Funding of Clinical Trials and Reported Drug Efficacy*

Tamar Oostrom†

February 18, 2022

Abstract

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. I use newly assembled data on psychiatric clinical trials to estimate that a drug appears substantially more effective when the trial is sponsored by that drug's manufacturer, compared with the same drug in the same set of drugs but without the drug manufacturer's involvement. This difference is not explained by observable characteristics of trial design. Publication bias is a key mechanism and pre-registration requirements may be effective in overcoming sponsorship bias.

JEL Codes: I11, I18, O31

Keywords: Innovation, Research and Development, Prescription Drugs

*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 and Frank Schilbach for helpful comments and support. This paper also benefited from discussions with Sarah Abraham, David Autor, Ivan Badinski, Jane Choi, Laura Dague, Joe Doyle, Colin Gray, Ryan Hill, Allan Hsiao, Simon Jaeger, Kurt Lavetti, Madeline Mckelway, Adrienne Sabety, Parinitha Sastry, Cory Smith, Amanda Starc, Carolyn Stein, Scott Stern, Sean Wang, Bruce Weinberg, Luigi Zingales, seminar participants at the AEA/ASSA Annual Meeting, Brookings Institute, BU/Harvard/MIT Health Seminar, Columbia Business School, Electronic Health Economics Colloquium, Harvard Business School, Ohio State, Reed College, UCLA Anderson, University of Illinois at Chicago, the Food and Drug Administration, 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 many markets, consumers and policy-makers have incomplete information on the effectiveness and quality of products. As a result, firms often fund information about their products themselves. For example, automakers run fuel-economy tests for new vehicles, sunscreen manufacturers pay laboratories to test their products, and clinical trials are often conducted by the drug's manufacturer. Firm's research may have welfare benefits as other parties can use the knowledge produced at minimal marginal cost and improve their consumption or production choices (Arrow, 1962). On the other hand, industry research may have specific, less-relevant characteristics, and the knowledge produced may not be shared with the public (Angell, 2000). This paper measures how industry and financial incentives shape available evidence in the pharmaceutical market.

Clinical trials are a key component of pharmaceutical research and development. These trials are expensive and risky investments: the average cost of a late-stage clinical trial is $35 million, an estimated 70% of trials are funded by industry, and pharmaceuticals are a $480 billion market in the United States (Yu et al., 2018; Moore et al., 2018; Wood, 2018). The results of trials form the basis of regulatory, prescribing, and medical treatment decisions for decades (Davidoff et al., 2001). For instance, trials have direct consequences for the health of the population, as seen by trials on the benefits of statins, the risks of hormone replacement therapy, and recent COVID-19 vaccines.

This paper quantifies how financial incentives affect the results of randomized control trials (RCTs), specifically clinical trials. I also estimate the downstream consequences of financial incentives in terms of trial characteristics and the availability of the research produced. My identification strategy uses the key insight that the exact same sets of drugs can be tested in *different* RCTs conducted by parties with different financial interests. This approach is useful for evaluating the bias and external validity of RCTs in other settings but is infrequently implemented due to data constraints.[1]

I construct a novel data set of psychiatric clinical trials where the exact same sets of drugs are examined in trials with different sponsorship interests. I focus on antidepressants and antipsychotics due their market size as well as data availability. Psychiatric drugs have a large market: 12.7% of the U.S. adult population takes antidepressant medication monthly, 1.6% take antipsychotics (Pratt et al., 2017; Moore and Mattison, 2017), and the economic burden of depressive disorders in the United States is estimated to be $210 billion annually (Greenberg et al., 2015). Antidepressants and antipsychotics also conveniently had several large and recently published meta-analyses on their efficacy (Cipriani et al.,

[1]Within economics, one notable exception is Allcott (2015) which estimates site selection bias using an energy conservation program. My findings therefore relate to a recent literature in medicine and economics on sources of bias and external validity in RCTs (e.g.Vivalt (2020); Abrams et al. (2021)).

2018; Leucht et al., 2013), which enable me to clearly define the relevant sample of drugs.[2]

As an example of my identifying variation, Wyeth Pharmaceuticals introduced a new antidepressant drug venlafaxine, also known as Effexor, in 1993. Over the next decade and a half, Wyeth funded RCTs that compared the effectiveness of venlafaxine with Eli Lilly's blockbuster drug Prozac. In 86 percent of the fourteen trials funded solely by Wyeth, venlafaxine was more effective than Prozac. In contrast, only 33 percent of the three trials with alternate funding found that venlafaxine was more effective. Each of these trials is a double blind RCT comparing the exact same two molecules and examining the same standard outcomes.[3] Motivated by this example—which might be due to idiosyncratic differences across these trials—I systematically investigate the effect of an RCT's funder on the reported efficacy of the tested drugs.

First, I use the variation in drug funding to show that financial incentives affect reported drug efficacy. I find that a drug is reported to be 36 percent more effective (0.16 standard deviations off a base of 0.45) 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 or marketer's involvement.[4] Sponsored drugs are also 49 percent more likely to report statistically significant improvements (0.12 off a base of 0.24), and 38 percent more likely to be the most effective drug in their trial (0.15 off a base of 0.39), again, compared with the same molecule tested against the same set of drugs, but without funding from the drug's manufacturer. I term my main effect a "sponsorship effect."

Identification of the causal effect of sponsorship requires that, within the same drug and drug combination, trials with alternate funding are equivalent tests of a drug's efficacy. My estimates use variation in funding both within-industry and across industry versus non-profit trials. I find consistent results across these different types of variation and my results hold when considering only within-industry trials. In balance checks, I show that, within a set of drugs, trials with different funding are similar in observable trial and enrollment characteristics.

In terms of external validity, the set of drugs that have variation in funding is broadly representative of the full set of psychiatric drugs, though older drugs are more likely to be included. I also find that the sponsorship effect is greater for drugs with a larger market size, as measured by Medicaid prescriptions.

[2]Each trial in my sample is a double-blind RCT. These trials were conducted before and, mostly, after the drugs gained regulatory approval. 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 the authors are academic researchers at a university or medical school. Section 2.4 contains more information.

[3]These trials often differed slightly in trial characteristics or examined additional outcomes. For example, these trials studied outpatients in Portugal, inpatients in France, patients in Latin America, looked at the association of treatment response with genetic markers in Taiwan, had an initially increased dosage of venlafaxine, looked at the activation of neural circuits in the United Kingdom, also examined two-year outcomes, or additionally examined readmission rates.

[4]I measure efficacy using the standard outcomes in the medical literature. See section 2.2.3. As a separate point, clinical trial results may choose to report and highlight a selected set of outcomes. In this analysis, I focus on a consistent set of outcomes to focus on differences in reported efficacy, not reporting decisions.

This result is consistent with the effect of sponsorship being driven by financial incentives of sponsors.

Secondly, I investigate the mechanisms of this sponsorship effect. There are two classes of potential mechanisms. Trials could either be planned or conducted differently ex-ante or presented and published differently ex-post. I show that the main effect is driven the second class of mechanisms, referred to as publication bias. Publication bias affects the availability of information after completion of the clinical trial. For example, the publication or data access decision may depend more on the manufacturer's drug's effectiveness than the effectiveness of the other drugs in the clinical trial. Trials in which the manufacturer's drug appears more effective are more likely to be published, while this relationship between outcomes and publication is attenuated for drugs without financial involvement. I incorporate data on unpublished clinical trials to quantify the importance of publication bias in explaining the main effect. The addition of unpublished trials attenuates the effect of sponsorship, and I find that approximately half of this sponsorship effect can be explained by publication bias.[5]

Another class of potential mechanisms is trial design, where trials are planned or conducted differently.[6] I test for this mechanism by incorporating data on trial characteristics including the length of the trial, the drug's dosage, total enrollment, recruitment location as well as the average age, gender, and baseline severity of the enrolled patients. For each of the trial characteristics, I estimate how this characteristic would affect each drug's effectiveness. I find no evidence that drugs that are tested in clinical trials funded by their manufacturer have more favorable trial design choices than when that same drug is tested against the same set of drugs but without the manufacturer's involvement. This analysis is constrained by characteristics which are observable, and some part of the sponsorship effect could be due to selection on aspects of trial design that are unobserved. However, these unobserved characteristics would have to be differently selection from the numerous observed characteristics, none of which explain this effect. The remaining unexplained share of the sponsorship effect could be due to underestimating the publication channels described above, data manipulation and reconciliation errors, or due to noise in estimating the mechanisms.

Finally, the relevance of publication bias in explaining the main sponsorship effect suggests an obvious policy implication: the required pre-registration of clinical trials. Starting in 2005, the International Committee of Medical Journal Editors (ICMJE) required pre-registration as a condition for publication in their journals (De Angelis et al., 2004). I quantify the significance of pre-registration in limiting

[5]This estimate is likely a lower bound. Some recent clinical trials were neither published nor pre-registered, and the unpublished trials I observe may be more favorable to the sponsors than all unpublished trials. In those cases, my estimate would underestimate the share explained by publication bias.

[6]Interviews with clinical trial managers highlight several potential mechanisms for conflicts of interest to manifest through trial characteristics, such as prematurely stopping the trials or manipulating the randomization or enrollment process (Østengaard et al., 2020).

publication bias and find that the effect of sponsorship on reported drug efficacy is statistically significantly lower after the introduction of ClinicalTrials.gov, the pre-registration platform, compared with the sponsorship effect before required pre-registration. In addition, the set of trials pre-registered in ClinicalTrials.gov has a statistically significantly lower sponsorship effect than the trials that were not pre-registered. While there may be other concurrent changes in social norms and transparency regarding clinical trials, these results suggest that pre-registration requirements may be effective in overcoming sponsorship bias and provide additional support for publication bias as a key mechanism.

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 (e.g., Bourgeois et al. (2010); Bekelman et al. (2003)). 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). 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. My evidence may also speak to the effects of financial incentives in the context of RCTs in e.g.. development economics, education, and environmental economics.

Broadly, this paper speaks to the value of pharmaceutical innovation and finds evidence that financial incentives have a large effect on the type and value of knowledge that is produced. My results on mechanisms suggest that clinical trial publications, or alternately the decision to not publish results, are valued by pharmaceutical firms. This is consistent with evidence on the effectiveness of direct-to-consumer advertising (Sinkinson and Starc, 2019; Shapiro, forthcoming) as well as detailing (Mizik and Jacobson, 2004), both of which rely on scientific publications in their material. Finally, this paper builds on a growing literature analyzing incentives for private research investments. Most of this research focuses on how financing and other policies affect the quantity of research activities (e.g. Budish et al. (2015); Gross and Sampat (Working Paper); Azoulay et al. (Working Paper). In contrast, this paper speaks to how incentives affect research quality, as measured uniquely by the bias in trial estimates. In terms of policy implications, this paper also finds suggestive evidence that ClinicalTrials.gov and pre-registration requirements reduced bias due to financial conflicts of interest.[7]

This paper provides evidence that the funder of a trial affects the reported efficacy of its tested drugs. Does this evidence have consequences for drug approval and prescription decisions? If physicians, pa-

[7]Within economics, pre-registration is not required and there are fewer conventions for consistent outcomes than among medical trials; accordingly, economics registries have arguably been less effective than the ICJME's pre-registration requirements (Abrams et al., 2021).

tients, and regulators already appropriately incorporate the role of the sponsor, then altering trial funding would not affect approvals and prescriptions. While there is some evidence that physicians discount trials with pharmaceutical funding (Kesselheim et al., 2012), evidence on how actual prescriptions respond to clinical trial results does not consider differences in funding (Azoulay, 2004; McKibbin, 2020). My results suggest that sponsored arms of trials should be discounted substantially about 30-50% in terms of efficacy. This magnitude is consequential for drug approvals – back of the envelope calculations relate this decrease in efficacy to three fewer psychiatric drug approvals (10%). In terms of prescriptions this reduction in efficacy is associated with 21% fewer Medicaid prescriptions.

