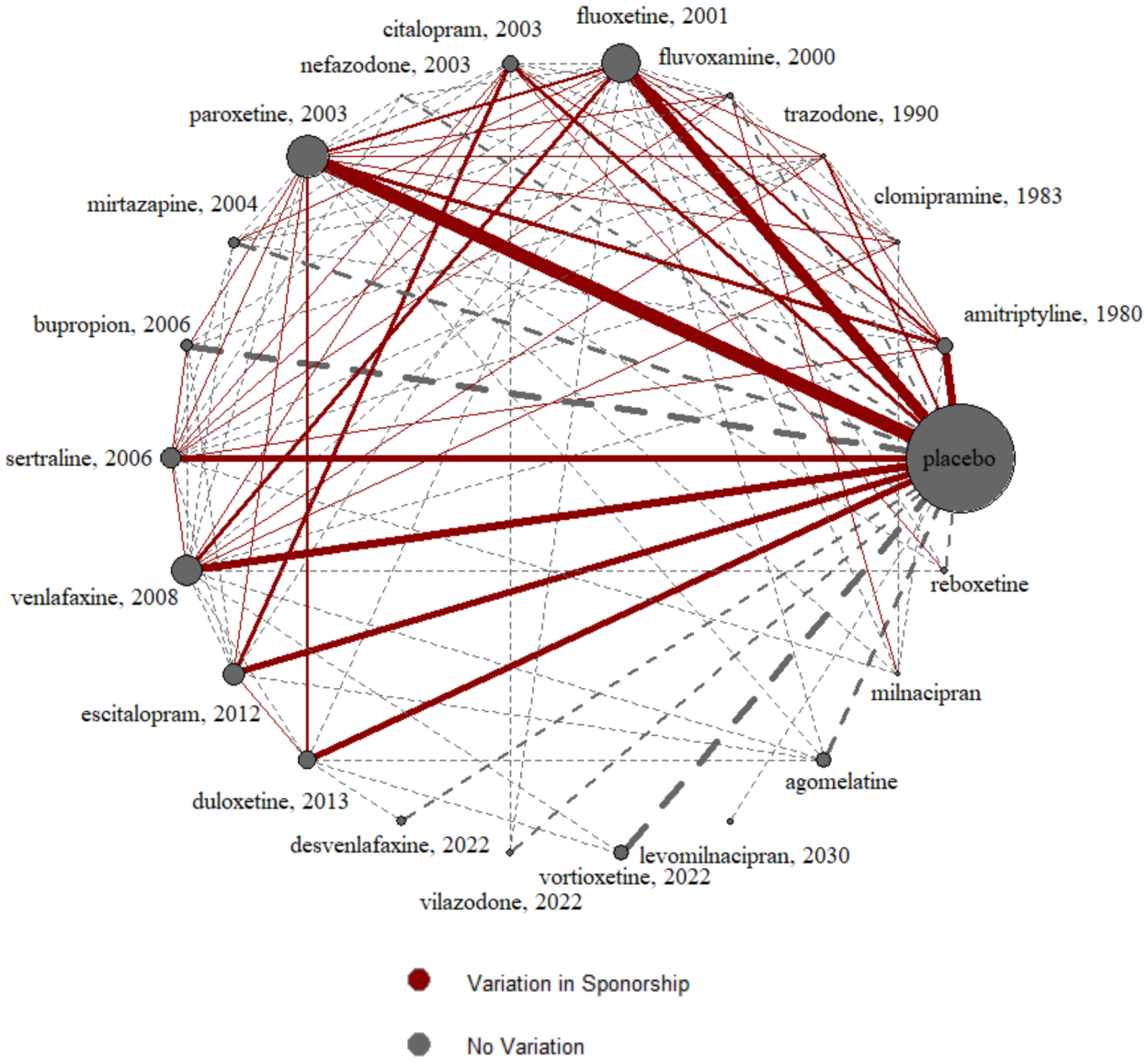
Notes: This figure presents the difference in characteristics and trial design outcomes for sponsored relative to unsponsored arms. The left panel presents the overall difference in trial characteristics between all sponsored and unsponsored arms. The right panel presents the difference in trial characteristics controlling for drug pairs. These differences were calculated using regression coefficients from the estimation of equation 3 and 4 as described in Section 3.4.2. Error bars represent 95 percent confidence intervals. Standard errors are clustered at the trial level. Details for each outcome are listed in Section 2.3.4.

Figure 3: Sponsorship Effect and Drug Sales



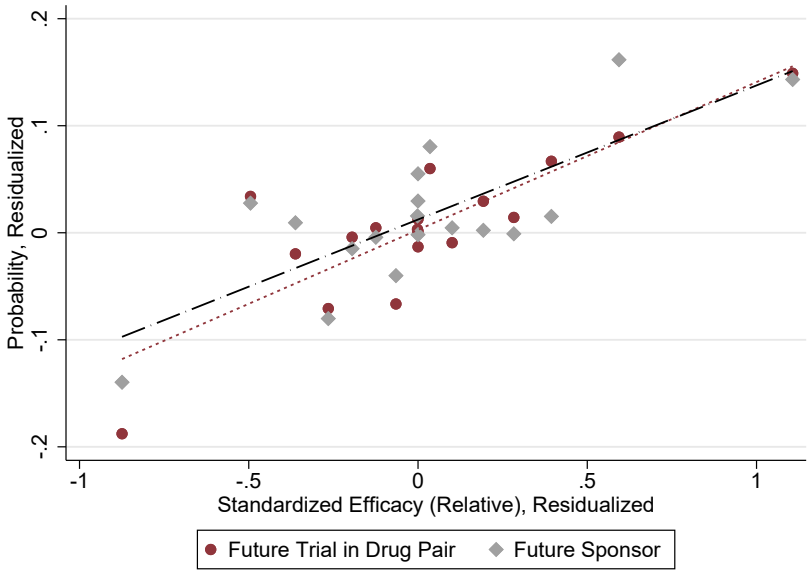
Notes: This figure plots the coefficient on sponsorship for each drug from the estimation of equation 5 against the total number of Medicaid prescriptions in the five years post-approval for that drug. The best-fit line is plotted in gray.

Figure 4: Network of Trials for Antidepressants



Notes: Figure presents the network of comparisons within antidepressants. Each node represents a drug and is labeled with the year that a generic formulation entered the United States market (years after 2019 are estimates). The size of the circle is proportional to the number of randomly assigned participants. Each line represents a clinical trial comparing the two drugs. A trial with three or more drugs would have a line between every pair of drugs tested. The width of the lines is proportional to the number of trials comparing every pair of treatments. Lines in solid red denote that the sponsorship status of at least one of the drugs varies within the trials; lines in dashed gray denote that the sponsorship status of both drugs is constant.

