Funding of Clinical Trials and Reported Drug Efficacy*

Tamar Oostrom
The Ohio State University

February 17, 2024

Abstract

This paper estimates the effect of financial sponsorship of clinical trials on reported drug efficacy, leveraging the insight that the exact same pairs of drugs are often compared in different trials conducted by parties with different financial interests. I assemble new psychiatric trial data to estimate that a drug appears substantially more effective when the trial is sponsored by that drug's manufacturer, compared with the same drug tested against the same combination of drugs but without sponsorship. This difference is not explained by observable characteristics, but publication bias is important. Pre-registration may be effective in overcoming this bias.

JEL Codes: I11, I18, O31
Keywords: innovation, health, research and development, pharmaceuticals, scaling

*Contact: oostrom.1@osu.edu. I am very grateful to Amy Finkelstein, Heidi Williams, and Jim Poterba for their enthusiasm and guidance. The helpful comments of the editor and three anonymous referees substantially improved the manuscript. This paper also benefited from discussions with Sarah Abraham, David Autor, Pierre Azoulay, Ivan Badinski, Jane Choi, Laura Dague, Joe Doyle, Colin Gray, Jonathan Gruber, Ryan Hill, Allan Hsiao, Simon Jaeger, Kurt Lavetti, Madeline Mckelway, Adrienne Sabety, Parinitha Sastry, Frank Schilbach, 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. Edited by John List.

1 Introduction

In many markets, consumers and policy-makers have incomplete information on product effectiveness and quality. Consequently, firms often finance research on their own products. For example, automakers run fuel-economy tests for new vehicles, sunscreen manufacturers pay laboratories to test their products, and drug manufacturers often conduct clinical trials. Firm's research may have welfare benefits as other parties can use the knowledge produced at minimal marginal cost. 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. Trials are also 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 the pharmaceuticals market in the United States alone is valued at $480 billion (Yu et al., 2018; Moore et al., 2018; Wood, 2018). The results of trials shape regulatory, prescribing, and medical treatment decisions for decades afterwards (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. It also estimates the downstream consequences of financial incentives on trial characteristics and the availability of the research. The identification strategy uses the key insight that the exact same pairs 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 pairs 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. The market for psychiatric drugs is significant, with 12.7% of the U.S. adult population using antidepressants monthly, and 1.6% using antipsychotics (Pratt et al., 2017; Moore and Mattison, 2017). Depressive disorders impose an estimated economic burden of $210 billion in the United States annually (Greenberg et al., 2015). Antidepressants and antipsychotics also conveniently had several large and recently published meta-analyses on their efficacy (Cipriani et al., 2018; Leucht et al., 2013), which enable me to clearly define the relevant sample of drugs.[2]

[1]A notable exception is Allcott (2015), which assesses site selection bias in an energy conservation program.

[2]Each trial in the sample is a double-blind RCT. These trials were conducted before and, mostly, after the drugs gained regulatory approval. Some trials are sponsored by the manufacturer of one of the drugs, while others receive funding from governments, alternate private firms, or the authors are academic researchers at a university or medical school. Section 2.3

As an example of the identifying variation, Wyeth Pharmaceuticals introduced a new antidepressant drug Effexor in 1993. Over the next decade and a half, Wyeth funded RCTs comparing the effectiveness of Effexor with Eli Lilly's blockbuster drug Prozac. In twelve of the fourteen trials funded solely by Wyeth, Effexor was more effective than Prozac. In contrast, only one of the three trials with alternate funding found Effexor to be more effective. Each of these trials is a double blind RCT comparing the exact same two molecules and examining the same standard outcomes.[3] Building on this illustrative example, I systematically investigate the effect of an RCT's funder on the reported efficacy of the tested drugs. As highlighted in this example, a drug's efficacy is usually reported relative to the other arms in the trial. Government, industry, and publication decisions are also based on a drug's relative efficacy.