Section 2 presents institutional background on clinical trials and psychiatric drugs and presents the data, sample construction, and summary statistics. I outline my empirical strategy and present my main results on the effect of sponsorship on reported drug efficacy in section 3. Section 4 investigates mechanisms, focusing on trial design and publication bias and estimates the share of the sponsorship effect explained by each component. Section 5 analyzes the effect of required pre-registration, and section 6 concludes and discusses implications for the funding of clinical trials.

2 Clinical Trials and Psychiatric Drugs

2.1 Clinical Trial Background

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 development. One estimate of the research and development spending per drug approved was $2.6 billion (DiMasi et al., 2016). Drug development begins with pre-clinical testing of new molecules in non-human subjects. Subsequent clinical trials in humans are organized into Phase I, Phase II, and Phase III clinical trials, which assess the safety and efficacy of new molecules with increasing numbers of participants. 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). 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]

Manufacturers submit these clinical trial reports for regulatory review. In the United States, the Food and Drug Administration (FDA) is the regulatory body that approves new drugs. For antidepressants, the

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

FDA recommends three to five controlled clinical trials demonstrating substantial evidence of efficacy 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 safety and efficacy, to produce marketing material, and to differentiate the drug against competitors.[9] Publications of clinical trial results provide material for pharmaceutical sales representatives to cite in the promotion of drugs to physicians, a process known as detailing. This material may also be used in medical journal advertisements and direct to consumer advertising, both of which are prevalent in the psychiatric drug market in the United States.

2.2 Psychiatric Clinical Trial Data

The clinical trial data in this paper contain all available double-blind RCTs for either antidepressants or antipsychotics.[10] The antidepressant clinical trial data is based on a comprehensive meta-analysis which includes all trials of 21 antidepressants (Cipriani et al., 2018). This meta-analysis searched clinical trial registries, the websites of regulatory agencies, data from FDA reports, Freedom of Information Act requests and data requested from pharmaceutical companies for all published and unpublished, double-blind RCTs. The included papers span from 1979 through 2015. This sample excludes clinical trials without a comparison, non-double blinded trials, trials with children, and trials for conditions other than major depressive disorder. Leucht et al. (2013) conducted a similar large meta-analysis of antipsychotic clinical trials for 14 antipsychotics from 1969 through 2012. These meta-analyses were multi-year projects of over a dozen authors and effectively contain the universe of all available clinical trials on these drugs. I rely on these meta-analyses to define my sample criteria since many psychiatric clinical trials were published in the 1980s and 1990s before the existence of centralized clinical trial registries.

I obtained the original publications or clinical trial reports for each of these trials, where possible. 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.[11] The final dataset

[9] As an example, Merck ran a post-approval trial for their drug Vioxx. The stated purpose of the trial was to show that Vioxx caused fewer stomach problems than naproxen. Merck's chief scientist characterized the trial as part of "small marketing studies which are intellectually redundant" (Berenson, 2005).

[10] Background on these drug classes is provided in Appendix A1.

[11] Occasionally, the original clinical trial reports contain additional arms that are not included in the meta-analyses. To correctly define the full set of drugs in a trial, I include these additional treatment arms as well. An example is a trial that compared duloxetine, placebo, and a third arm "AZD7268." The trial was supported by AstraZenca, which was developing AZD7268 and that arm would be considered sponsored. The meta-analyses did not include this arm, but it is included in

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.

2.2.1 Defining Terminology

First, I use the term *drug set* to refer 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 paroxetine 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 randomization unit for a randomized control trial. In most cases, each arm in a trial corresponds to a unique drug. In a few cases, a trial may contain several treatment arms with the same drug but different dosages in different arms.

2.2.2 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 or acknowledgements. Any of a drug's manufacturers or marketers in any country are considered sponsors.[12] 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.[13] For each antipsychotic drug, I constructed a list of that

my paper for completeness. In practice, these additions add no new variation as the additional arms all have consistent sponsorship. Removing these additions marginally improves the statistical significance of the drug-by-drug set estimates, since removing arms allows more drugs to be compared within a drug set.

[12]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 5.

[13]In three cases, I revised the Cipriani et al. (2018) sponsorship definitions based on likely errors after reviewing the

drug's global manufacturers and marketers each year. In robustness specifications, I also re-define sponsorship to exclude cases of conflict of interest.

2.2.3 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 antidepressant 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 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). In robustness checks, I also consider the percent decline in these antipsychotic scales, rather than the absolute change.

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.

2.3 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

initial publications. Using exclusively the original coding for antidepressants increases most point estimates and makes no significant difference in my results.

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 is run by the United States National Library of Medicine at the National Institutes of Health and contains the conditions, drugs, interventions, authors, funders, and many trial characteristics for over 300,000 clinical trials. This registry was first available to the public in February of 2000; initially just over a thousand clinical trials were added annually. The registry grew substantially with the ICMJE requirement beginning in 2005 that clinical trials published in any of their affiliated journals had to be registered before patient enrollment. In recent years, ClinicalTrials.gov has been growing by ten to twenty thousand clinical trials annually.

2.4 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, as well as unpublished clinical trials, my analysis sample contain 509 trials and 1,215 treatment arms.[14]

Table 1 presents summary statistics on trial characteristics. Year refers to the year the trial was published. 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 seven percent are considered sponsored due to conflicts of interest alone.[15] Approximately three-quarters of the data are from antidepressant trials and the remaining quarter are from antipsychotic trials. Registered is an indicator for whether the trial was registered and has a National Clinical Trial (NCT) number. Only 12% of my sample was ever registered. Among the full sample, 74% were published or made available after the drug in that arm gained FDA approval.[16]

The average treatment arm enrolled 100 patients and the average trial length was nine weeks. As several different scales are used to measure the severity of major depressive disorder and schizophrenia (see section 2.2.3), the baseline severity of enrolled patients is standardized to have mean zero and standard deviation one within each scale. On average 29% of patients dropped out of each arm before the trial completed. These arms enrolled 51% women on average, and the mean patient was 42 years old. Since my identification strategy uses variation in sponsorship within either a drug set or drug pair, I present summary statistics for the subset of trials with variation in sponsorship separately. These subsets are similar to the full sample in terms of trial characteristics.

[14] See section 2.2.1 for definitions of terminology.

[15] A total of 206 arms (17% of the total) are placebo arms and would never be considered sponsored.

[16] This outcome, as well as "Year relative to FDA Approval" is missing for placebo arms.

Table 1: Summary Statistics: Full and Variation Samples

	Full Sample			Sample with Variation Within: Drug Sets			Drug Pairs		
	Mean	Std Dev.	% Missing	Mean	Std Dev.	% Missing	Mean	Std Dev.	% Missing
Year	2001	8.8	0	1999	7.7	0	2000	8.1	0
Year relative to FDA approval	10.1	10.6	29	11.4	11.4	18	11.8	10.8	24
Share:									
Sponsored	0.48	0.50	0	0.51	0.50	0	0.42	0.49	0
Sponsored w/o COI	0.41	0.49	0	0.40	0.49	0	0.32	0.47	0
Antidepressant	0.74	0.44	0	0.79	0.41	0	0.78	0.41	0
Registered	0.12	0.33	0	0.05	0.21	0	0.09	0.28	0
Post approval	0.83	0.37	21	0.90	0.30	14	0.92	0.27	22
Trial design:									
# of patients	100	86	0	89	101	0	92	91	0
Length (weeks)	9.0	8.0	0	8.6	6.6	0	9.3	8.6	0
Dosage (mg)	69	104	23	59	92	15	58	87	23
Baseline severity	-0.0	1.0	6	0.0	1.0	4	-0.1	1.0	6
% Dropout	29	15	11	29	15	12	30	16	13
Mean age	42	9	16	44	11	15	43	10	16
% Female	51	21	45	51	20	52	50	20	50
Total arms	1,215			453			780		
Total trials	509			208			349		

Notes: This table presents the mean and standard deviation for trial arm characteristics, along with the percent of trial arms with missing values. These 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. Year refers to the year the trial was published; this is also presented relative to the year the drug obtained FDA approval. Sponsored is defined as in section 2.2.2. Details for each trial design outcome are listed in section 1.

In total, the analysis contains 37 unique drugs, 24 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. Older drugs are sponsored the least often in my analysis sample. These drugs often no longer have patent protection during my sample, so the manufacturer has weaker incentives to produce new marketing material. These older drugs might also no longer be the comparison standard of treatment against which new drugs are tested.

3 The Effect of Sponsorship: Empirical Strategy and Results

3.1 Description of Sponsorship Variation

The main types of drug combinations are presented graphically in table 2. The final clinical trial data contain 41 unique drug sets with variation in sponsorship. There are 207 total unique drug sets in my sample, but in the other drug sets, each drug is either always sponsored or always unsponsored. These 41 drug sets with variation in sponsorship contain 208 trials and 453 treatment arms. Turning to the subsample with variation within drug pairs, there are 59 drug pairs with variation in sponsorship, which corresponds to 349 unique trials and 780 unique treatment arms.

The first category ("Active vs. Placebo") directly compares a psychiatric drug ("drug A") to a placebo only. Among the sample with variation in sponsorship, this category contains 5 drug sets, 24% of the trials, and 24% of the treatment arms. Some of these trials are sponsored by the company that manufactures drug A ("company A").[17] The other unsponsored trials have alternative funding not provided by company A.

The second category in table 2 ("Active vs. Active") contains drug sets that compare an active drug to another active drug. This contains 33 drug sets, 70% of the trials, and 66% of the treatment arms. In all cases, "drug A" varies in funding interests. There are three main subgroups considered. 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. The last category contains drug sets with three drugs.

This analysis includes both variation within industry as well as industry to non-industry comparisons. Drugs can be included in clinical trials as a control group in an alternate pharmaceutical firm's analysis, as in the example in the introduction. In the latter two subgroups in the "Active vs Active" category in table 2, multiple pharmaceutical firms with different financial interests are conducting clinical trials on the same sets of drugs. Some clinical trials are also funded by government and academic organizations. For example, my paper includes trials from the National Institutes of Health and international organizations such as the S*ã*o Paulo Research Foundation.

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. 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

[17] These trials could have additional funders; it is sufficient that company A is affiliated with the trial in some capacity.

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.[18] While on average sponsored arms occur earlier in a drug's life cycle than non-sponsored arms, systematic differences in the age of a drug are not a primary contributor to the sponsorship effect, as the efficacy of a drug does not substantially change in the years after FDA approval (see appendix figure B3). Additional robustness that considers the age of drugs is also shown in section 3.6.

Table 2: Types of Variation

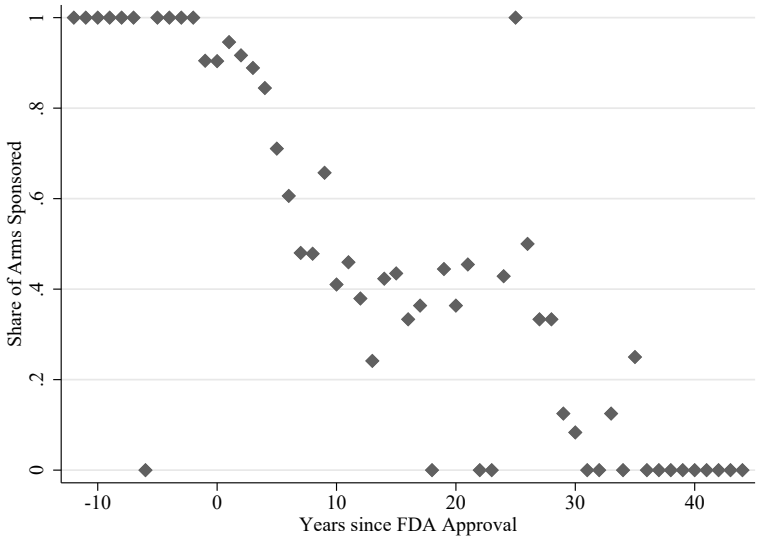
	Variation within:					
	Drug Set			Drug Pair		
	Drug Sets	Trials	Arms	Drug Pairs	Trials	Arms
Active vs. Placebo						
Company A: **Drug A** vs. Placebo ⟷ Unsponsored: Drug A vs. Placebo	5	49	108	12	175	406
Active vs. Active						
One Drug Never Sponsored						
Company A: **Drug A** vs. Drug B ⟷ Unsponsored: Drug A vs. Drug B	10	41	85	10	37	79
One Drug Always Sponsored						
Company A & Company B: **Drug A** vs. **Drug B** ⟷ Company B: Drug A vs. **Drug B**	14	65	131	17	83	175
Both Drugs Vary in Sponsorship						
Company A: **Drug A** vs. Drug B ⟷ Company B: Drug A vs. **Drug B**	9	40	84	20	86	190
Subtotal	33	146	300	47	203	440
Three or More Drugs						
Company A: **Drug A** vs. Drug B vs. Placebo ⟷ Unsponsored: Drug A vs. Drug B vs. Placebo	3	13	45	0	0	0
Total	41	208	453	59	349	780

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, and the number of treatment arms, which is a unique randomization arm of a trial. See section 2.2.1 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.