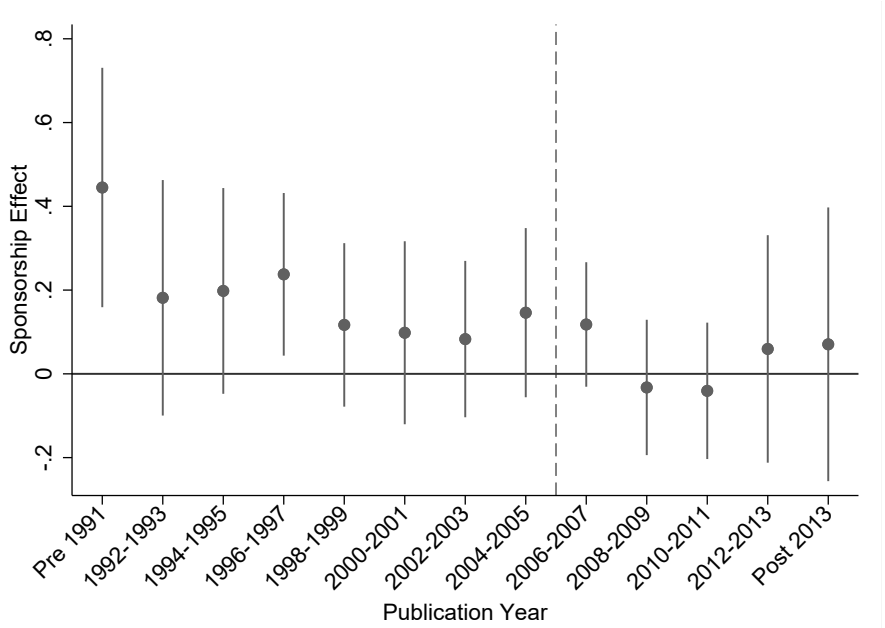
Figure 5: Predictors of Future Papers and Sponsorship



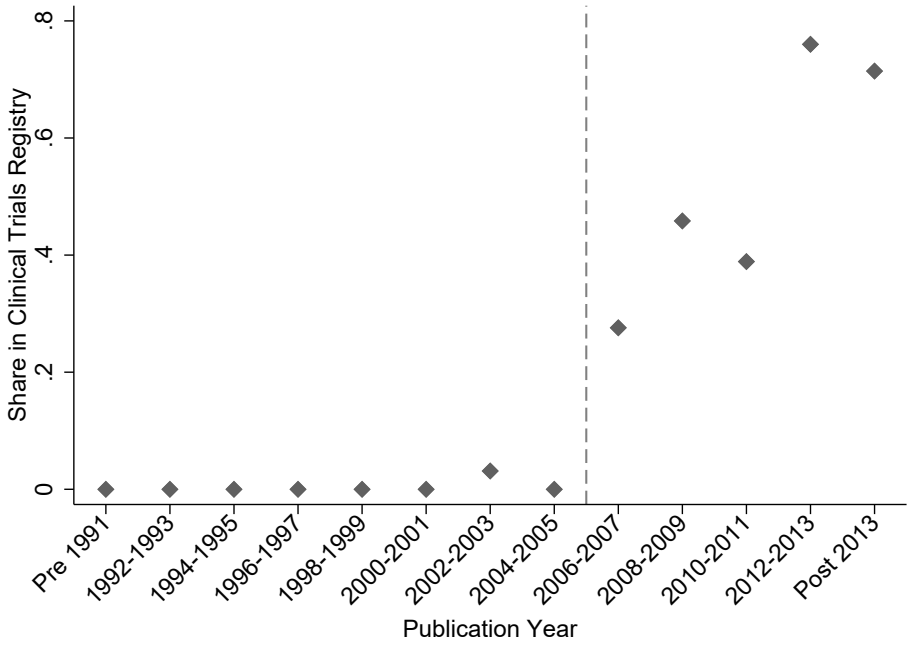
Notes: This figure presents the relationship between effectiveness and future papers and sponsorship. "Future Trial in Drug Pair" is an indicator for whether there is a future trial assessing that same drug in the same drug pair. "Future Sponsor" is an indicator for whether that drug is ever sponsored in any future trials assessing the same drug in that same drug pair. Future refers to any trial published in a year after the original publication year; concurrent publications are excluded. The x-axis plots the standardized efficacy of the original arm, controlling for each separate drug in each drug pair. The y-axis presents the probability that there is either a future trial or a future sponsorship, again controlling for each separate drug in each separate drug pair.

Figure 6: Introduction of Clinical Trial Pre-registration

(a) Sponsorship Effect by Calendar Year

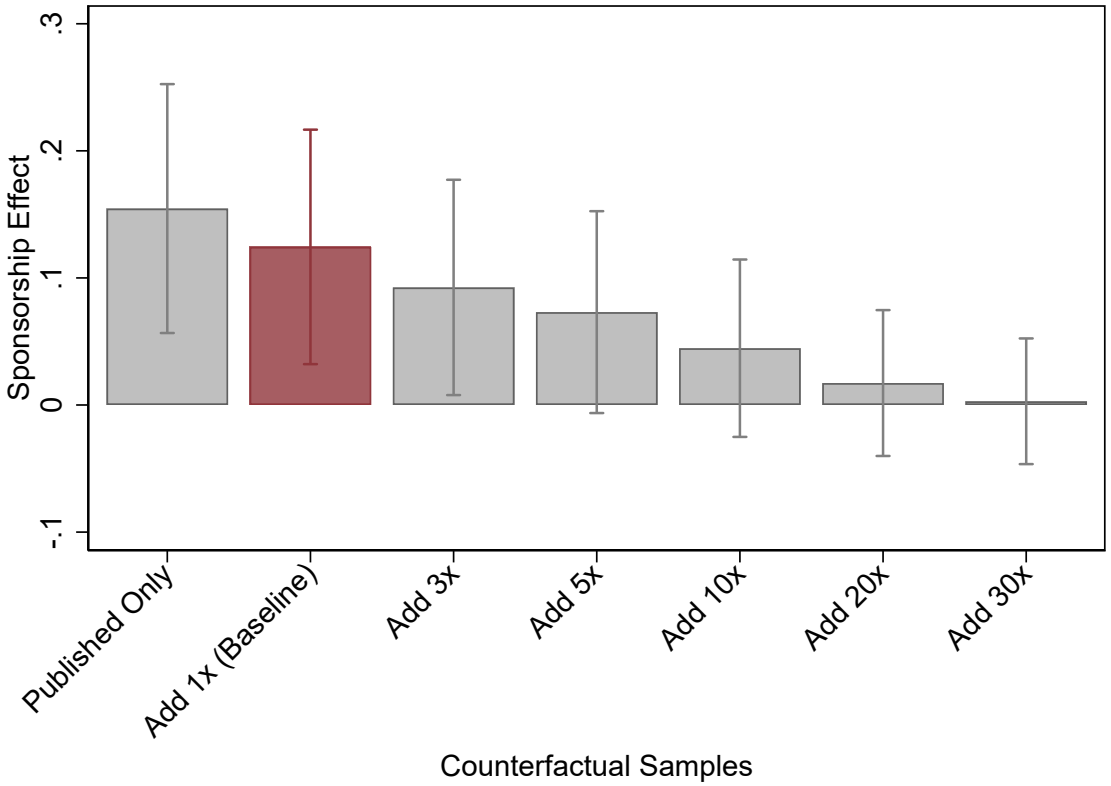


(b) Treatment Intensity by Calendar Year



Notes: Panel A presents the coefficients β_y from the estimation of equation 10. The vertical dashed line midway between 2005 and 2006 represents July 1st, 2005, when the International Committee on Medical Editors agreed to only publish clinical trials that had been registered before patient enrollment. Standard errors are clustered at the trial level. Panel B plots the share of antidepressant trials in my sample that were pre-registered in ClinicialTrials.gov by publication year.