First, I use variation in trial funding to show that financial incentives affect reported drug efficacy. Efficacy is based on standard outcomes in the medical literature, measured relative to the other arms in the trial, and standardized to have a mean of zero and standard deviation of one. I find that a drug is reported to be 49 percent more effective (0.17 standard deviations off a base of 0.35) 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 43 percent more likely to report statistically significant improvements (0.10 off a base of 0.24), and 73 percent more likely to be the most effective drug in their trial (0.28 off a base of 0.39), again, compared with the same molecule tested against the same pair of drugs, but without funding from the drug's manufacturer. I term the 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. Potentially trials with industry funding occurred early in the drug's life cycle and coincided with idiosyncratically high effectiveness, while later trials had lower effectiveness simply due to mean reversion. In robustness checks, I find similar results after controlling for time since approval, as well as restricting to only post-approval trials.

This paper focuses on financial incentives rather than academic or government incentives, since financial incentives can be more directly assigned to a drug. This analysis does use variation in funding both within-industry and across industry versus academic or government-run trials. I find a sponsorship

contains more information on the trials.

[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. This example uses brand names, but the rest of the paper uses generic names interchangeably.

[4]Clinical trial results may selectively report and highlight specific outcomes. In this analysis, I focus on a consistent set of outcomes to focus on differences in reported efficacy, not reporting decisions.

effect in both categories separately. Estimates using only within-industry variation are larger, which is consistent with within-industry trials having two sets of opposing incentives compared to industry versus unsponsored trials.

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. 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 sponsorship effect. The addition of unpublished trials attenuates the effect of sponsorship, and most of the sponsorship effect can be explained by publication bias.

Another class of potential mechanisms is trial design, where trials are planned or conducted differently. I test for this mechanism by incorporating data on trial characteristics including the length of the trial, the drug's dosage, and total enrollment as well as the average age, gender, and baseline severity of the enrolled patients. In balance checks, I show that, within a pair of drugs, trials with different funding are similar in observable trial and patient characteristics. Controlling for trial and patient characteristics also does not materially change the sponsorship effect, and the sponsorship effect within same drug, drug combination, *and* dosage or patient characteristics is still positive and statistically significant. I also find no evidence that sponsors chose trial design features that favor their drugs based on each characteristic separately, and for all patient and trial characteristics combined. This analysis is constrained by characteristics which are observable, and part of the sponsorship effect may be due to selection on unobserved trial design. 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 a natural 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 publication bias and find that the effect of sponsorship on reported drug efficacy is statistically significantly lower after the introduction of pre-registration, 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.[5] While there were other

[5]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 require-

concurrent changes in social norms, results reporting, 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). I demonstrate that both are true: pharmaceutical companies test more effective drugs and select worse comparison drugs, leading to bias in the correlation between industry funded trials and efficacy outcomes. In this paper, I measure the causal effect of changing sponsorship *for a given drug and evaluated against the same competitors*, a novel contribution.

This paper builds on a growing literature on implementation science and replicability. Medical evidence has long been based on clinical trials, but recent work has highlighted issues of bias and external validity in RCTs (e.g.Vivalt (2020); Abrams et al. (2021)). Previous studies in economics (Camerer et al., 2016), psychology (Open Science Collaboration, 2015), and finance (Menkveld et al., forthcoming) have shown that treatment effects can vary substantially in different contexts. This is phenomenon is also called the scaling problem (List, 2022).

In my paper, the scaling problem is due to false positives. There are fewer degrees of freedom in medical trials than in the social sciences, and I estimate the effect of funding while holding the efficacy outcome, duration, drug, and drug combination in a trial fixed, limiting sources of non-standard errors (Menkveld et al., forthcoming). I also find evidence that the experimental population, as measured by patient characteristics, and the experimental situation, such as trial characteristics, are not substantially different between different funders (Al-Ubaydli et al., 2017). Consistent with theoretical results in scaling, I find that the sponsorship effect is greater for drugs with a larger market size and for more novel drugs. Additionally, the sponsorship effect decreases as the costs for non-replicability increase through required pre-registration, aligning with existing theoretical predictions (Al-Ubaydli et al., 2020).

This research underscores the impact of financial incentives on pharmaceutical innovation and the types of knowledge generated. The findings suggest that clinical trial publications are valuable resources for pharmaceutical firms, consistent with the effectiveness of direct-to-consumer advertising (Sinkinson and Starc, 2019; Shapiro, 2022) and detailing (Mizik and Jacobson, 2004), both of which rely on sci-

ments (Abrams et al., 2021).

entific publications. Furthermore, this study contributes to the literature on private research investments and incentives (Budish et al., 2015).