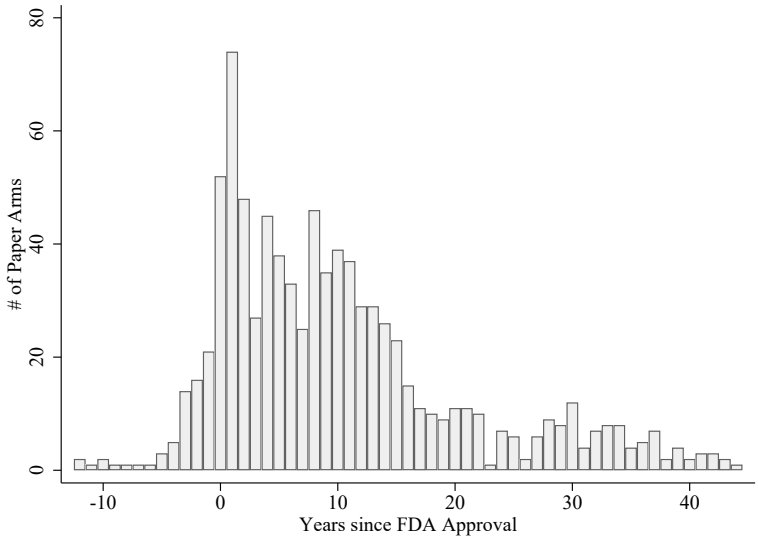
[18] In contrast, the share of arms sponsored by calendar year has remained fairly constant within my sample (see appendix figure B2).

Figure 1: Variation in Sponsorship by Year Relative to Drug Approval

(a) Sponsorship



(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.

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 outcome y_{ij} is computed *relative* to the placebo arm in trial j, if available, or least effective arm, otherwise.[19] 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.[20]

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. Each trial j can be mapped to a single drug set $s(j)$. As described in section 2.2.1, a drug set is the unique combination of drugs in a clinical trial. In this case, for example, 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 trials comparing the exact same set of drugs. Appendix table B1, column (1), provides a more detailed example of this fixed effects structure. 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.

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 $G_{d(i),p(j)}$ in equation 2 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 $p(j)$. In this case, 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. Conceptually, this specification assumes that the presence of an additional arm should not affect the comparison between an existing drug pair. One technical point regarding this fixed effect structure is that a trial with e.g. three unique drugs will contain three drug pairs. Therefore, each arm in that trial will be counted in two separate drug pairs.[21] Therefore, I re-weight the observations so that each treatment

[19] The effectiveness of an arm within a clinical trial is usually stated relative to the other arms in the trial. 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. Conceptually, this is similar to adding trial fixed effects. Appendix table B5, panel A, includes results for non-relative outcomes.

[20] As described in section 2.2.3, some trials report efficacy using alternative 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.

[21] In the trials with n treatment arms, each drug will be counted in $n-1$ drug pairs. Thus each treatment arm is weighted

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. Panel A contains all antidepressant drug sets that compare an active drug to a placebo and have variation in sponsorship. Panel B presents a sample of the five most common drug sets that compare an active drug to another active drug.[22] Each row is a unique drug set and, in my initial empirical specification, each drug in each row would receive its own fixed effect.

Table 3: Difference in Difference

Panel A: Active versus Placebo									
	Sponsored				Not Sponsored				
	Drug	Placebo	*Diff*	# Arms	Drug	Placebo	*Diff*	# Arms	DD
All Drug Sets	0.491	0.290	0.201	51	0.441	0.301	0.140	8	0.061
Paroxetine	0.465	0.305	0.160	29	0.250	0.226	0.024	1	0.137
Sertraline	0.460	0.361	0.099	11	0.476	0.433	0.042	2	0.057
Citalopram	0.458	0.158	0.300	6	0.568	0.353	0.215	1	0.085
Trazodone	0.509	0.350	0.160	4	0.303	0.209	0.095	1	0.065
Amitriptyline	0.564	0.278	0.286	1	0.607	0.282	0.325	3	-0.039
Panel B: Active versus Active									
	Sponsored				Not Sponsored				
	Drug	Other Arm	*Diff*	# Arms	Drug	Other Arm	*Diff*	# Arms	DD
All Drug Sets	0.647	0.597	0.049	50	0.567	0.583	-0.016	60	0.066
Amitriptyline vs. Fluoxetine	0.653	0.564	0.088	3	0.500	0.522	-0.022	10	0.111
Amitriptyline vs. Paroxetine	0.658	0.648	0.010	1	0.466	0.473	-0.008	8	0.017
Citalopram vs. Escitalopram	0.794	0.815	-0.021	6	0.639	0.760	-0.120	3	0.099
Fluoxetine vs. Venlafaxine	0.764	0.745	0.018	1	0.613	0.687	-0.074	7	0.092
Venlafaxine vs. Fluoxetine	0.687	0.613	0.074	7	0.704	0.707	-0.003	1	0.077
...									

Notes: This table presents the difference-in-difference estimate of the sponsorship effect for "Active versus Placebo" drug sets (panel A) and "Active versus Active" drug sets (panel B) . 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 (or 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.

In the first row, I consider trials that directly compare paroxetine to a placebo. There are 30 such

by $\frac{1}{n-1}$, where n is the number of treatment arms in the trial.

[22] The first category corresponds to the "Active vs Placebo" row in table 2, while the second category corresponds to the "Active vs Active" row in the same table.

trials; 29 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 31% of patients respond to the placebo. Therefore, on average, paroxetine is 16 percentage points more effective than the placebo. Turning to trial in which paroxetine is 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 14 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 6.1 percentage points (table 3, row 1).[23]

Panel B presents the analogous estimates for a subset of 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 fluoxetine. If amitriptyline is sponsored and fluoxetine is not, an average of 65% of patients respond to amitriptyline, while 56% of patients respond to fluoxetine, an average of 8.8 percentage points higher efficacy for amitriptyline. In the eight trials where neither amitriptyline nor fluoxetine were sponsored, an average of 50% of patients responded to amitriptyline, while 52% responded to fluoxetine, an average of 2 percentage points lower efficacy for amitriptyline. Thus, the sponsorship effect, i.e. the "difference in difference," is 11 percentage points. Averaging across all antidepressant drug sets in this category, and weighting by the number of trials, the average sponsorship effect is 6.6 percentage points. The full set of antidepressant drug sets is presented in appendix table B2; the antipsychotic trials are shown in appendix table B3.

The coefficient on sponsorship in equations 1 and 2 is analogous to the average of the difference-in-difference values in tables 3, appendix table B2, and appendix table B3, weighted by the number of arms in each estimate.

3.4 The Effect of Sponsorship on Reported Efficacy

Table 4 presents the regression estimates from equations 1 and 2. In column (1a), I find that a sponsored drug is 0.16 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

[23] This table does not contain standard errors, since some of the categories only have a single observation. The regression specification presents standard errors for very similar estimates.

[24] If both drugs vary in sponsorship, they are included as two separate entries, as in the last two rows. In the regression specifications, each trial is given the same weight.

no difference in this result. The sponsorship effect is 36% of the average difference in efficacy of 0.45 standard deviations between a treatment arm and the placebo arm. Therefore, the funding interests of a given drug can explain a third of the relative efficacy of that drug.

Table 4: 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)
$Sponsor_{ij}$	0.159*	0.162**	0.118***	0.071	0.152*
	(0.082)	(0.081)	(0.045)	(0.045)	(0.091)
Controls		X	X	X	X
Drug by Drug Set F.E.	X	X	X	X	X
Mean Outcome	0.45	0.45	0.24	0.28	0.39
N	1,215	1,215	1,087	1,087	1,215
Panel B: Drug Pair	Standardized Outcome (Relative)		Significant (5%)	Significant (10%)	Most Effective Arm
	(1b)	(2b)	(3b)	(4b)	(5b)
$Sponsor_{ij}$	0.169***	0.155***	0.107***	0.107***	0.269***
	(0.052)	(0.050)	(0.040)	(0.040)	(0.054)
Controls		X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X
Mean Outcome	0.36	0.36	0.24	0.28	0.39
N	1,990	1,990	1,741	1,741	1,990
Weighted N	1,215	1,215	1,087	1,087	1,215

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 $^{***}p < 0.01$.

In column (3a), the outcome 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 section A2 provides

details on the construction of this variable. On average, sponsored arms are 12 percentage points more likely to be statistically significant at the 5% level. This represents a 49% increase over the baseline of 24% statistical significance. As described in section , 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 0.15 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 39% 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. In this specification, trials with differences in sponsorship in any of the occurrences of a drug pair can be used to identify the sponsorship effect.[26] My preferred specifications are columns (2a) and (2b), which use the relative standardized outcome and control for the measurement scale and calendar year. In column (2b), the effect of sponsorship on reported drug efficacy is 0.16 standard deviations, or 43% of the average relative efficacy.

3.5 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

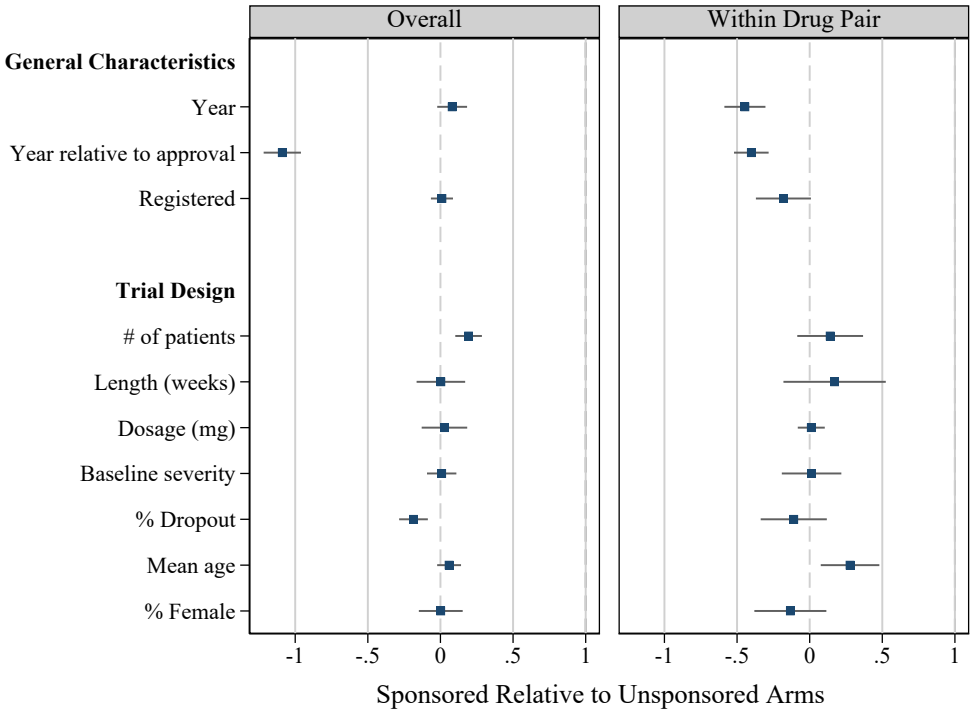
$$c_{ij} = \alpha + \beta Sponsor_{ij} + \varepsilon_{ij} \tag{3}$$

[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.