Figure 7: Counterfactual Sponsorship Effect under Alternate Publication Assumptions



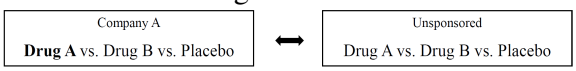
Notes: This figure presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 2 with alternate samples. The left-most bar presents estimates including only published trials. The second bar in solid red presents estimates including each unpublished trial one time. This is my baseline sample and this estimate replicates Table 5, column (7). Subsequent columns include additional unpublished trials as described in Section 4.3. 95% confidences intervals are presented as lines on each bar graph. Standard errors are clustered at the trial level.

Table 1: Summary Statistics: Full and Variation Samples

	Full Sample				Sample with Variation within: Drug Sets				Drug Pairs			
	Mean	Median	Std Dev.	% Missing	Mean	Median	Std Dev.	% Missing	Mean	Median	Std Dev.	% Missing
Year	2001	2001	9	14	1999	1999	8	8	2000	2000	8	22
Year relative to approval	10.1	8.0	10.6	39	11.4	8.0	11.4	24	11.8	9.0	10.8	41
Share:												
Sponsored	0.49	0.00	0.50	0	0.52	1.00	0.50	0	0.42	0.00	0.49	12
Sponsored w/o	0.43	0.00	0.50	0	0.41	0.00	0.49	0	0.34	0.00	0.47	12
COI												
Antidepressant	0.77	1.00	0.42	0	0.81	1.00	0.39	0	0.81	1.00	0.39	12
Registered	0.13	0.00	0.33	0	0.05	0.00	0.22	0	0.09	0.00	0.29	12
Trial design:												
# of patients	100	89	84	0	88	70	98	0	94	78	90	12
Length (weeks)	8.7	8.0	7.5	0	8.5	6.0	6.4	0	9.1	8.0	8.1	12
Standard outcome	0.90	1.00	0.30	0	0.90	1.00	0.31	0	0.90	1.00	0.30	12
Baseline severity	-0.0	-0.0	1.0	7	0.0	-0.0	1.0	6	-0.1	-0.1	1.0	19
Dosage (mg)	68	25	102	26	59	20	91	17	58	20	85	35
% Dropout	29	27	15	10	29	27	15	11	30	28	15	22
Mean age	42	41	9	18	44	41	11	17	43	41	9	28
% Female	52	58	20	44	52	55	20	54	52	57	20	56
Total arms	1,412				499				900			
Total trials	586				230				400			

Notes: This table presents summary statistics at the trial arm level. Summary statistics are shown for the full sample, the subsample with variation in sponsorship within drug sets, and the subsample with variation in sponsorship within drug pairs. Details for each outcome are listed in Section 2.3.4.

Table 2: Types of Variation

	Variation within:					
	Drug Set			Drug Pair		
	Drug Sets	Trials	Arms	Drug Pairs	Trials	Arms
Active vs. Placebo						
	5	55	122	12	211	495
Active vs. Active						
One Drug Never Sponsored						
	10	42	87	10	38	81
One Drug Always Sponsored						
	16	76	153	18	95	199
Both Drugs Vary in Sponsorship						
	9	44	92	20	96	210
Subtotal	35	162	332	48	226	486
Three or More Drugs						
	3	13	45	0	0	0
Total	43	230	499	60	400	900

Notes: This table presents the different categories of variation in funding. The first set of columns reports estimates for the sample with variation in funding within a drug set, which is a unique combination of drugs within a trial. The second set of columns reports estimates for the sample with variation in funding within a drug pair. The columns report the number of drug sets or drug pairs, the number of trials, which are either a published or an unpublished RCT, and the number of treatment arms, which is a unique randomization arm of a trial. See section 2.3.1 for more details. The rows describe different types of variation. The boxes represent examples of trials for each type. In each box, the first line refers to the funding source. Sponsored arms are in bold. Unsponsored arms are unbolded. Trials are only directly compared to the analogous trials in the same row.

Table 3: Difference in Difference: Active versus Placebo Antidepressants

	Sponsored				Not Sponsored				
	Share Respond				Share Respond				
	Drug	Placebo	*Diff*	# Trials	Drug	Placebo	*Diff*	# Trials	DD
All Drug Sets	0.491	0.303	0.188	59	0.441	0.301	0.140	8	0.048
Paroxetine	0.469	0.320	0.149	32	0.250	0.226	0.024	1	0.126
Sertraline	0.453	0.360	0.093	12	0.476	0.433	0.042	2	0.051
Citalopram	0.513	0.399	0.114	8	0.303	0.209	0.095	1	0.019
Trazodone	0.458	0.158	0.300	6	0.568	0.353	0.215	1	0.085
Amitriptyline	0.564	0.278	0.286	1	0.607	0.282	0.325	3	-0.039

Notes: This table presents the difference-in-difference estimate of the sponsorship effect for "Active vs. Placebo" drug sets. The first set of columns compares the share of patients that respond to treatment when the drug is sponsored; the next set compare these results when the drug is not sponsored. The difference between the share of patients that respond to a given drug and the share that respond to the placebo group is given in the column labeled "Diff" for "Difference." The last column reports the difference between the two difference columns. This difference in difference (DD) is analogous to the sponsorship effect in equation 1.