Removing the sponsorship effect would reduce the difference in efficacy between a sponsored drug and other drugs in the trial by about 50%. This may have important consequences for drug approval and prescription decisions. However, if physicians, patients, 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, forthcoming; Ching et al., 2016). My results suggest that sponsored arms of trials should be discounted substantially. Back of the envelope calculations suggest that discounting sponsored arms appropriately would relate to 10% fewer psychiatric drug approvals and 8-18% fewer prescriptions.

Section 2 presents institutional background on clinical trials and psychiatric drugs and introduces the data. I outline the empirical strategy and present estimates of the effect of sponsorship on reported drug efficacy in section 3. Section 4 investigates mechanisms, focusing on publication bias and trial design. Section 5 tests theoretical predictions on incentives in scaling and the effect of required pre-registration. 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 large financial stakes. There are the direct costs of conducting clinical trials, high failure rates, and the opportunity cost of capital. The research and development spending per drug approved can be $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.

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 FDA recommends three to five controlled clinical trials demonstrating substantial evidence of efficacy to support approval. The FDA recommends testing new antidepressants both in trials against a placebo and against the current standard of treatment. 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, produce marketing material, and differentiate the drug against competitors. Publications of clinical trial results provide material for pharmaceutical sales representatives to cite in the promotion of drugs to physicians, medical journal advertisements and direct to consumer advertising.[6]

2.2 Psychiatric Clinical Trial Data

The clinical trial data in this paper contain all available double-blind RCTs for either antidepressants or antipsychotics.[7] 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 the 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.[8] The final dataset contains efficacy and sponsorship information, as well as the length of the trial, the drug's dosage, total enrollment and patient characteristics such as the mean age, gender, dropout rate and baseline severity.

Supplemental data include the Medical Expenditure Panel Survey (MEPS) from 1996-2019 and 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,

[6]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).

[7]Background on these drug classes is provided in appendix A1.

[8]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 the paper for completeness. In practice, these additions add no new variation as the additional arms all have consistent sponsorship and the estimates are essentially the same.

authors, funders, and many trial characteristics for over 300,000 clinical trials as of 2020.

2.2.1 Defining Terminology

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 contain the same drug pair of paroxetine versus venlafaxine, though they test different drug sets. 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 the same drug but different dosages in different arms.

2.2.2 Defining Sponsorship

A treatment arm is sponsored if any of the following conditions are met: the trial was funded by drug's manufacturer or marketer, one of the authors had an affiliation with the company, or the data came from documents on the company website, or the drug manufacturers were listed in the author's conflicts of interest statement or acknowledgements.[9][10] For example, consider a trial that compares escitalopram to venlafaxine and a placebo in which one author 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.

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 behavior to a numeric score. The most common scale for antidepressants

[9] 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 A1. I also consider robustness to the definition of sponsorship in table 3. In three cases, I revised the Cipriani et al. (2018) sponsorship definitions based on likely errors after reviewing the initial publications. Using the original coding for antidepressants increases most point estimates and makes no significant difference in the results.

[10] This paper focuses on financial incentives, since these can be assigned to one drug within a trial. Academic and government-run trials may also have incentives, but incentives to simply find larger effects would apply to either drug in the trial.

is the Hamilton Score for Depression (HAMD); this scale is available for 85% of the antidepressant sample. 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).

The standard efficacy measure for antipsychotics is the mean change in the total Positive and Negative Syndrome Scale (PANSS) score or, if the PANSS score is not available, the Brief Psychiatric Rating Scale or the Clinical Global Impressions–Schizophrenia Scale, in that order (Leucht et al., 2013). In robustness checks, I consider the percent decline in either the total depression or the antipsychotic scores.

For both drug types, 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).

2.3 Sample Construction and Summary Statistics

The antidepressant and antipsychotic meta-analyses contain 732 total clinical trials. I obtained the original publications or clinical trial reports for 656 trials. After dropping observations with missing efficacy or sponsorship information, the sample contains 586 trials and 1,412 treatment arms. In the initial analysis, I focus on only published papers, which consists of 509 trials and 1,215 treatment arms.