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. 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 differences in characteristics between sponsored and unsponsored arms. Sponsored arms occur in trials 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.

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 3.5. Error bars represent 95 percent confidence intervals. Standard errors are clustered at the trial level. Details for each outcome are listed in section 2.1.

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 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. Within drug pairs, sponsored arms occur only 0.4 standard deviations or

about four years earlier. Similarly, while in panel (a) sponsored arms enroll 0.2 standard deviation or 15 more patients per arm, 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 0.19 standard deviation smaller dropout rate, while within drug pairs, the difference in dropout rates is statistically insignificant and lowered to -0.11 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 B3 plots the average efficacy of drugs relative to their approval year. After approval, drug efficacy appears stable over time. 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 obtained unexpectedly high efficacy draws and therefore were able to gain FDA approval. After approval, their mean efficacy decreases to match their true efficacy. The next section assesses the importance of mean reversion.

3.6 Robustness

Table 5 accounts for any systematic changes in efficacy over the drug's life cycle. Column (2a) and (2b) control for the publication order of the trial within the drug set. In both, controlling for the publication order slightly decreases 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 in magnitude, though the estimates vary in significance. 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 5, column (4a) and (4b). The point estimates decrease slightly but the estimate of the sponsorship effect remains statistically significant. In all cases, the sponsorship effect is similar, suggesting that mean reversion cannot explain most of the sponsorship effect.

As described in section 2.2.2, 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 slightly, suggesting that conflict of interest is a relevant component of sponsorship.

Table 5: 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	Sponsor w/o COI	Weight by Enrollment
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
$Sponsor_{ij}$	0.162**	0.146*	0.130	0.147*	0.132	0.131**
	(0.081)	(0.083)	(0.090)	(0.084)	(0.089)	(0.066)
Controls	X	X	X	X	X	X
Drug by Drug Set F.E.	X	X	X	X	X	X
Mean Outcome	0.45	0.45	0.45	0.55	0.45	0.40
N	1,215	1,215	1,215	795	1,215	1,215

Panel B: Drug Pair

	Baseline	Control for Trial Order	Control for Year Relative to Approval	Restrict to Post Approval	Sponsor w/o COI	Weight by Enrollment
	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
$Sponsor_{ij}$	0.155***	0.143***	0.119**	0.129**	0.138**	0.088**
	(0.050)	(0.050)	(0.050)	(0.052)	(0.053)	(0.039)
Controls	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X
Mean Outcome	0.36	0.36	0.36	0.43	0.36	0.30
Weighted N	1,215	1,215	1,215	795	1,990	1,215

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 4, 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 $^{***}p < 0.01$.

Finally, I consider robustness to alternative weighting schemes. My analysis weights each treatment

arm equally, as the conceptual counterfactual involves changing the funding for a drug within a clinical trial. However, an alternate counterfactual may randomize funding of drugs at the patient level. This weighting may also correspond to physicians interpreting 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) and (6b)). The drug-by-drug set fixed effect estimates are very similar to the baseline estimates in direction and significance, 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.

The effect of funding could differ by drug class. To explore this distinction, appendix table B4 separates the analysis by the class of drug—antidepressant or antipsychotic. The sponsorship effect for only antidepressants is similar in magnitude to the combined baseline specification. I also present estimates using alternate outcomes, which are similarly positive and statistically significant. The sponsorship effect is smaller in the antipsychotic subsample. In addition, antipsychotics are a small share of the analysis sample, so results within this subset are not statistically significant.

Prior literature has used either only drug fixed effects or no fixed effects to estimate the sponsorship effect. For completeness, appendix table B5 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.

3.7 Sponsorship Effect and Financial Incentives

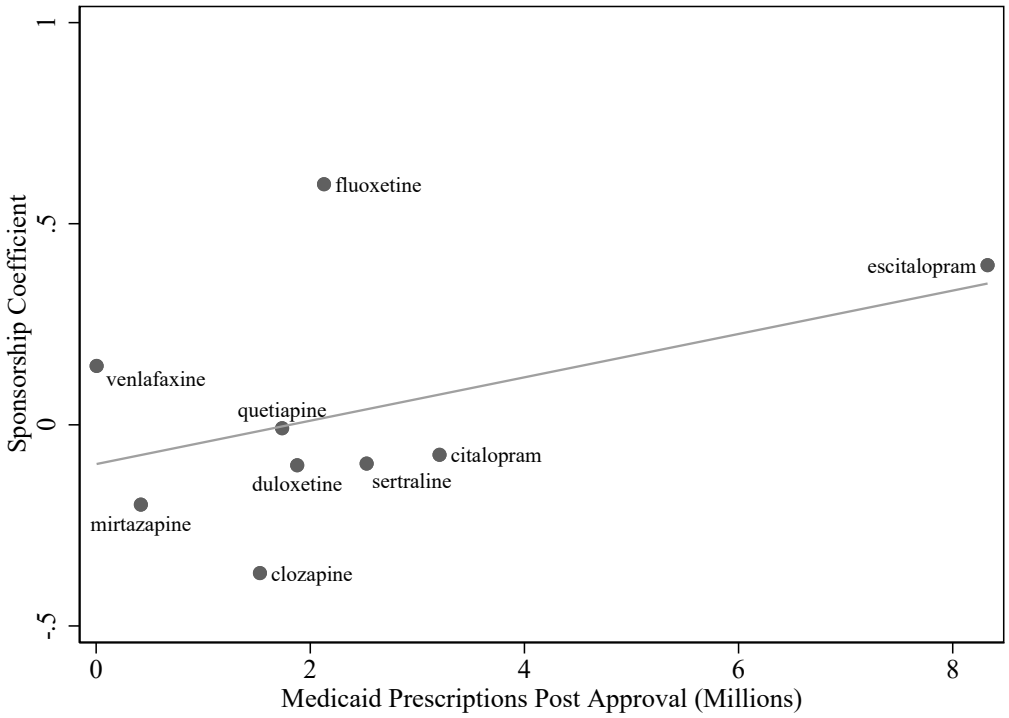
The sponsorship effect may be related to the financial incentives of pharmaceutical firms. If the potential market for a given drug is larger, due to higher patient demand or fewer competitors within a subclass, there might be additional incentives to obtain higher reported efficacy for that drug. To assess this correlation, I plot the sponsorship effect for each drug relative to its market size. I compute the drug-specific sponsorship effect by estimating:

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

[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 B5, panel C.

where $d(i)$ is an indicator for each drug. Each term is the same as equation 2, but the $Sponsor_{ij}$ indicator is now interacted with each drug separately. Figure 3 plots the coefficients for each drug against a proxy for market size: the total Medicaid prescriptions in the five years after FDA approval for that drug. The positive relationship could be driven either by high projected sales incentivizing a high sponsorship effect or a high sponsorship effect driving higher sales. In either case, the positive correlation between the sponsorship effect and prescriptions shows the sponsorship effect is related to market factors.

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.

3.8 Heterogeneity by Variation Type

My estimates use variation both within industry and across industry vs non-industry trials. As shown in Table 2, the most common type of variation involves comparing one active drug that varies in sponsorship with an active drug that is always sponsored. In this case, I am comparing funding incentives using variation only within industry. In other cases, such as in trials comparing an active drug to a placebo, I am comparing industry-funded trials to unsponsored trials. Table B6 shows that the sponsorship effect is similar across these two types of variation. Within active versus placebo comparisons, the sponsorship effect is 0.225 (column 2b) and statistically significant. Using only within industry variation in column

(4b), the sponsorship effect is 0.318, which is also statistically significant.

3.9 Which Drug Trials Have Variation in Sponsorship?

The identification is driven by the subset of drug sets or pairs that have variation in sponsorship. To assess which types of trials have variation in sponsorship, table 6 plots the share of arms that have variation in funding within drug pairs. 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 prescribed class of antidepressant. Drugs with a larger market share might be more likely to be included in marketing trials. Placebo arms are mechanically less likely to be part of a drug pair with variation in funding since the placebo arm is always unsponsored.

Table 6: Sponsorship Variation by Characteristics

	# Arms	Share with Variation
Full Sample	1,215	0.64
Drug Type - Antidepressants		
Tricyclic	67	0.88
Atypical	160	0.28
SSRI	333	0.79
SNRI	124	0.60
Drug Type - Antipsychotics		
1st Gen	52	0.75
2nd Gen	201	0.63
Placebo	260	0.67
Approval Year		
Prior to 1990	278	0.77
1990 - 1996	305	0.69
1997 or after	231	0.50
Patent Expiry Year		
Prior to 2000	167	0.80
2000 - 2007	395	0.71
2008 or after	311	0.59

Note: This table presents the share of arms with each characteristic that have variation in sponsorship. In this table, variation in sponsorship is defined at the drug pair level.

The strongest predictor of variation in sponsorship is the age of a drug. Drugs that were approved in earlier years or already had their patents expire are the most likely to have variation in funding. Drugs

that were approved later have less time to be included in different trials. Drugs are also more likely to be included as control arms in other industry trials once their patent expires.[28]

The subsample of trials with variation in sponsorship might also depend on the results of previous trials. Firms might be more likely to test drugs that had higher efficacy in previous trials. Appendix figure B5 shows 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. The set of trials with variation in sponsorship are potentially more effective and more commonly prescribed than the rest of the sample. However, most drug pairs have variation in sponsorship, and those without will potentially acquire variation in future years.

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). The second class of mechanisms covers all cases that occur after data collection (i.e. ex-post mechanisms).

4.1 Trial Design

Trial design and patient selection can substantially affect reported efficacy for psychiatric medications.[29] To test whether these characteristics systematically explain the sponsorship effect, I first assess whether controlling for these characteristics affects the estimates. The first column in Table 7 replicates the baseline estimates. Controlling for trial characteristics (total enrollment, length of trial, and dosage) increases the point estimate, though the results remain very similar to the baseline. Controlling for patient characteristics (mean age, share female, baseline severity, and dropout share) slightly decreases the point estimate, though again these results are similar to the baseline. With the full set of controls, the estimate is 0.146, statistically insignificantly different from the baseline of 0.155. These results are consistent with evidence from figure 2, which showed that sponsored arms are not different in most

[28] This pattern is also shown in appendix figure B4, which presents the network of comparisons between drugs. One of the best predictors of variation in sponsorship is the generic entry year. Among the drugs with earlier generic entrants, most drug pairs have variation in sponsorship (marked by solid maroon lines). Among the drugs which do not yet have generic entrants, none of the drug pairs have variation in sponsorship (marked by dashed gray lines).

[29] 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 antidepressant drug 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).

observable characteristics.

Table 7: Sponsorship Effect with Trial and Patient Characteristic Controls

Drug Pair	Baseline	Trial Characteristics	Patient Characteristics	All
	(1)	(2)	(3)	(4)
$Sponsor_{ij}$	0.155***	0.166***	0.137***	0.146***
	(0.050)	(0.050)	(0.050)	(0.050)
Controls	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X
Mean Outcome	0.36	0.36	0.36	0.36
$Weighted\,N$	1,215	1,215	1,215	1,215

Note: Panel A presents the coefficients on $Sponsor_{ij}$ from the estimation of equation 2, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. Column (1) replicates the main result from table 4, columns (2a) and (2b). Column (2) includes controls for trial characteristics: the length of the trial in weeks, the number of patients, and the initial dosage. Column (3) includes controls for patient characteristics: the mean age, share female, baseline severity, and dropout share. Missing values for these characteristics are imputed as the mean value for each characteristic. Column (4) includes both sets of controls. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

While controlling for trial and patient characteristics does not change the sponsorship effect, these characteristics may 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, controlling for the share female would not affect the overall 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 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:

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

where the terms were defined in section 3.2. The coefficient on $Sponsor_{ij}$ can now 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 8 shows these results separately by each characteristic. Sponsored arms do not have higher predicted efficacy for any individual characteristic. 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, this predicted sponsorship effect is not statistically significant. I also combine all covariates in one prediction, using LASSO to select the most predictive characteristics.[30] As shown in table 8, column (8), sponsored arms are not predicted to have higher relative efficacy based on the full set of observable characteristics.