Table 4: Difference in Difference: Active versus Active Antidepressants

	Sponsored				Not Sponsored				
	Share Respond				Share Respond				
	Drug	Other Arm	*Diff*	# Trials	Drug	Other Arm	*Diff*	# Trials	DD
All Drug Sets	0.640	0.595	0.045	56	0.560	0.578	-0.019	73	0.064
Amitriptyline vs. Paroxetine	0.658	0.648	0.010	1	0.465	0.474	-0.008	14	0.018
Amitriptyline vs. Fluoxetine	0.653	0.564	0.088	3	0.500	0.522	-0.022	10	0.111
Fluoxetine vs. Venlafaxine	0.764	0.745	0.018	1	0.587	0.636	-0.049	10	0.067
Venlafaxine vs. Fluoxetine	0.636	0.587	0.049	10	0.704	0.707	-0.003	1	0.052
Citalopram vs. Escitalopram	0.794	0.815	-0.021	6	0.639	0.760	-0.120	3	0.099
Paroxetine vs. Fluoxetine	0.525	0.473	0.052	7	0.683	0.565	0.119	1	-0.067
Clomipramine vs. Paroxetine	0.535	0.371	0.164	1	0.566	0.657	-0.091	5	0.255
Mirtazapine vs. Fluoxetine	0.713	0.518	0.196	4	0.667	0.444	0.222	1	-0.027
Sertraline vs. Fluoxetine	0.559	0.505	0.054	4	0.673	0.464	0.209	1	-0.155
Amitriptyline vs. Sertraline	0.500	0.529	-0.029	1	0.526	0.452	0.074	3	-0.104
Amitriptyline vs. Trazodone	0.557	0.435	0.122	2	0.566	0.467	0.099	2	0.023
Clomipramine vs. Fluoxetine	0.733	0.800	-0.067	1	0.552	0.665	-0.113	3	0.046
Trazodone vs. Fluoxetine	0.765	0.476	0.289	1	0.431	0.496	-0.065	3	0.353
Amitriptyline vs. Fluvoxamine	0.618	0.371	0.246	1	0.368	0.507	-0.139	2	0.385
Sertraline vs. Venlafaxine	0.549	0.628	-0.079	1	0.570	0.622	-0.052	2	-0.028
Amitriptyline vs. Citalopram	0.650	0.625	0.025	1	0.516	0.548	-0.031	1	0.056
Clomipramine vs. Venlafaxine	0.672	0.533	0.139	1	0.400	0.577	-0.177	1	0.316
Fluvoxamine vs. Milnacipran	0.537	0.660	-0.123	1	0.571	0.702	-0.130	1	0.007
Paroxetine vs. Bupropion	0.395	0.400	-0.005	1	0.507	0.507	-0.000	1	-0.005
Paroxetine vs. Escitalopram	0.564	0.621	-0.057	1	0.698	0.675	0.023	1	-0.080
Paroxetine vs. Fluvoxamine	0.436	0.369	0.067	1	0.533	0.567	-0.033	1	0.101
Reboxetine vs. Citalopram	0.421	0.557	-0.136	1	0.609	0.600	0.009	1	-0.145
Sertraline vs. Citalopram	0.695	0.680	0.015	1	0.231	0.360	-0.129	1	0.144
Sertraline vs. Fluvoxamine	0.583	0.725	-0.142	1	0.479	0.551	-0.072	1	-0.070
Trazodone vs. Paroxetine	0.873	0.906	-0.033	1	0.413	0.560	-0.148	1	0.115
Venlafaxine vs. Citalopram	0.645	0.667	-0.022	1	0.429	0.840	-0.411	1	0.390
Venlafaxine vs. Sertraline	0.628	0.549	0.079	1	0.667	0.709	-0.042	1	0.122

Notes: This table presents the difference-in-difference estimate of the sponsorship effect for "Active vs. Active" drug sets. The first set of columns compares the share of patients that respond to treatment when the first listed drug is sponsored; the next set compare the share of patients that respond when the first listed drug is not sponsored. In all cases, the second listed drug has no change in sponsorship interests. The difference between the share of patients that respond to a given drug and the share that respond to the other arm is given in the column labeled "Diff" for "Difference." The last column reports the difference between the two difference columns. This difference in difference (DD) is analogous to the sponsorship effect in equation 1.

Table 5: Effect of Sponsorship on Drug Efficacy

Panel A: Drug Set

	Standardized Outcome (Relative)		Significant (5%)	Significant (10%)	Most Effective Arm
	(1a)	(2a)	(3a)	(4a)	(5a)
$Sponsor_{ij}$	0.152**	0.153**	0.104**	0.059	0.164*
	(0.077)	(0.076)	(0.040)	(0.043)	(0.087)
Controls		X	X	X	X
Drug by Drug Set F.E.	X	X	X	X	X
Mean Outcome	0.42	0.42	0.22	0.26	0.39
N	1,412	1,412	1,284	1,284	1,412

Panel B: Drug Pair

	Standardized Outcome (Relative)		Significant (5%)	Significant (10%)	Most Effective Arm
	(1b)	(2b)	(3b)	(4b)	(5b)
$Sponsor_{ij}$	0.135***	0.124***	0.081**	0.084**	0.241***
	(0.049)	(0.047)	(0.037)	(0.039)	(0.051)
Controls		X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X
Mean Outcome	0.34	0.34	0.22	0.26	0.39
N	2,332	2,332	2,083	2,083	2,332
$Weighted\,N$	1,412	1,412	1,284	1,284	1,412

Note: Panel A presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 1, where the fixed effects $G_{d(i),s(j)}$ control for each drug in each unique drug combination. Panel B presents coefficients from the estimation of equation 2, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. See section 3.2 for more detail. In columns (1a),(2a),(1b), and (2b), the dependent variable y_{ij} is the standardized efficacy measure, relative to the placebo arm if available or least effective arm in that trial otherwise. In columns (3a), (4a), (3b) and (4b) the dependent variable y_{ij} is an indicator for whether arm i in trial j was statistically significantly different from the other arms in that trial. The statistical significance level was computed using a one-sided test in placebo-controlled trials, and a two-sided test in trials with only active drugs. In column (5a) and (5b), the dependent variable y_{ij} is an indicator for whether arm i was the most effective arm in trial j. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 6: Robustness of Sponsorship Effect

Panel A: Drug Set

		Mean Reversion Tests				
	Baseline	Control for Trial Order	Control for Year Relative to Approval	Restrict to Post Approval	Remove COI Sponsorship	Weight by Enrollment
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
$Sponsor_{ij}$	0.153**	0.142*	0.146*	0.131*	0.093	0.125**
	(0.076)	(0.078)	(0.084)	(0.079)	(0.087)	(0.062)
Controls	X	X	X	X	X	X
Drug by Drug Set F.E.	X	X	X	X	X	X
Mean Outcome	0.42	0.42	0.42	0.50	0.42	0.38
N	1,412	1,412	1,412	992	1,412	1,412

Panel B: Drug Pair

	Baseline	Control for Trial Order	Control for Year Relative to Approval	Restrict to Post Approval	Remove COI Sponsorship	Weight by Enrollment
	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
$Sponsor_{ij}$	0.124***	0.110**	0.102**	0.099**	0.104**	0.057
	(0.047)	(0.047)	(0.047)	(0.048)	(0.050)	(0.039)
Controls	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X
Mean Outcome	0.34	0.34	0.34	0.39	0.34	0.29
Weighted N	1,412	1,412	1,412	992	2,332	1,412