Figure 1(a) plots the average share of treatment arms that are sponsored by the time since the drug gained FDA approval. Prior to FDA approval, most drugs are tested by that drug's manufacturer. For the two decades after FDA approval, a drug is sponsored about half the time. Thirty or more years after FDA approval, almost none of the drugs are still sponsored. Figure 1(b) plots the share of arms by the year relative to the FDA approval year. The majority of the trials occur just before and in the ten years immediately after FDA approval and would be classified as Phase IV trials.[11] On average sponsored arms occur earlier in a drug's life cycle than non-sponsored arms. The difference in age between sponsored and non-sponsored arms is reduced with drug and drug pair controls, and additional robustness that considers the age of drugs is shown in section 3.5.

Appendix table A1 presents summary statistics on trial characteristics. The average trial in the sample was published in 2001. Just under half of all arms are considered sponsored, and seven percent are considered sponsored due to conflicts of interest alone. Approximately three-quarters of the data are from antidepressant trials and the remaining quarter are from antipsychotic trials. Only 12% of the sample is ever pre-registered, as measured by having a National Clinical Trial (NCT) number listed on

[11]In contrast, the share of arms sponsored by calendar year has remained fairly constant within the sample (see appendix figure A1).

ClinicalTrials.gov. Among the full sample, 86% were published after the drug in that arm gained FDA approval. The average treatment arm enrolled 100 patients and the average trial length was nine weeks. On average 29% of patients dropped out of each arm before the trial completed. These arms enrolled 51% women on average, and the average patient was 42 years old. Since the identification strategy uses variation in sponsorship, I present summary statistics for the subset of trials with variation in sponsorship separately, which are similar to the full sample.

3 The Effect of Sponsorship: Empirical Strategy and Results

3.1 Description of Sponsorship Variation

The main types of drug combinations are presented in table 1. Each box refers to an example trial, where the funder is listed at the top and the treatment arms listed below. Trials are only compared with others in the same row. In each row, one drug varies in sponsorship while the other drugs remain constant in funding. Only comparing trials across rows is key to the analysis because it ensures that the sponsorship effect is estimated using only differences in funding among trials with the exact same drug combinations.

The first category ("Active vs. Placebo") directly compares a psychiatric drug ("Drug A") to a placebo. Some of these trials are sponsored by the company that manufactures drug A ("Company A"). The other unsponsored trials have alternative funding not provided by company A.[12] Thirty percent of trials are in this category.

The second category in table 1 ("Active vs. Active") contains drug combinations that compare an active drug to another active drug. This occurs in 45% of trials. In all cases, "drug A" varies in funding. 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. Multiple pharmaceutical companies can be involved in a trial if the authors have several conflicts of interest or affiliations. In the third subgroup, the sponsorship interests of both drugs vary.

3.2 Difference in Difference Framework

The key finding in this paper can be succinctly summarized using raw means in figure 2. This figure, along with all the results in section 3, uses only published papers, which consists of 509 trials and 1,215

[12] While most trials are conducted with financial assistance from one of the drug's manufacturers, 54 trials (11%) have no sponsored arms. Twenty of these are funded by a governmental agency, such as the National Institute of Mental Health (5), or the Department of Health of Taiwan (2). Thirty-two papers list no government or industry funding and have a first author with an academic or hospital affiliation, such as the Medical College of Georgia (2) or the University of Munich (2). The remaining two papers have industry funding from an unrelated firm.

treatment arms. Panel A presents all drug sets that compare an active drug to a placebo and have variation in sponsorship. Each row represents a unique drug set, where the first listed drug varies in sponsorship across trials and the second listed drug has the same sponsorship status in all trials.