Table 8: Predicted Sponsorship Effect Using Individual Characteristics

	Trial Characteristics			Patient Characteristics				All
	N	Length	Dose	Baseline Severity	Dropout Rate	Age	Gender	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Sponsor_{ij}$	-0.01	-0.01	0.02	-0.03	0.04	0.00	-0.02	0.00
	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.04)
Controls	X	X	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X	X	X
Predicted R^2	0.22	0.26	0.20	0.13	0.35	0.33	0.32	0.22
Mean Outcome	0.21	0.24	0.23	0.11	0.26	0.30	0.29	0.33
Weighted N	1,215	1,215	1,215	1,215	1,215	1,215	1,215	1,215

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, or all trial and patient characteristics (column 8). Missing values for these characteristics are imputed as the mean value for each characteristic. Controls include the trial's publication year and the type of psychiatric score used. Standard errors are reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

I conclude that the observable characteristics of trial design and patient enrollment do not explain the sponsorship effect. Differential trial design might be less prevalent in psychiatric drugs because identifying characteristics that are favorable for psychiatric medications is difficult. An important caveat of my analysis is there are many characteristics of trial design not included in these observable characteristics,

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

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.

4.2 Publication Bias

4.2.1 General Tests for Publication Bias

The results of clinical trials may affect their publication outcomes. For example, 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). This paper assesses whether publication bias differs for trials by funding source.

First, I test whether sponsored arms are more likely to be published if they have higher reported efficacy, compared to unsponsored arms. As noted in section 2.2, I observe data on approximately a hundred unpublished clinical trials. These unpublished trials are a subset of the universe of all unpublished trials ever conducted. Most unpublished clinical trials are never made available.

The unconditional relationship between reported efficacy and the share of arms published is presented in figure 4. Among a combination of my analysis sample and the observed unpublished papers, 86% of arms are published. The publication share remains high among arms with low relative efficacy, suggesting that there are journal outlets for null results. Among non-sponsored arms, efficacy is weakly positively related to the share of arms published. As predicted, the relationship between efficacy and publication status is much stronger among sponsored arms, shown in red.

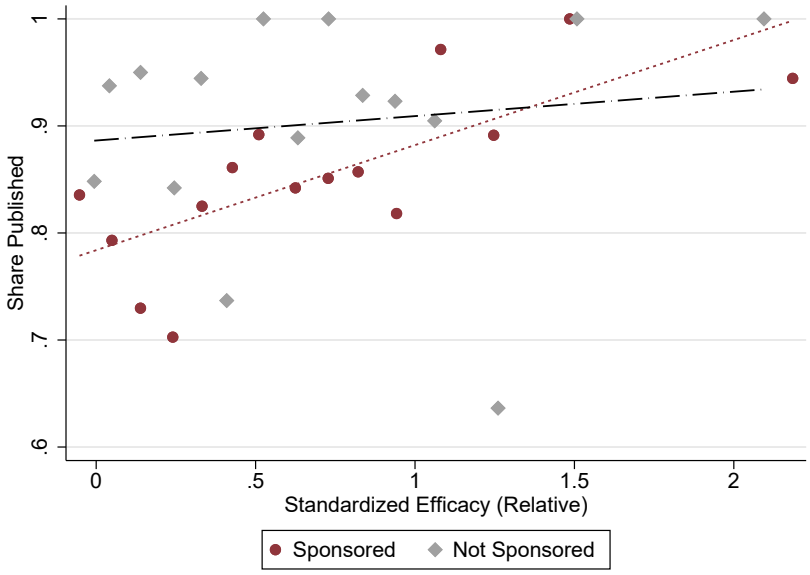
Appendix table B7 shows these results hold within a drug pair. Specifically, I estimate:

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

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 4, column (2b). Sponsored arms are statistically significantly less likely to be published overall, but sponsored arms with higher efficacy are *more* likely to be published, which corroborates the results from figure 4.[31]

[31] Another standard test for publication bias is to measure the level of bunching around z-score cutoffs (Brodeur et al., 2016). Appendix figure B6 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. These are phase III or phase IV clinical trials and are conducted based on the results of earlier stage trials. Therefore, estimating the underlying distribution of results requires additional assumptions and I do not implement the Andrews and Kasy (2019) procedure for adjusting for publication bias.

Figure 4: The Relationship Between Efficacy and Publication



Notes: This figure presents the relationship between effectiveness and publication status. The x-axis plots the standardized efficacy of an arm, relative to the placebo or least effective arm in that trial. Efficacy is binned into ventiles separately for sponsored and non-sponsored arms. The y-axis presents the probability that arms in the given efficacy bin are published.

4.2.2 Magnitude of Publication Bias

To determine the share of the sponsorship effect explained by publication bias, I estimate how the sponsorship effect would change if I observed data from the full set of trials ever conducted. In recent years and particularly after the mid-2000s, pre-registration requirements have been enforced, and authors are required to pre-register their clinical trial before patient enrollment.[32] Therefore, to get an approximation of the full set of clinical trials, I rely on the set of pre-registered clinical trials in recent years.

Pre-registered trial data are taken from ClinicalTrials.gov, as described in section 2.3.[33] To assess the role of publication bias, I restrict the ClinicalTrials.gov sample to trials that were submitted to the registry after 2005 and before 2011. This latter restriction allows five years for the antidepressant trials to potentially be published. I also restrict to trials that assess the conditions of major depressive disorder or depression and that tested at least one of the antidepressant drugs included in my sample. Further, I restrict to trials with a stated purpose of either treatment or basic science, with randomized allocation and parallel treatment assignment, and enrollment restricted to depressed patients. Finally, I excluded trials with children, trials that assessed chronically depressed patients, and trials that assessed only a single drug without a placebo or alternate treatment arm. These inclusion criteria are all outlined in

[32] See section 5 for more information.

[33] This analysis is restricted to antidepressant trials since the inclusion criteria in the antidepressant meta-analysis corresponds closely to Clinicaltrial.gov variables.

Cipriani et al. (2018). This criteria successfully includes over 90% of the trials in my analysis sample that were linked to ClinicalTrials.gov.[34] Equally importantly, only 1% of included trials had results that were not in my analysis sample.[35]

Out of the 163 pre-registered trials that fit this criteria, my clinical trial data contain results for just 25% of these trials. Therefore, I estimate that there exist four times more unpublished trials in the universe of all clinical trials than I observe in my analysis sample. This estimate is consistent with evidence that only 22% of pre-registered drug trials reported results (Prayle et al., 2012). Therefore, I create a counterfactual sponsorship effect under the assumption that I observe four times more unpublished trials. These additional unpublished trials are randomly drawn from my set of existing unpublished trials.

Using this estimate to benchmark the share of the sponsorship effect explained by publication bias relies on two strong assumptions. First, I assume that the sponsorship effect among the unpublished trials observed in my analysis sample has the same magnitude as among all unpublished clinical trials. Second, I assume that the clinical trial registry contains the full universe of trials conducted during 2005-2011. To the extent that the unpublished trials I observe are more favorable to funders than the unpublished trials with unobserved results, or that the clinical trials registry undercounts trials, this estimate would underestimate the true share of the sponsorship effect explained by publication bias. Both caveats are likely true, so my estimate of the share of the sponsorship effect explained by publication bias should be considered a lower bound. Figure 5 accommodates alternative assumptions about the magnitude of publication bias. Under my preferred assumption that there are four times more unpublished trials, the sponsorship effect would fall from 0.155 to 0.082, or just under 50%. The 95% confidence interval for the sponsorship effect with four times more unpublished trials ranges from 0.00 to 0.164.

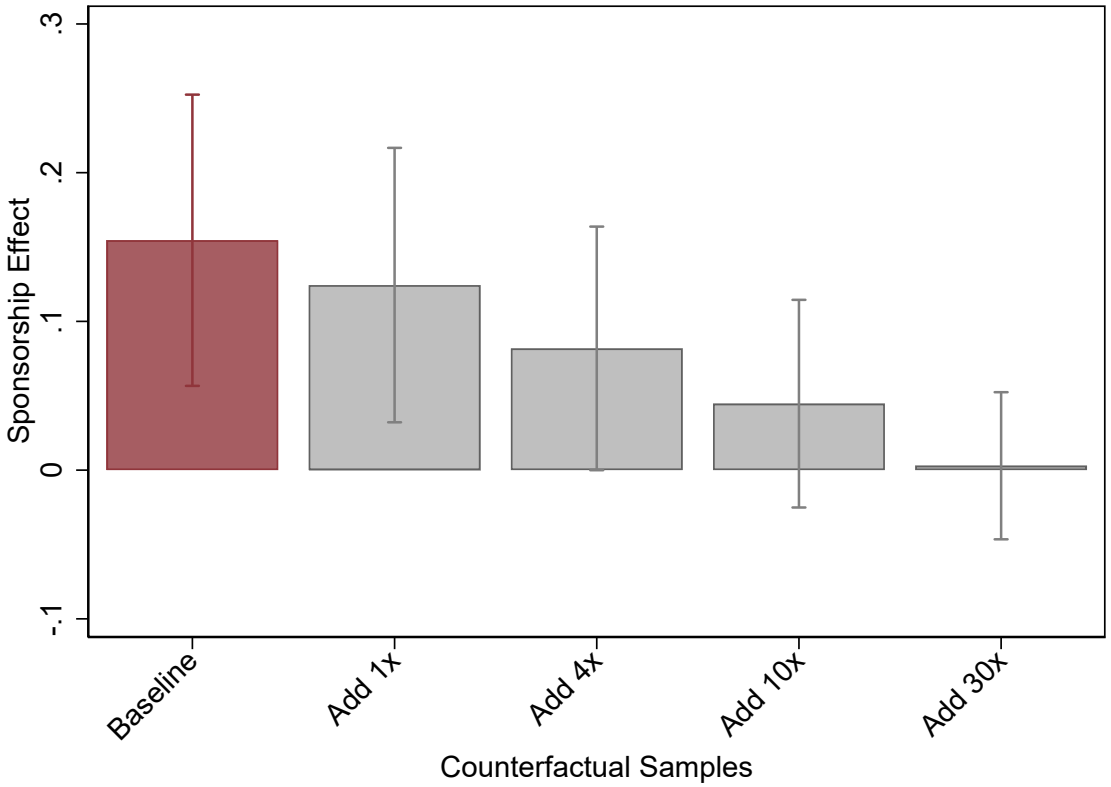
In contrast, the point estimate for the share of sponsorship effect explained by trial design from table 8, column (8) is just 0.002 off of a base of 0.155.[36] The remaining unexplained share of the sponsorship effect could be due to underestimating the publication channels described above, noise in these estimates, selection on aspects of trial design that are unobserved, or data manipulation or reconciliation errors.

[34] Specifically, this criterion includes 64 of the 71 trials in my analysis sample that were linked to ClinicalTrials.gov. Of the seven trials in my analysis sample that were excluded, one trial was categorized by the registry as related to cognition, two referred to the drugs by their development codes rather than generic names, two did not list the allocation as random, one stated they included children, and one stated they enrolled healthy patients rather than depressed patients. In all cases, the contents of these trials fit the inclusion criteria above but the ClinicalTrials.gov labels were incorrect.

[35] Specifically, of the 314 trials with this inclusion criteria that were not included in the analysis sample, only four had available results or publications. In each of these cases, the primary outcomes in these trials were not assessing depression symptoms and thus they were correctly not a part of the analysis sample.

[36] However, the 95% confidence interval means I can only rule out that trial design explains more than 51% of the sponsorship effect.

Figure 5: 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 in solid maroon presents the baseline estimates including only published trials, replicating replicates table 4, column (7). The second bar presents estimates including each unpublished trial once. Subsequent columns include additional unpublished trials as described in section 4.2.2. 95% confidences intervals are presented as lines on each bar graph. Standard errors are clustered at the trial level.