Note: Panel A presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 1, where the fixed effects $G_{d(i),s(j)}$ control for each drug in each unique drug combination. Panel B presents coefficients from the estimation of equation 2, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. Columns (1a) and (1b) replicates the main results from Table 5, columns (2a) and (2b), where the outcome is the standardized efficacy measure, relative to the placebo arm if available or least effective arm in that trial otherwise. The dependent variable is the same in all subsequent columns. Columns (2a) and (2b) include controls for the order that the trial occurred within the drug set., while columns (3a) and (3b) include control for the year the trial was published relative to the drug approval year. Columns (4a) and (4b) restrict the sample to exclude trials that were published before one of the drugs in the trial was approved by the FDA. Columns (5a) and (5b) exclude trials for which the only sponsorship indication is a conflict of interest (COI) statement. Columns (6a) and (6b) weight each trial's arm by the total enrollment in that arm. Baseline controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 7: Sponsorship Effect by Drug Type and Outcome

Panel A: Drug Set

		Antidepressants			Antipsychotics		
	Baseline	Standardized Outcome	Share Respond	% Decline	Standardized Outcome	PANSS Decline	% Decline
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)	(7a)
$Sponsor_{ij}$	0.153**	0.195*	0.031*	0.021*	0.122	1.196	0.007
	(0.076)	(0.100)	(0.016)	(0.011)	(0.088)	(0.819)	(0.008)
Controls	X	X	X	X	X	X	X
Drug by Drug Set F.E.	X	X	X	X	X	X	X
Mean Outcome	0.42	0.40	0.06	0.05	0.49	4.62	0.08
N	1,412	1,090	1,090	960	322	218	287

Panel B: Drug Pair

	Baseline	Standardized Outcome	Share Respond	% Decline	Standardized Outcome	PANSS Decline	% Decline
	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)	(7b)
$Sponsor_{ij}$	0.124***	0.160***	0.025***	0.025***	0.061	0.913	0.006
	(0.047)	(0.060)	(0.009)	(0.009)	(0.059)	(0.684)	(0.005)
Controls	X	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X	X
Mean Outcome	0.34	0.35	0.05	0.04	0.31	2.94	0.05
Weighted N	1,412	1,090	1,090	964	322	218	287

Note: Panel A presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 1, where the fixed effects $G_{d(i),s(j)}$ control for each drug in each unique drug combination. Panel B presents coefficients from the estimation of equation 2, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. See section 3.2 for more detail. Columns (1a) and (1b) replicate the main results from Table 5, columns (2a) and (2b), where the outcome is the standardized efficacy measure, relative to the placebo arm if available or least effective arm in that trial otherwise. In columns (2a) and (2b), I include only antidepressants. In columns (3a) and (3b), I present results using the unstandardized antidepressant outcome: the share of patients that responded to treatment for arm i in trial j. I also use the percent decline in the depression score as an outcome in columns (4a) and (4b). The last three columns consider only antipsychotics. Columns (5a) and (5b) use the baseline outcome. Columns (6a) and (6b) use the unstandardized antipsychotic outcome: the mean decline in the PANSS for arm i in trial j. This restricts the sample since many antipsychotic trials consider different scales. Finally, columns (7a) and (7b) use the percent decline in the psychotic score as an outcome. In all cases, outcomes are reported relative to the placebo or least effective arm in that trial. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 8: Sponsorship Variation by Trial Characteristics

	# Trials	Share with Variation
Full Sample	586	0.39
Drug Type - Antidepressants		
Tricyclic	78	0.58
Atypical	173	0.22
SSRI	316	0.53
SNRI	110	0.29
Drug Type - Antipsychotics		
1st Gen	48	0.52
2nd Gen	102	0.43
Placebo	319	0.21
Earliest Approval Year		
Prior to 1990	151	0.54
1990 - 1999	202	0.51
2000 or after	233	0.19
# Trials in a Drug Set		
Less than 5	293	0.27
5-9 trials	160	0.46
10 or more	133	0.59

Note: This table presents the share of trials with each characteristic that have variation in sponsorship. Drug types are not mutually exclusive, since a trial can compare drugs of different types. A drug set refers to a unique combination of drugs, so "# Trials" refers to how many total trials examine the same drug set as a given trial.

Table 9: Characteristics of Sponsored Arms

	N	Length	Base Severity	Dropout Rate	Dosage	Age	Gender
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Sponsor_{ij}$	18.96	0.82	-1.42	-0.09	-3.18	2.94*	0.01
	(12.72)	(0.71)	(7.38)	(0.15)	(2.37)	(1.37)	(0.04)
Controls	X	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X	X
Mean Outcome	100.07	8.74	68.11	-0.01	29.43	42.36	0.52
Weighted N	1,412	1,412	1,051	1,310	1,269	1,152	792

Note: This table presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 2, where dependent variable y_{ij} is now a given trial characteristic. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 10: Predicted Sponsorship Effect Using Individual Characteristics

	Trial Characteristics			Patient Characteristics			
	N	Length	Dosage	Baseline Severity	Dropout Rate	Age	Gender
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Sponsor_{ij}$	-0.02	-0.01	0.01	-0.02	0.04	0.01	-0.02
	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.01)	(0.02)
Controls	X	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X	X
Mean Outcome	0.19	0.22	0.20	0.09	0.23	0.27	0.26
Weighted N	1,412	1,412	1,412	1,412	1,412	1,412	1,412

Note: This table presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 7, where the dependent variable is predicted drug efficacy. Each column predicts drug-specific efficacy using a different trial characteristics, as shown in equation 6. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 11: Predicted Sponsorship Effect Using All Characteristics

	Trial Characteristics		
	All Trial	All Patient	Both
	(1)	(2)	(3)
$Sponsor_{ij}$	-0.08**	0.02	-0.01
	(0.03)	(0.03)	(0.04)
Controls	X	X	X
Drug by Drug Pair F.E.	X	X	X
Mean Outcome	0.28	0.31	0.32
Weighted N	1,412	1,412	1,412

Note: This table presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 7, where the dependent variable is predicted drug efficacy. Predicted drug-specific efficacy is estimated using all trial characteristics (column 1), all patient characteristics (column 2), or both (column 3) as in equation 6. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 12: Publication by Efficacy

	Published	
	(1)	(2)
$Sponsor_{ij}$	-0.006	-0.065**
	(0.025)	(0.032)
Standardized Outcome (Relative)		0.043
		(0.029)
$Sponsor_{ij}$ x Standardized Outcome (Relative)		0.091**
		(0.035)
Controls	X	X
Drug by Drug Pair F.E.	X	X
Mean Outcome	0.85	0.85
Weighted N	1,412	1,412

Note: This table presents the coefficients from the estimation of equation 8, where the outcome is an indicator for whether the trial was published. Column (1) presents the coefficient on $Sponsor_{ij}$, excluding the interaction term. Column (2) presents the coefficients from the estimation of equation 8 with the interaction term. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 13: Publication by Pre-Registration

	Standardized Outcome (Relative)	
	(1)	(2)
$Sponsor_{ij}$	0.124**	0.135**
	(0.047)	(0.048)
Linked to ClinicalTrials.gov		0.048
		(0.038)
$Sponsor_{ij}$ x Linked to ClinicalTrials.gov		-0.094
		(0.082)
Controls	X	X
Drug by Drug Pair F.E.	X	X
Mean Outcome	0.48	0.48
Weighted N	2,332	2,332