As an example, consider the second row, which considers trials that compare paroxetine to a placebo. In the trials where paroxetine is sponsored, an average of 47% of patients receiving paroxetine respond to treatment. This corresponds to the solid black circle. In those trials, an average of 31% of patients respond to the placebo, shown in the hollow black circle. Therefore, on average, paroxetine is 16 percentage points more effective than the placebo in sponsored trials. Turning to trials in which paroxetine is not sponsored, 25% of patients receiving paroxetine respond to treatment as shown in the solid gray triangle, while 23% of patients responded to the placebo, as shown in the hollow gray triangle. On average, paroxetine is 2 percentage points more effective than the placebo in unsponsored trials. The difference in difference estimate of the sponsorship effect for paroxetine versus a placebo is 14 percentage points. This is shown in black dots on the left. The following rows present estimates for other drug sets and the first row presents the average effect across all trials in this category, weighting by the number of trials.[13]

Panel B presents the analogous estimates for the "Active vs. Active" category in table 1. The row labels now list both drugs in the drug set. In the majority of drug sets, the difference-in-difference estimate is positive. This means that a drug is more effective when it is sponsored, relative to the other arm, compared with the same unsponsored drug in the same drug set, relative to the other arm. This positive sponsorship effect holds for four out of five active versus placebo drug sets, eighteen out of twenty-five active versus active antidepressant drug sets, and nine out of twelve active versus active antipsychotic drug sets. Appendix tables A2 and A3 present the individual components for these difference-in-difference estimates in a table, along with the number of trials in each drug set.

3.3 Estimating Equations

The regression specification is conceptually similar to figure 2. Both compare the efficacy of a drug when it is sponsored versus not sponsored, relative to other arms in those trials. The regression specification includes a few components that improve precision. First, I standardize the efficacy measure to combine the estimates for both antidepressants and antipsychotics. Secondly, the regression is at the arm level, so drug combinations with more trials and arms receive more weight. Finally, the main regression specification uses variation within drug pairs, while figure 2 presents comparisons within drug sets. For

[13] This figure does not contain standard errors since some of the categories only have a single observation. The regression specification in the next section presents standard errors for very similar estimates.

all trials with two arms (75%), drug sets and drug pairs are identical. However, drug sets with three unique arms can contribute to three drug pairs. This allows for more variation in sponsorship since a trial with three arms can be included in some of the comparisons shown in figure 2.

In the main analysis, I estimate the following specification:

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

where y_{ij} is the efficacy for arm i in trial j. The coefficient of interest is on $Sponsor_{ij}$, which is a dummy for whether arm i was sponsored in trial j. I control for X_{ij} which denotes the type of measurement scale for arm i and the year published for trial j.[14]

Most importantly, $G_{d(i),p(j)}$ is a dummy for each unique drug $d(i)$ in each separate drug pair $p(j)$. Each arm i can be mapped to a unique drug $d(i)$. Each trial j can be mapped to at least one and potentially multiple drug pairs $p(j)$. As described in section 2.2.1, a drug pair is a combination of two drugs in a clinical trial. This is key to the analysis, because it ensures that the sponsorship effect is estimated using differences in funding sources among trials comparing the exact same pairs of drugs. These fixed effects for each drug combination are analogous to the separate rows in table 1 and figure 2. Appendix table A4, column (2), provides a more detailed example of this fixed effects structure, and compares this specification with drug set fixed effects, which are included in robustness checks.[15]

In most cases, the outcome y_{ij} is computed *relative* to the placebo arm in the drug pair $p(j)$, if available, or least effective arm, otherwise.[16] 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.

3.4 The Effect of Sponsorship on Reported Efficacy

Table 2 presents the regression estimates from equation 1. In column (1), I find that a sponsored drug is 0.18 standard deviations more effective than the same drug in the same drug pair without sponsorship. Controlling for the publication year and the type of psychiatric score in column (2) reduces the sponsorship effect slightly to 0.17. The sponsorship effect in column (2) is 49% of the average relative efficacy

[14] 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.

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

[16] 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 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. I show estimates using the absolute efficacy and other outcomes measures in table 2.

of 0.35 standard deviations. Therefore, the funding interests of a given drug can explain almost half of the relative efficacy of that drug.

Column (3) presents estimates using the absolute efficacy, rather than the relative efficacy. Sponsored arms are 0.26 standard deviations more effective in absolute efficacy than non-sponsored arms of the same drug and drug pair. The main analysis focuses on relative efficacy as regulatory decisions, publication decisions, and the papers themselves focus on the efficacy of drugs relative to the other arms in the trial (see appendix section C and appendix table A5). Within a drug pair, sponsored trials increase the efficacy of both the sponsored drug and the least effective drug in the trial (see appendix section D and appendix table A6). Therefore, the absolute efficacy sponsorship effect is larger than the relative efficacy effect as it does not incorporate changes in the other arms of the trial.