5 Mitigation and Pre-Registration

One major policy lever in regulating clinical trials is pre-registration requirements. These require that investigators register the existence of a clinical trial as a condition of publication or funding. Requirements often include pre-specifiying outcomes, reporting results, and pre-registration prior to patient enrollment. Arguably the most significant of these requirements is the ICMJE's agreement to only publish clinical trials in affiliated journals that were registered before patient enrollment. This condition applied to all trials that began on or after July 1st, 2005; trials that began earlier had to be adequately registered before journal submission as of September 13, 2005 (De Angelis et al., 2004).[37]

[37] A later requirement is Section 801 of the Food and Drugs Amendements Act. This was initially passed in 2007 and compliance was required as of April 18, 2017. This law requires applicable clinical trials to register no later than 21 days after the enrollment of the first participant and to report their outcomes no later than one year after the primary completion date. Compliance rates are estimated to be below 50%, and no fines have ever been levied (Piller, 2020).

If the sponsorship effect is largely due to publication bias, then pre-registration and outcome reporting requirements would expand the availability of clinical trial results and mitigate these effects. To test whether pre-registration changed the sponsorship effect, I estimate the following specification:

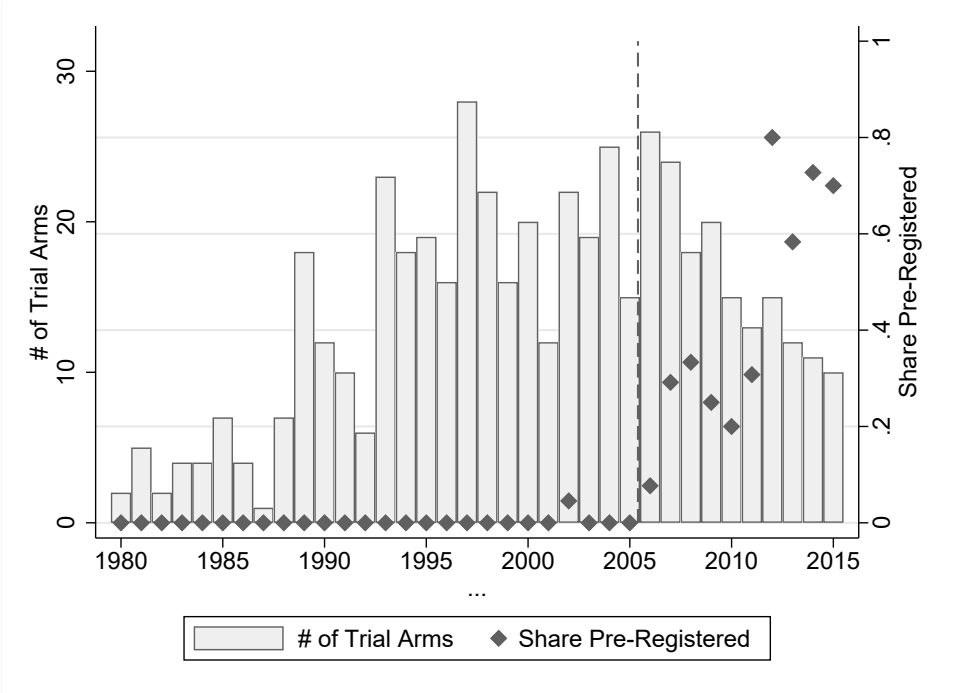
$$y_{ij} = \alpha + Sponsor_{ij} + \sum_{y} \beta_y Sponsor_{ij} * y(j) + \sum_{y} y(j) + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{9}$$

where the sponsorship effect is interacted with publication year bins $y(j)$. The controls X_{ij} are indicators for the measurement scale and all other terms are the same as in equation 2. 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 as shown in figure 6a. Accordingly, figure 6b plots the coefficients β_y on the sponsorship effect over time. The coefficients decrease in magnitude gradually after the 2005 pre-registration requirements. The effect of sponsorship on reported drug efficacy is statistically significant before required pre-registration and decreases after required pre-registration (see table 9, column (2)). This difference in the effect of sponsorship before relative to after required pre-registration is statistically significant. In addition, among the set of trials that were not pre-registered, the effect of sponsorship remains large after required pre-registration, though this estimate has large standard errors (appendix figure B7).

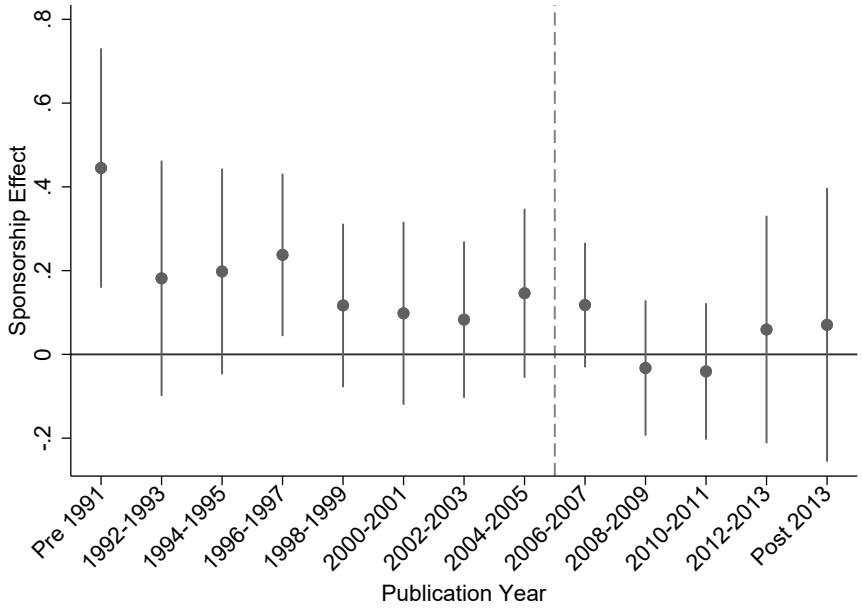
If pre-registration were effective at mitigating the sponsorship effect, then the sponsorship effect should be smaller among trials that have been pre-registered. Table 9 column (3) assesses this relationship and, accordingly, finds that the sponsorship effect among trials pre-registered on ClinicalTrials.gov is -0.172 standard deviations, relative to sponsored but not pre-registered trials. Relative to the baseline sponsorship effect of 0.135 among trials that were not pre-registered, the point estimate for the sponsorship effect among pre-registered trials is below zero. The evidence that the sponsorship effect decreases with the advent of required pre-registration is suggestive that pre-registration may be effective at mitigating conflicts of interest and publication bias.

Figure 6: Introduction of Clinical Trial Pre-registration

(a) Pre-Registration by Calendar Year



(b) Sponsorship Effect by Calendar Year



Notes: Panel A plots the share of antidepressant trials in my sample that were pre-registered in ClinicialTrials.gov by publication year. The gray bars plot the sample size of treatment arms by publication year. 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. Panel B presents the coefficients β_y from the estimation of equation 9. Standard errors are clustered at the trial level.

Table 9: Publication by Pre-Registration

	Standardized Outcome (Relative)		
	(1)	(2)	(3)
$Sponsor_{ij}$	0.155***	0.202***	0.171***
	(0.050)	(0.058)	(0.052)
Post 2005		-0.114***	
		(0.141)	
$Sponsor_{ij}$ x Post 2005		-0.150**	
		(0.066)	
Pre-Registered on ClinicalTrials.gov			0.015
			(0.042)
$Sponsor_{ij}$ x Pre-Registered on ClinicalTrials.gov			-0.172*
			(0.102)
Controls	X	X	X
Drug by Drug Pair F.E.	X	X	X
Mean Outcome	0.34	0.34	0.34
Weighted N	1,215	1,215	1,215

Note: Table presents the coefficients from the estimation of equation 2 with $Sponsor_{Ij}$ interacted with an indicator for after 2005 or an indicator for whether the trial was pre-registered . Column (1) presents the coefficient on $Sponsor_{ij}$, excluding the interaction terms. Column (2) presents the coefficients on $Sponsor_{ij}$ interacted with an indicator for whether the trial was published after 2005. Column (3) presents the coefficients on $Sponsor_{ij}$ interacted with an indicator for whether the trial was pre-registered on ClinicalTrials.gov. 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 $^{***}p < 0.01$.

6 Conclusion

This paper provides empirical evidence that financial incentives affect the reported 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 my 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, noise in the estimates, 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 0.15 standard deviations, then the approval rate would fall from 79% 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 0.52 million fewer Medicaid prescriptions per drug, or a 21% decrease.

This paper also finds that a major policy change regarding clinical trials – required pre-registration as a condition for publication – coincides with a statistically significant decrease in the effect of sponsorship on reported drug effectiveness. This suggests that pre-registration may be beneficial at reducing the effect of trial sponsorship. There are also non-negligible costs of pre-registration such as the investigator's time, unforeseen changes to the analysis, or unexpected circumstances. As the modal clinical trial in my sample involves numerous investigators, reports standard outcome differences across treatment arms and takes place over the course of weeks, the extent that investigators must respond to changing methodological or environmental concerns is limited. In addition, the time costs of registration should be compared with the large stakes of a clinical trial, suggesting that pre-registration may be more worthwhile for clinical trials relative to other fields.

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 these 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 alternative drug classes. Classes which also have numerous substitutable drugs and variation in sponsorship could be viable candidates.[38] 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.

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. 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 how these restrictions would limit the total amount of innovative research. Alternate funding schemes should also consider how sponsored clinical research is interpreted by physicians and

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

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.

References

Abrams, Eliot, Jonathan Libgober, and John A. List, "Research Registries: Taking Stock and Looking Forward," *Working Paper*, 2021.

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

Andrews, Isaiah and Maximilliam Kasy, "Identification of and correction for publication bias," *American Economic Review*, 2019, *109* (8), 2766–94.

Angelis, Catherine De, Jeffrey M. Drazen, Frank A. Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, Christine Laine, Ana Marusic, A. John P.M. Overbeke, Torben V. Schroeder, Hal C. Sox, and Martin B. Van Der Weyden, "Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors," *New England Journal of Medicine*, 2004, *351* (12), 1250–1251.

Angell, Marcia, "Is Academic Medicine for Sale?," *New England Journal of Medicine*, 2000, *342*, 1516–1518.

Arrow, Kenneth, "Economic Welfare and the Allocation of Resources for Invention," 1962.

Azoulay, P, M Heggeness, and J Kao, "The Impact of Health Care Financing on Research Outcomes: Evidence from Academic Medical Centers," Working Paper.

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.

Berenson, Alex, "Evidence in Vioxx Suits Shows Intervention by Merck Officials," *The New York Times*, 2005.

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.

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.

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.

Davidoff, Frank, Catherine D. DeAngelis, Jeffrey M. Drazen, John Hoey, Liselotte Hï¿œjgaard, 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.

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.

Gross, Daniel P. and Bhaven N. Sampat, "Inventing the Endless Frontier: The Effects of the World War II Research Effort on Post-war Innovation," Working Paper.

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).

Kesselheim, Aaron S., Christopher T. Robertson, Jessica A. Myers, Susannah L. Rose, Victoria Gillet, Kathryn M. Ross, Robert J. Glynn, Steven Joffe, and Jerry Avorn, "A Randomized Study of How Physicians Interpret Research Funding Disclosures," *New England Journal of Medicine*, 2012, *367* (12), 1119–1127. PMID: 22992075.

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.

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

Mizik, Natalie and Robert Jacobson, "Are Physicians "Easy Marks"? Quantifying the Effects of Detailing and Sampling on New Prescriptions," *Management Science*, 2004, *50* (12), 1704–1715.

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 Solï¿œne 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 andï¿œSuhayb 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.

Piller, Charles, "Transparency on trial," *Science*, 2020, pp. 240–243.

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.

Prayle, Andrew P, Matthew N Hurley, and Alan R Smyth, "Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study," *BMJ*, 2012, *344*.

Shapiro, Brad, "Promoting Wellness or Waste? Evidence from Antidepressant Advertising," *American Economic Journal: Microeconomics*, forthcoming.

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.

Sinkinson, Michael and Amanda Starc, "Ask Your Doctor? Direct-to-Consumer Advertising of Pharmaceuticals," *The Review of Economic Studies*, 2019, *86* (2), 836–881.