Note: Table presents the coefficients from the estimation of equation 9. Column (1) presents the coefficient on $Sponsor_{ij}$, excluding the interaction term. Column (2) presents the coefficients on $Sponsor_{ij}$ interacted with an indicator for whether the trial was linked to ClinicalTrials.gov from the estimation of equation 9. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Appendix A: Theory Appendix

Statistical Significance Calculation

In Table 5 columns (3), (4), (8), and (9), the outcome is an indicator for whether the drug was statistically significantly more effective than the placebo arm or least effective arm in that trial. The efficacy outcome—the proportion of patients that responded to treatment—was considered statistically significant if the Z-score, computed as

$$Z = \frac{p_1 - p_2}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{11}$$

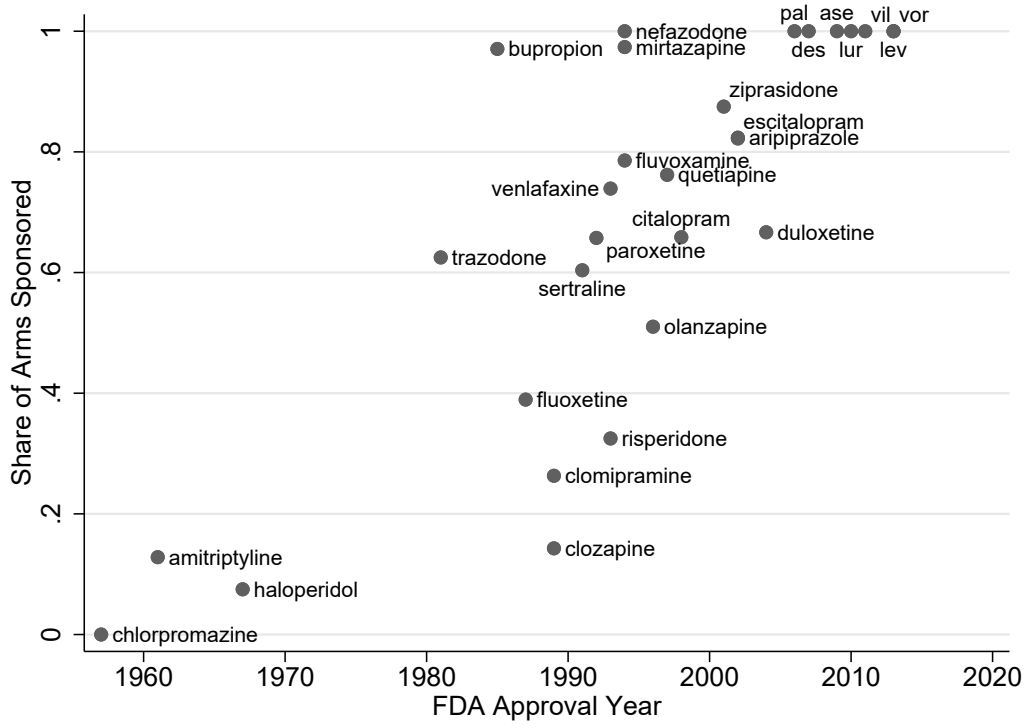
was significant at the 5% level. With an infinite sample, this Z-score cutoff was 1.64 for placebo-controlled trials and 1.96 for head to head trials. Here p is the proportion of patients that respond to treatment. The numeric indexing in equation 11 refers to the first or second arm, and $\hat{p}$ is the overall proportion for both arms. The variable n refers to the number of patients in each arm. For schizophrenia trials, the Z-score was computed as

$$Z = \frac{e_1 - e_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)}} \tag{12}$$

where e is the decline in schizophrenia score, σ is the standard deviation of this decline, and n is the sample size in that arm.