In column (4), the outcome is an indicator for whether the arm was statistically significantly more effective than the other arms in that trial. Appendix section B provides details on the construction of this variable. On average, sponsored arms are 10 percentage points more likely to be statistically significant at the 5% level. This represents a 43% increase over the baseline 24% of arms that are statistically significant. 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. In column (5), the outcome is an indicator for whether the given arm was the most effective arm in that trial. Sponsored arms are 0.28 percentage points more likely to be the most effective arm, compared with that same drug evaluated in the drug pair, but without sponsorship. This is a 73% increase over a baseline of 0.39.[17] Column (6) uses the percent decline in the psychotic score, relative to the placebo or least effective arm. While this is not the standard efficacy measure used in columns (1)-(3), it also shows a positive sponsorship effect. In appendix table A7, I show that including drug-by-set fixed effects, rather than drug-by-pair fixed effects, yields very similar estimates in magnitude, with less statistical precision.

Appendix section D presents results with alternate specifications for completeness. I show that industry choses to fund more effective drugs than government or academic trials, which yields a positive unconditional relationship between sponsorship and efficacy. In addition, sponsored trials choses to test their drugs against worse competitors as shown in appendix table A6. Therefore, using only drug fixed effects or no fixed effects, as in prior literature and appendix table A8, does not capture the sponsorship effect of interest.

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

3.5 Robustness

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 A2 plots the average efficacy of sponsored arms by the year since approval. 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.

Table 3 accounts for any systematic changes in efficacy over the drug's life cycle and mean reversion. Column (1) replicates the baseline estimate from Table 2, column (2). Column (2) controls for the publication order of the trial within the drug pair. This slightly decreases the sponsorship effect estimate by 6%. Column (3) controls for the year relative to the drug's approval year; this estimate is 0.14 compared to the baseline effect of 0.17 but is still statistically significant. As an additional test of whether the FDA approval benchmark is distortionary, I restrict the sample to only post-approval trials (column (4)). The point estimate decreases by 15% and the estimate of the sponsorship effect remains statistically significant. In all cases, the sponsorship effect is similar though a bit smaller, 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 direct funding. I examine robustness to excluding conflicts of interest from the definition of sponsorship (column (5)). In this case the sponsorship effect is a bit smaller at 0.15 standard deviations, but still statistically significant.

The 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 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 (6)). This estimate is smaller than the baseline estimate, but also statistically significant.

3.6 Heterogeneity by Variation Type

There are two main types of drug pairs – pairs that compare an active drug to a placebo drug and pairs that compare two active drugs. Table 4 presents estimates for these two subsamples in columns (1) and

(2). The sponsorship effect in the active versus placebo sample is larger, but this group has a larger average relative efficacy as well. In percent terms, the sponsorship effect in column (1) is 55% (0.27 off of a base of 0.49), the same as the active vs. active column's estimate of 50% (0.12 off of a base of 0.25). Columns (3) and (4) separate the analysis by the type of drug—antidepressant or antipsychotic. Most of the trials are antidepressant clinical trials and this sample drives the sponsorship results. Antipsychotics are a small share of the analysis sample, so results within this subset are not statistically significantly different from zero, nor statistically different from the antidepressant sample.

Column (5) restricts to the subset of the drug pairs that have at least one unsponsored trial. Unsponsored trials are almost always funded by a governmental agency or have authors with academic affiliations. The sponsorship effect in this subset is estimated by comparing industry-funded and unsponsored trials. The sponsorship effect is 39% (0.16 off of a base of 0.41), which is lower than the baseline. In contrast, column (6) only uses variation across industry-funded trials and has a much larger sponsorship effect of 83% (0.25 off of a base of 0.30). Industry versus unsponsored trials have incentives for the industry-funded drug to appear more effective in one set of trials, but the unsponsored trials are not incentivized to make either drug more effective. In contrast, within industry variation has two sets of opposing incentives and a much larger effect.