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.

Vivalt, Eva, "How Much Can We Generalize From Impact Evaluations?," *Journal of the European Economic Association*, 2020, *18*, 3045–3089.

Wood, Shelley, "The Price of Knowledge: Industry-Sponsored Studies in the Era of Evidence-Based Medicine," *TCTMD*, 2018.

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.

Appendix A

A.1 Antidepressant and Antipsychotic Drugs

Antidepressants and antipsychotics are both large and lucrative drug classes. 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 sales of more than a billion dollars (Ioannidis, 2008).[39] Total revenue fell in later years as some of these blockbusters went off patent, but the quantity of antidepressant prescriptions has increased over time. For example, the share of the U.S. adult population that takes antidepressants has increased 64% from 1999–2014 (Moore and Mattison, 2017).

Both drug classes have many substitutable drugs within the class and vibrant debate regarding their efficacy. Antidepressants were developed in several waves, beginning with the monoamine oxidase inhibitors in 1958 (Hillhouse and Porter, 2015). The earliest drugs 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 also includes all second-generation antidepressants approved either in the United States, Europe, or Japan, plus 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). For antipsychotics, this analysis includes the first-generation antipsychotics chlorpromazine (approved in 1957) and haloperidol (approved in 1967) along with thirteen second generations antipsychotics. The full sample of included drugs is shown in appendix figure B1.

A.2 Statistical Significance Calculation

In table 4 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

$$Z = \frac{p_1 - p_2}{\sqrt{\hat{p}\,(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{10}$$

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

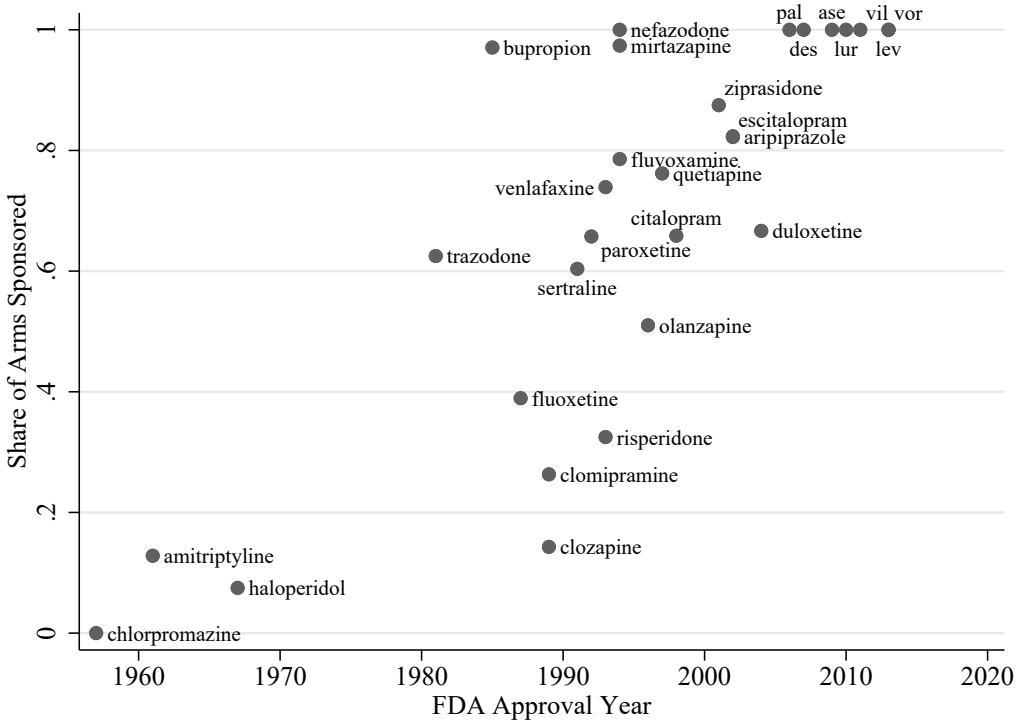
was significant at the 5% level. With an infinite sample, this Z-score cutoff was 1.64 for placebo-controlled trials and 1.96 for head-to-head trials. Here p is the proportion of patients that respond to treatment. The numeric indexing in equation 10 refers to the first or second arm, and $\hat{p}$ is 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

$$Z = \frac{e_1 - e_2}{\sqrt{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)}} \tag{11}$$

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

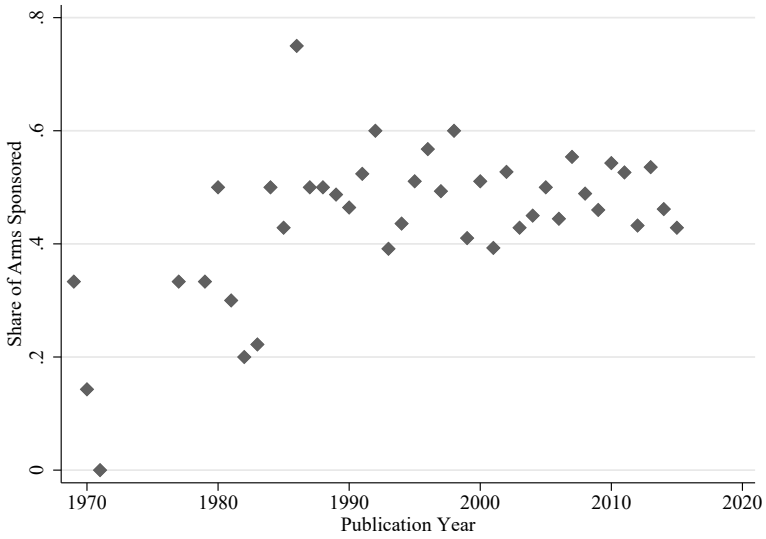
Figure B1: Included Drugs



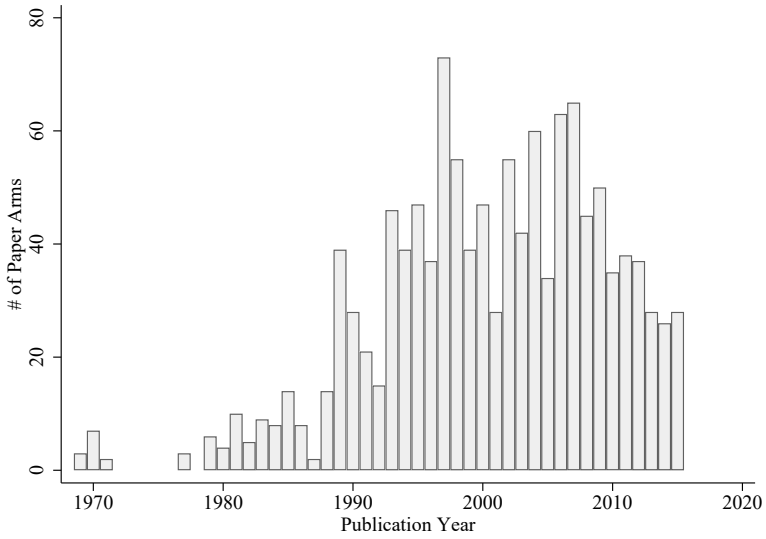
Notes: This figure presents the antidepressant and antipsychotic drugs included in this analysis. The x-axis presents the year of FDA approval for the drug, while the y-axis plots the share of arms in which that drug is sponsored by its manufacturer or marketer. The label "ase" refers to asenapine, "lur" 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: Variation in Sponsorship by Calendar Year

(a) Sponsorship

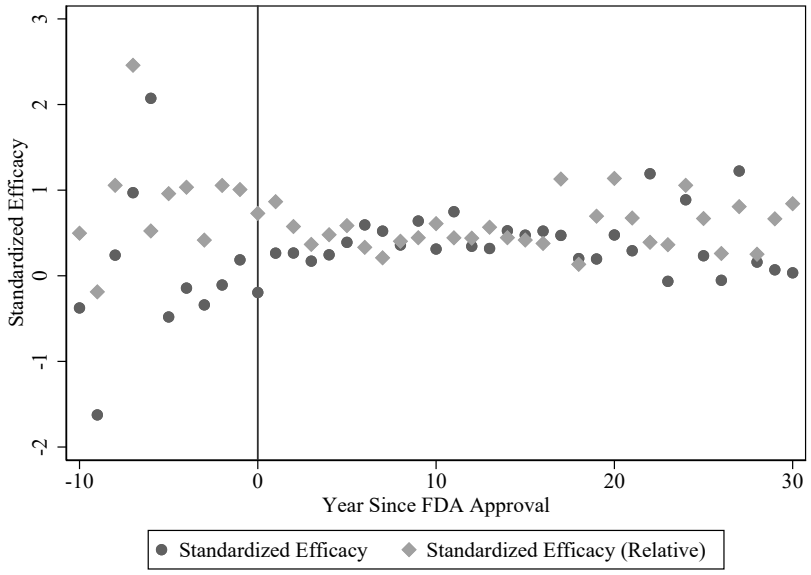


(b) Sample Size



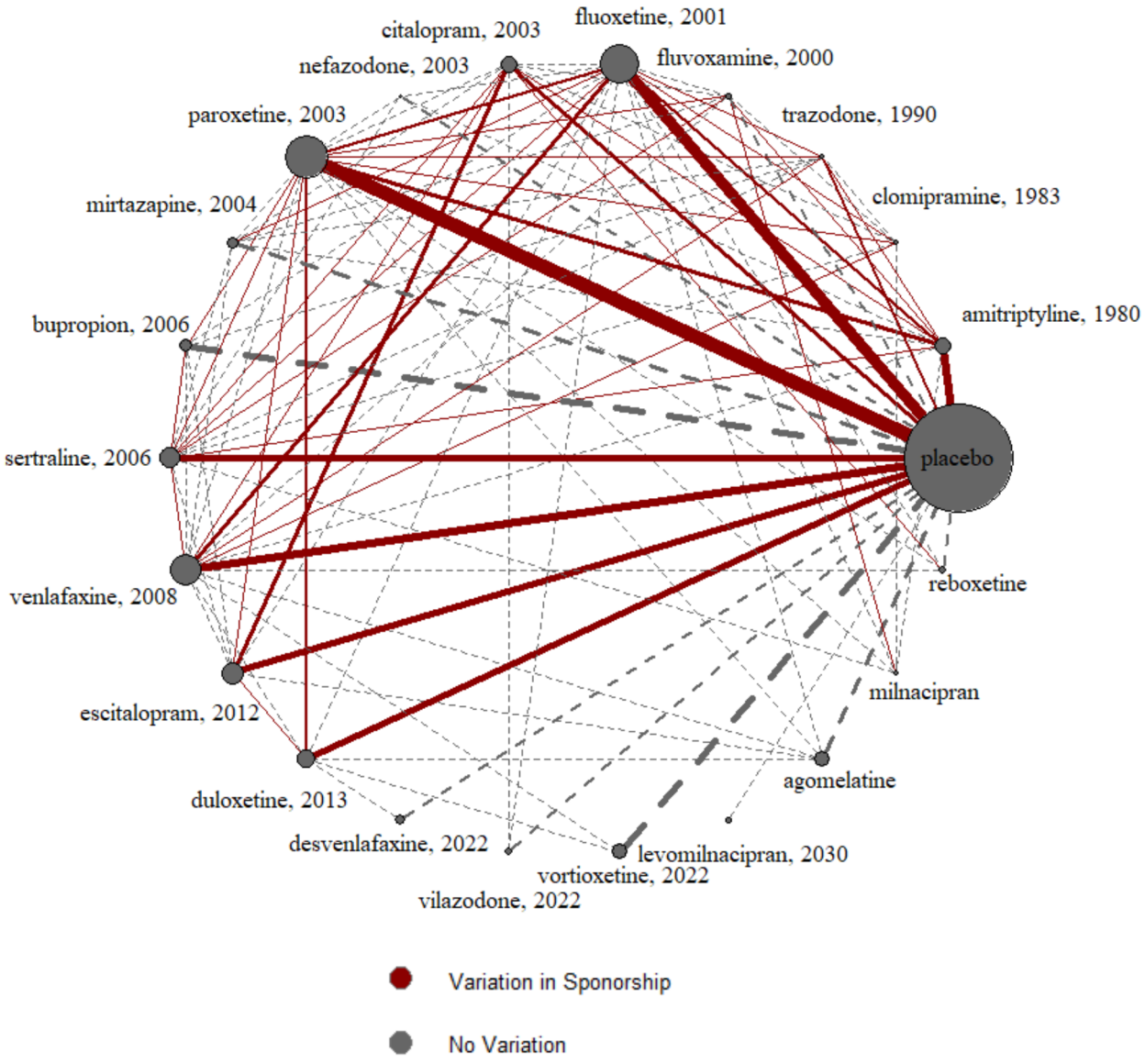
Notes: This figure (panel A) presents the average share of sponsored arms over time. The x-axis plots the publication year of the arm's trial. The y-axis plots the share of those arms that are sponsored. This figure excludes drugs that are not approved by the FDA (agomelatine, amisulpride, milnacipran, reboxetine, sertindole, and zotepine). Panel B presents the number of trial arms in my sample by their publication year.