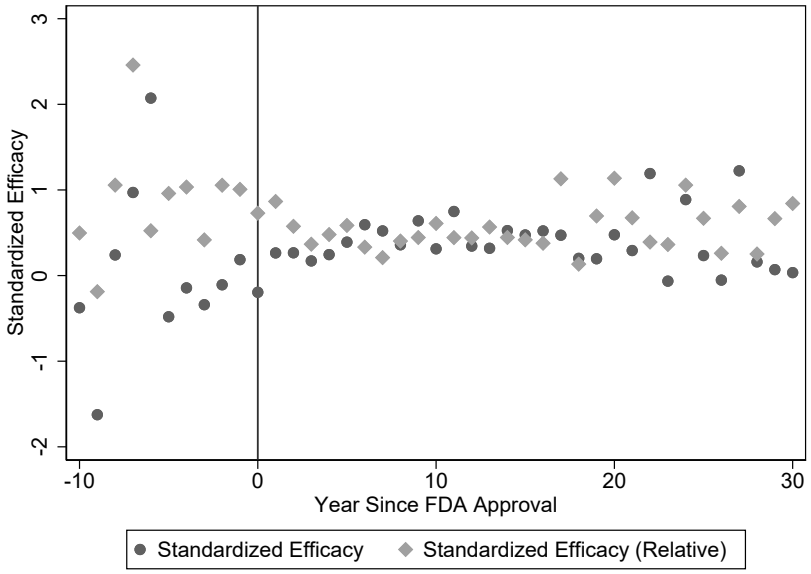
Appendix B: Tables and Figures

Figure B1: Included Drugs



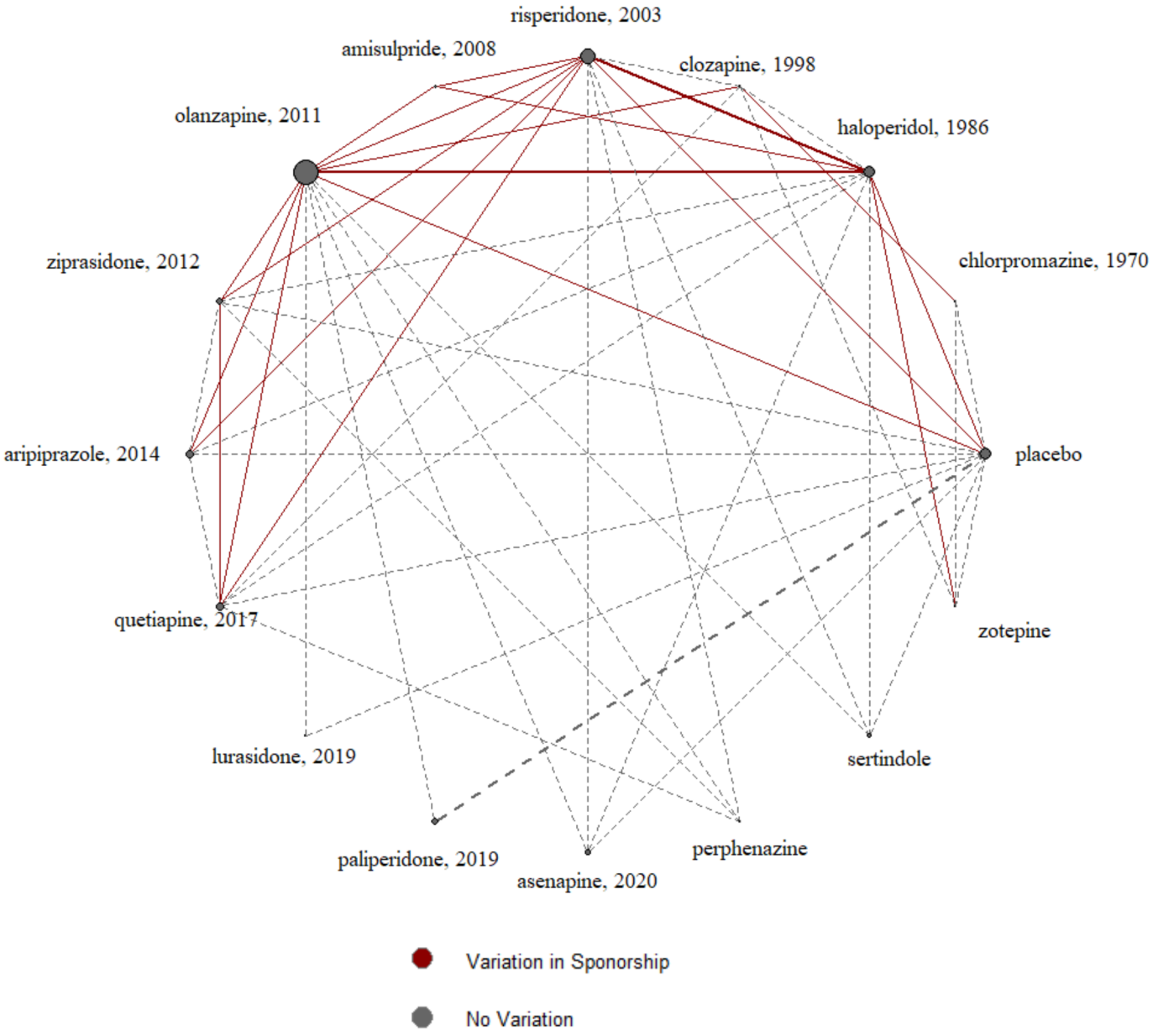
Notes: This figure presents the antidepressant and antipsychotic drugs included in this analysis. The x-axis presents the year of FDA approval for the drug, while the y-axis plots the share of arms in which that drug is sponsored by its manufacturer or marketer. The label "ase" refers to asenapine, "lur" refers to lurasidone, "vil" refers to vilazodone, "lev" refers to levomilnacipran, and "vor" refers to vortioxetine. My analysis sample also includes agomelatine, amisulpride, milnacipran, reboxetine, sertindole, and zotepine which are not yet approved in the United States and thus not shown in this figure.

Figure B2: Efficacy by Year Since Drug Approval



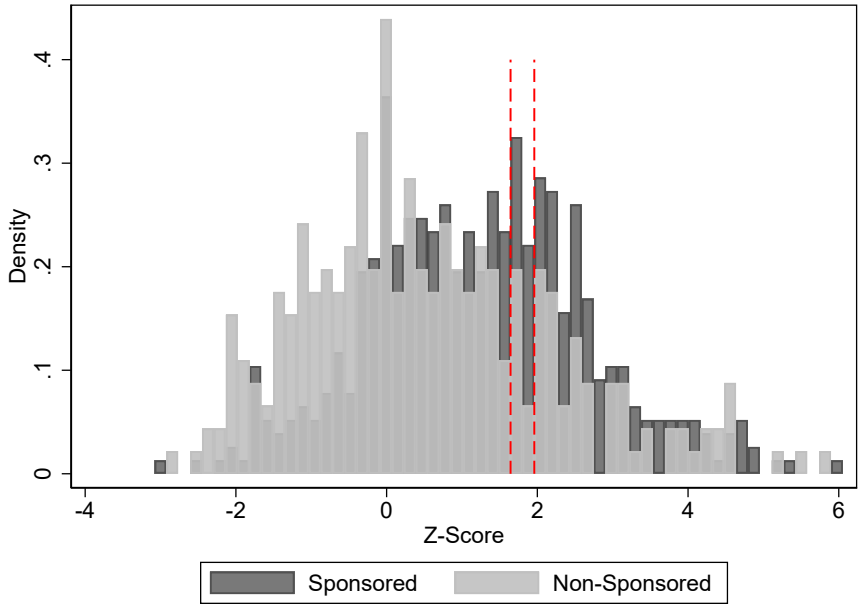
Notes: This figure presents the relationship between effectiveness and year since FDA approval. The x-axis plots the year the arm was published or made publically available, relative to the FDA approval year for that drug. The y-axis plots the mean standard efficacy, or the mean standard efficacy, relative to the least effective or placebo arm in the drug pair. Details on these outcomes are listed in Section 2.3.3.

Figure B3: Network of Trials for Antipsychotics



Notes: This figure presents the network of comparisons within antipsychotics. Each node represents a drug and is labeled with the year that a generic formulation entered the United States market (years after 2019 are estimates). The size of the circle is proportional to the number of randomly assigned participants. Each line represents a clinical trial comparing the two drugs. A trial with three or more drugs would have a line between every pair of drugs tested. The width of the lines is proportional to the number of trials comparing every pair of treatments. Lines in solid red denote that the sponsorship status of at least one of the drugs varies within the trials; lines in dashed gray denote that the sponsorship status of both drugs is constant.

Figure B4: Distribution of Z-Scores Conditional on Publication



Notes: This figure presents the distribution of z-scores for drug efficacy in published trials. Both placebo-controlled and head-to-head trials are included. I test for bunching at $Z = 1.645$ (5%, one sided, 10%, two sided) and $Z = 1.96$ (5%, two sided).