3.7 Which Drug Trials Have Variation in Sponsorship?

The identification is driven by the subset of drug combinations that have variation in sponsorship. Appendix table A9 presents the share of arms that have variation in funding by characteristics. Among antidepressants, the drug classes of tricyclics and selective serotonin reuptake inhibitors (SSRIs) are most likely to have variation in funding. The former are the first antidepressants and the latter are the most prescribed class of antidepressant. 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.

This pattern is also shown in appendix figure A3, 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).

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 occurs after data collection (i.e. ex-post mechanisms).

4.1 Trial Design

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).[18] To test whether these characteristics systematically explain the sponsorship effect, I assess whether sponsored arms differ in trial or patient characteristics in appendix section E and appendix figure A4. Within a drug pair, sponsored trials occur about four years earlier and have slightly older patients; they are statistically indistinguishable in terms of registration, the number of patients, length, dosage, baseline severity, dropout share, or share female.

I also test whether controlling for these characteristics affects the estimates. The first column in table 5 replicates the baseline estimates. Controlling for trial characteristics (total enrollment, length of trial, and dosage) increases the point estimate slightly, while controlling for patient characteristics (mean age, share female, baseline severity, and dropout share) slightly decreases the point estimate. With the full set of controls, the estimate is 0.16, which is similar to the baseline estimate of 0.17.

4.1.1 Sponsorship Effect Within Patient and Trial Characteristics

Simply controlling for patient and trial characteristics does not account for the concern that characteristics might be differentially predictive of efficacy *within a given drug and drug pair.* I conduct two analyses to assess this mechanism. First, I compute the sponsorship effect within drug, drug pair and certain characteristics. I focus on dosage, age, gender, and baseline severity since these are commonly featured in heterogeneity analyses for other drug types. I estimate

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

[18] 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).

which is identical to equation 1, except instead of drug by drug pair fixed effects, I include fixed effects for each drug by drug pair and characteristic group k of arm i. In column (5), the characteristic group is the exact minimum dosage in arm i. In column (6), the characteristic group includes the dosage, two bins for the average female share in the trial and two bins for mean age.[19] Column (7) includes all the earlier characteristics and adds two bins for baseline severity. That column can be interpreted as the sponsorship effect within a given drug, drug pair, dosage, share female, mean age, and baseline severity. In all columns (5)-(7) the sponsorship effect is positive, statistically significant, and ranges from 0.16 to 0.19 standard deviations. The specificity of the fixed effects limits the variation that can be used to identify the sponsorship effect and increases the standard errors. [20]

4.1.2 Predicted Efficacy

As a last test, I estimate whether sponsored arms chose characteristics that are predicted to be more effective for their drugs. I create drug-specific predicted efficacy by regressing

$$y_{ij} = \alpha + \sum_{k(i)} \sum_{d(i)} \beta_{k(i),d(i)} Z_{k(i)} d(i) + X_{ij}\gamma + \varepsilon_{ij} \tag{3}$$

where y_{ij} is the outcome for arm i in trial j, $Z_{k(i)}$ is each characteristic k (e.g. baseline severity, share female) interacted with each drug $d(i)$, and X_{ij} controls for the type of measurement scale and the year published as in section 3.3.

I use the estimates from equation 3 to compute $\hat{y}_{ij}$, the predicted efficacy for arm i in trial j for every characteristic. Then, I re-estimate the main regression from equation 1 with relative predicted efficacy on the left hand side:

$$\hat{y}_{ij} = \alpha + \beta Sponsor_{ij} + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{4}$$

The coefficient on $Sponsor_{ij}$ can now be interpreted as “how large would we expect the sponsorship effect to be, simply because sponsored arms are more or less likely to enroll characteristic k?” I first estimate these results separately by each characteristic. Table 6 shows that sponsored arms do not have higher predicted efficacy for any individual characteristic. The largest coefficient is on the dropout rate, though this is not statistically significant. Trials with lower dropout rates generally have higher efficacy, and sponsored arms are more likely to have lower dropout rates. I also combine all covariates

[19]Mean age among trials is bimodal, with two peaks in the early 40s and in the 60s. Similarly, the share female is bimodal, with distributions just below and above 50%.