Figure B3: Efficacy by Year Since Drug Approval



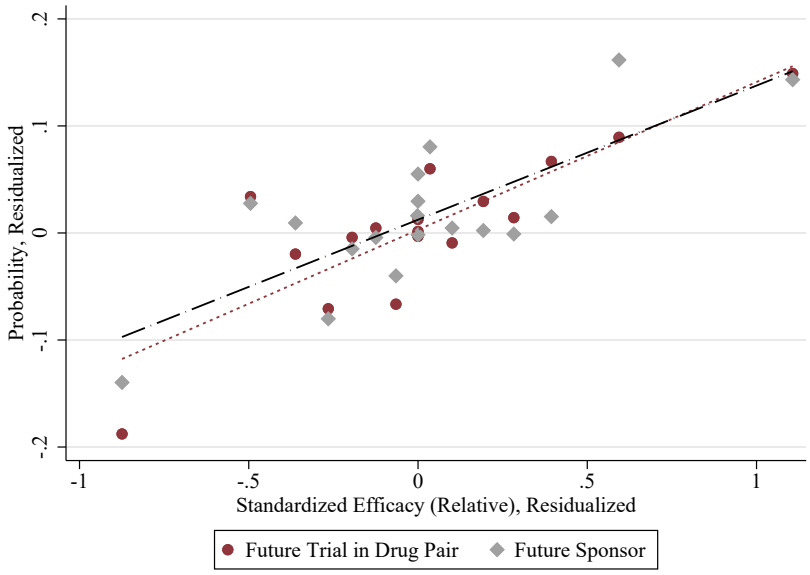
Notes: This figure presents the relationship between effectiveness and year since FDA approval. The x-axis plots the year the arm was published or made publicly available, relative to the FDA approval year for that drug. The y-axis plots the mean standard efficacy, or the mean standard efficacy, relative to the least effective or placebo arm in the drug pair. Details on these outcomes are listed in section 2.2.3.

Figure B4: Network of Trials for Antidepressants



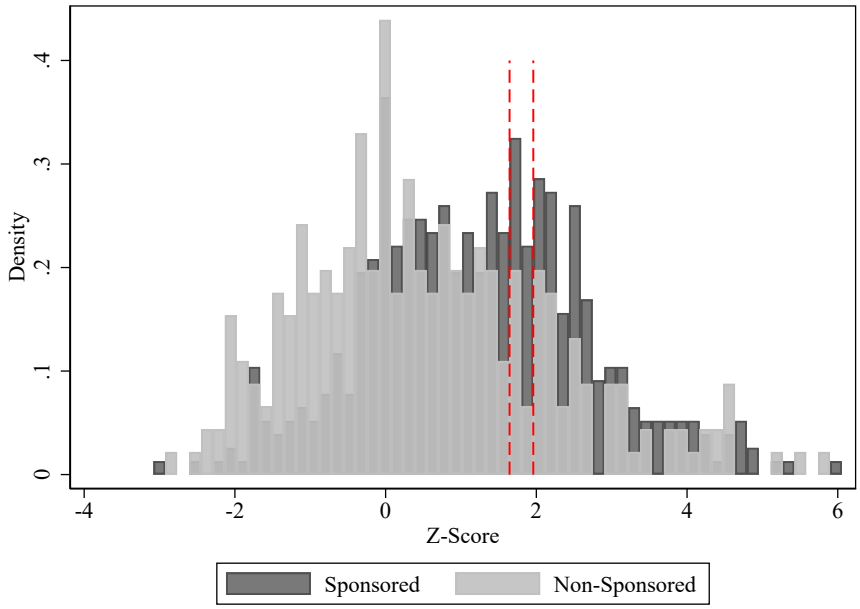
Notes: Figure presents the network of comparisons within antidepressants. Each node represents a drug and is labeled with the year that a generic formulation entered the United States market (years after 2019 are estimates). The size of the circle is proportional to the number of randomly assigned participants. 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 B5: Predictors of Future Papers and Sponsorship



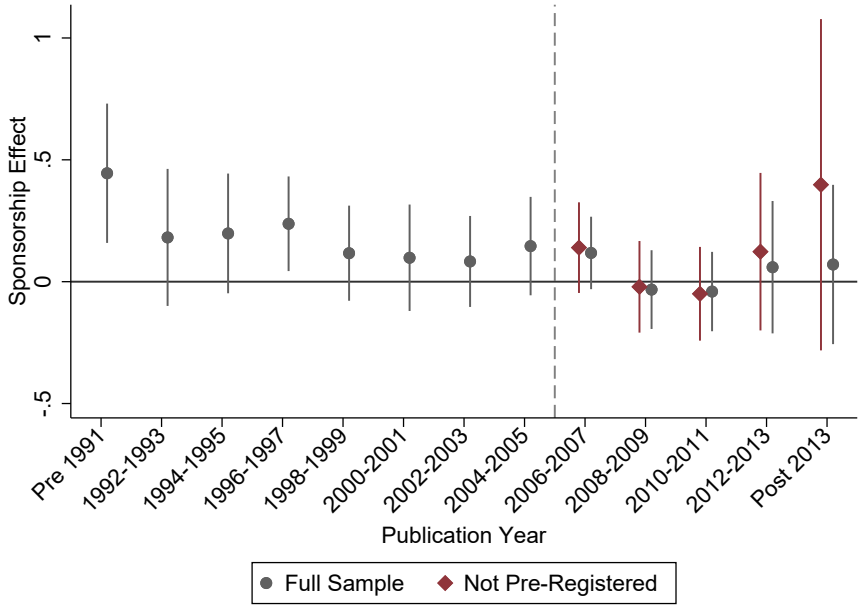
Notes: This figure presents the relationship between effectiveness and future papers and sponsorship. "Future Trial in Drug Pair" is an indicator for whether there is a future trial assessing that same drug in the same drug pair. "Future Sponsor" is an indicator for whether that drug is ever sponsored 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 B6: 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).

Figure B7: Sponsorship Effect by Calendar Year and Pre-registration



Notes: Figure presents the coefficients β_y from the estimation of equation 9 separately for all trials and for the subset of trials linked to ClinicalTrials.gov 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.

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 Antidepressants

	Sponsored				Not Sponsored				
	Drug	Other Arm	*Diff*	# Arms	Drug	Other Arm	*Diff*	# Arms	DD
All Drug Sets	0.647	0.597	0.049	50	0.567	0.583	-0.016	60	0.066
Amitriptyline vs. Fluoxetine	0.653	0.564	0.088	3	0.500	0.522	-0.022	10	0.111
Amitriptyline vs. Paroxetine	0.658	0.648	0.010	1	0.466	0.473	-0.008	8	0.017
Citalopram vs. Escitalopram	0.794	0.815	-0.021	6	0.639	0.760	-0.120	3	0.099
Fluoxetine vs. Venlafaxine	0.764	0.745	0.018	1	0.613	0.687	-0.074	7	0.092
Venlafaxine vs. Fluoxetine	0.687	0.613	0.074	7	0.704	0.707	-0.003	1	0.077
Paroxetine vs. Fluoxetine	0.531	0.475	0.056	6	0.683	0.565	0.119	1	-0.063
Clomipramine vs. Paroxetine	0.535	0.371	0.164	1	0.607	0.649	-0.042	4	0.205
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
Amitriptyline vs. Citalopram	0.650	0.625	0.025	1	0.516	0.548	-0.031	1	0.056
Fluvoxamine vs. Milnacipran	0.537	0.660	-0.123	1	0.571	0.702	-0.130	1	0.007
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
Sertraline vs. Venlafaxine	0.549	0.628	-0.079	1	0.569	0.653	-0.084	1	0.005
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 B3: Difference in Difference: Active versus Active Antipsychotics

	Sponsored				Not Sponsored				
	Drug	Other Arm	*Diff*	# Arms	Drug	Other Arm	*Diff*	# Arms	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 decline in the schizophrenia score 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 B4: 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)
$Sponsor_{ij}$	0.162**	0.213*	0.034*	0.022*	0.122	1.196	0.007
	(0.081)	(0.110)	(0.017)	(0.012)	(0.088)	(0.822)	(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	0.45	0.43	0.07	0.05	0.49	4.57	0.08
N	1,215	900	900	798	315	211	287

Panel B: Drug Pair

	Baseline	Standardized Outcome	Share Respond	% Decline	Standardized Outcome	PANSS Decline	% Decline
	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)	(7b)
$Sponsor_{ij}$	0.155***	0.207***	0.033***	0.027***	0.062	0.913	0.006
	(0.050)	(0.066)	(0.010)	(0.010)	(0.059)	(0.686)	(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	0.36	0.37	0.06	0.04	0.31	2.92	0.05
Weighted N	1,215	900	900	802	315	211	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 4, 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 B5: Alternate Specifications

Panel A:	Standardized Outcome (Relative)			
	(1a)	(2a)	(3a)	(4a)
$Sponsor_{ij}$	0.162**	0.155***	0.169***	0.371***
	(0.081)	(0.050)	(0.048)	(0.034)
Controls	X	X	X	X
Drug Combination Fixed Effects	Drug by Drug Set	Drug by Drug Pair	Drug	None
Mean Outcome	0.45	0.36	0.45	0.45
Weighted N	1,215	1,215	1,215	1,215
Panel B:	Standardized Outcome			
	(1b)	(2b)	(3b)	(4b)
$Sponsor_{ij}$	0.343**	0.229**	0.086	0.416***
	(0.158)	(0.102)	(0.086)	(0.053)
Controls	X	X	X	X
Drug Combination Fixed Effects	Drug by Drug Set	Drug by Drug Pair	Drug	None
Mean Outcome	0.06	0.06	0.06	0.06
Weighted N	1,215	1,215	1,215	1,215

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 4, 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 $^{***}p < 0.01$.

Table B6: Sponsorship by Drug Set Type

Panel A: Drug Set		Drug Set Type				
	Baseline	Active vs. Placebo	Active vs. Active			Only Variation
			Never	Always	Both	
	(1a)	(2a)	(3a)	(4a)	(5a)	(6a)
$Sponsor_{ij}$	0.162**	0.376	-0.027	0.337*	0.086	0.171**
	(0.081)	(0.301)	(0.221)	(0.200)	(0.080)	(0.070)
Controls	X	X	X	X	X	X
Drug by Drug Set F.E.	X	X	X	X	X	X
Mean Outcome	0.45	0.54	0.27	0.26	0.23	0.37
N	1,215	433	266	315	84	453
Panel B: Drug Pair						
	(1b)	(2b)	(3b)	(4b)	(5b)	(6b)
$Sponsor_{ij}$	0.155***	0.255**	0.031	0.318**	0.031	0.160***
	(0.050)	(0.099)	(0.172)	(0.143)	(0.053)	(0.048)
Controls	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X
Mean Outcome	0.36	0.51	0.27	0.31	0.22	0.38
$Weighted\ N$	1,990	840	448	1,116	256	986

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 4, 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)–(5a) and (2b)–(5b) 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 B7: 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 $^{***}p < 0.01$.