Table B1: Fixed Effect Example

	(1) Drug by Drug Set Fixed Effects		(2) Drug by Drug Pair Fixed Effects	
Trial	$G_{d(i),s(j)}$	Drug	$G_{d(i),p(j)}$	Drug
X	1	Drug A	1	Drug A
X		Placebo		*Placebo*
Y	1	Drug A	1	Drug A
Y		Placebo		*Placebo*
Z	2	Drug A	1	Drug A
Z		Herbal Supplement		Herbal Supplement
Z		Placebo		*Placebo*
W	3	Drug A	1	Drug A
W		Drug B		Drug B
W		Placebo		*Placebo*
W			2	Drug A
W				Drug B
W				Placebo
K	4	Drug A	2	Drug A
K		Drug B		Drug B
Q	5	Drug A	3	Drug A
Q		Drug C		Drug C

Notes: This table provides an example of the fixed effects in equation 1 and 2 based on six hypothetical trials: X,Y, Z, W, K, and Q. Each row represents a treatment arm (i.e. drug) in my sample. The $G_{d(i),s(j)}$ and $G_{d(i),p(j)}$ columns present the fixed effects for Drug A; each number represents a different fixed effect. The fixed effects for the other drugs are omitted. Column (1) presents the more restrictive drug by drug set fixed effects $G_{d(i),s(j)}$. In this case, each different drug set has a separate fixed effect for Drug A. The first two trials assess the same drug set, so Drug A has the same fixed effect in those two trials. Each of the other four trials assess a different drug set, so Drug A has four separate fixed effect in these trials. Column (2) presents the less restrictive drug by drug pair fixed effects $G_{d(i),p(j)}$. In this case, Drug A gets a separate fixed effect for each different drug it is directly compared against. Here, Drug A has the same fixed effect for the first four trials, where it is compared with a placebo. In trial W, Drug A also has a separate fixed effect since it is compared with Drug B as well; this is the same fixed effect as in trial K. In this case, trial W would be re-weighted so that this arm is not double counted.

Table B2: Difference in Difference: Active versus Active Antipsychotics

	Sponsored				Not Sponsored				
	Decline in Score				Decline in Score				
	Drug	Other Arm	*Diff*	# Trials	Drug	Other Arm	*Diff*	# Trials	DD
All Drug Sets	20.50	19.53	0.97	31	17.86	16.99	0.87	23	0.10
Olanzapine vs. Haloperidol	21.09	16.51	4.57	10	6.57	4.37	2.20	2	2.37
Risperidone vs. Haloperidol	16.52	15.00	1.52	5	25.44	23.07	2.37	7	-0.85
Amisulpride vs. Risperidone	24.47	23.17	1.30	3	24.10	28.40	-4.30	1	5.60
Olanzapine vs. Aripiprazole	31.50	27.30	4.20	1	24.32	23.93	0.39	3	3.81
Olanzapine vs. Amisulpride	35.00	45.00	-10.00	1	22.56	20.85	1.72	2	-11.72
Risperidone vs. Olanzapine	11.25	11.00	0.25	2	4.90	4.70	0.20	1	0.05
Ziprasidone vs. Olanzapine	13.13	14.53	-1.40	2	26.00	35.70	-9.70	1	8.31
Zotepine vs. Haloperidol	13.82	14.78	-0.97	2	5.00	6.20	-1.20	1	0.24
Amisulpride vs. Haloperidol	27.30	21.90	5.40	1	20.90	17.30	3.60	1	1.80
Amisulpride vs. Olanzapine	25.00	28.00	-3.00	1	45.00	35.00	10.00	1	-13.00
Clozapine vs. Chlorpromazine	21.10	20.80	0.30	1	19.94	14.48	5.46	1	-5.16
Haloperidol vs. Risperidone	4.60	13.80	-9.20	1	15.00	16.52	-1.52	1	-7.68
Olanzapine vs. Risperidone	28.10	24.90	3.20	1	4.70	4.90	-0.20	1	3.40

Notes: This table reports the difference-in-difference estimate of the sponsorship effect for "Active vs. Active" schizophrenia drug sets. The first set of columns compares the efficacy rates when the first listed drug is sponsored; the next set compare these results when the first listed drug is not sponsored. In all cases, the second listed drug has no change in sponsorship interests. The difference between the share of patients that respond to a given drug and the share that respond to the other arm is given in the column labeled "Diff" for "Difference." The last column reports the difference between the two difference columns. This difference in difference (DD) is analogous to the sponsorship effect in equation 1.

Table B3: Sponsorship by Drug Set Type

Panel A: Drug Set

		Drug Set Type			
	Baseline	Active vs. Placebo	Active vs. Active	Three or More Arms	Only Variation
	(1a)	(2a)	(3a)	(4a)	(5a)
$Sponsor_{ij}$	0.153**	0.335	0.148*	0.233	0.168**
	(0.076)	(0.279)	(0.083)	(0.248)	(0.068)
Controls	X	X	X	X	X
Drug by Drug Set F.E.	X	X	X	X	X
Mean Outcome	0.42	0.49	0.27	0.56	0.37
N	1,412	541	529	342	499

Panel B: Drug Pair

	Baseline	Active vs. Placebo	Active vs. Active	Only Variation
	(1b)	(2b)	(3b)	(5b)
$Sponsor_{ij}$	0.124***	0.173	0.095*	0.132***
	(0.047)	(0.086)	(0.055)	(0.046)
Controls	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X
Mean Outcome	0.34	0.46	0.29	0.37
Weighted N	2,332	1,044	1,764	1,138

Note: Panel A presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 1, where the fixed effects $G_{d(i),s(j)}$ control for each drug in each unique drug combination. Panel B presents coefficients from the estimation of equation 2, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. See section 3.2 for more detail. Columns (1a) and (1b) replicates the main results from Table 5, columns (2a) and (2b), where the outcome is the standardized efficacy measure, relative to the placebo arm if available or least effective arm in that trial otherwise. The dependent variable is the same in all columns. Columns (2a)–(4a) and (2b)–(3b) present results split by drug set type, as described in Table 2. In columns (5a) and (5b), I restrict to only drug sets or drug pairs with variation in sponsorship. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table B4: Alternate Specifications

Panel A:	**Standardized Outcome (Relative)**			
	(1a)	(2a)	(3a)	(4a)
$Sponsor_{ij}$	0.153**	0.124***	0.156***	0.353***
	(0.076)	(0.047)	(0.045)	(0.031)
Controls	X	X	X	X
Drug Combination Fixed Effects	Drug by Drug Set	Drug by Drug Pair	Drug	None
Mean Outcome	0.42	0.34	0.42	0.42
Weighted N	1,412	1,412	1,412	1,412
Panel B:	**Standardized Outcome**			
	(1b)	(2b)	(3b)	(4b)
$Sponsor_{ij}$	0.311**	0.150	0.010	0.375***
	(0.151)	(0.094)	(0.082)	(0.047)
Controls	X	X	X	X
Drug Combination Fixed Effects	Drug by Drug Set	Drug by Drug Pair	Drug	None
Mean Outcome	0.00	0.00	0.00	0.00
Weighted N	1,412	1,412	1,412	1,412

Note: This table presents estimates of the sponsorship effect with alternate specifications. Column (1) presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 1, where the fixed effects $G_{d(i),s(j)}$ control for each drug in each unique drug combination. Column (2) presents coefficients from the estimation of equation 2, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. In column (3) I include only drug fixed effects, and column (4) has no drug-specific fixed effects. See section 3.2 for more detail. In the top panel, the dependent variable is the standardized efficacy measure, relative to the placebo arm if available or least effective arm in that trial otherwise. In the bottom panel, the dependent variable y_{ij} is the standardized efficacy measure for arm i in trial j. Columns (1a) and (2a) replicate the main results from Table 5, columns (2a) and (2b). Controls include the trial's publication year and the type of psychiatric score used. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.