[20]The inclusion of even more specific fixed effects with additional characteristics leads to even larger standard errors. Including drug by drug pair by all characteristic fixed effects leaves no variation left to estimate the sponsorship effect and the coefficient on sponsorship is not identified.

in one prediction, using LASSO to select the most predictive characteristics. As shown in table 6, column (8), sponsored arms are not predicted to have higher relative efficacy based on the all observable characteristics.

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 the analysis is there are many characteristics of trial design not included in these observable characteristics, such as the patient's willingness to adhere to treatment, their underlying health conditions, or the level of monitoring during treatment. These might be notable components of the sponsorship effect.

4.2 Publication Bias

4.2.1 General Tests for Publication Bias

Another potential mechanism for the sponsorship effect is publication bias. To test for publication bias by sponsorship, I assess whether sponsored arms are more likely to be published if they report higher efficacy, compared to unsponsored arms. As noted in section 2.2, I observe data on seventy-seven unpublished antidepressant or antipsychotic clinical trials. These unpublished trials are a subset of the universe of all unpublished trials ever conducted, as most unpublished clinical trials are never made available. The unconditional relationship between reported efficacy and the share of arms published is presented in figure 3.[21]

Among a combination of the 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.

Table 7 shows these results hold within a drug pair. Specifically, I estimate:

$$1\{Published_j\} = \alpha + y_{ij} + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{5}$$

where the outcome is an indicator for whether trial j was published. The coefficient of interest is on y_{ij}, the relative efficacy of a given arm i in trial j. The rest of the terms are the same as in equation 1, though X_{ij} now includes only the type of measurement scale. I estimate this equation separately for sponsored

[21]This figure compares all sponsored arms to all unsponsored arms and combines information across all drugs in the sample. It is therefore not informative of the overall sponsorship effect, which is computed within a drug and drug pair.

and unsponsored arms. The relationship between relative efficacy and publication is much stronger for sponsored arms than for non-sponsored arms, which corroborates the results from figure 3.[22][23]

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 all conducted trials. After 2005, many journals required that authors pre-register their clinical trial before patient enrollment. Therefore, I use the sample of all pre-registered antidepressant clinical trials as an approximation of the full set of trials.

I further narrow down the sample to trials assessing major depressive disorder or depression, testing at least one of the antidepressant drugs in the sample, with a purpose of treatment or basic science. I include trials with randomized allocation, parallel treatment assignment, and enrollment limited to depressed patients. Excluded are trials involving children, chronically depressed patients, and trials testing a single drug without a placebo or alternate treatment arm. These criteria align with Cipriani et al. (2018). This registry sample includes 90% of the trials in the analysis sample that were registered on ClinicalTrials.gov.[24] In the other direction, only 6% of the registered trials had results that were not in the analysis sample.[25]

I then restrict the registry sample to trials submitted between 2006 and 2010, to allow time for registered trials to be observed in the analysis sample.[26] Out of the 163 pre-registered trials meeting this criteria, the analysis sample contains results for just 23% of them. Therefore, I estimate that there are approximately four times more trials for each trial observed in the analysis sample. This estimate aligns with previous evidence indicating that only 22% of pre-registered trials report results (Prayle et al., 2012), though estimates of results reporting vary widely. To approximate the sponsorship effect in the presence of additional trials, I randomly draw from the unpublished trials in the analysis sample to

[22] The difference between the sponsored and not sponsored arms is the main take-away from table 7. There are many more unpublished papers that I do not observe so interpretation of the magnitude requires additional assumptions as in section 4.2.2.

[23] Another standard test for publication bias is to measure the level of bunching around z-score cutoffs. Appendix figure A5 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.

[24] The registry sample includes 64 of the 71 registered trials in the analysis sample. Of the seven trials in the 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.

[25] Specifically, of the 314 trials with this inclusion criteria that were not included in the analysis sample, nineteen had available results or publications. In many the trial was not assessing depression symptoms or started too late to be included in the analysis sample.

[26] The median time from submission to the registry to publication is four years. The 90th percentile is five years. A five-year gap from submission to potential publication allows a trial submitted in 2010 to potentially be observed in 2015 in the analysis sample. This analysis is restricted to antidepressant trials since the inclusion criteria in the antidepressant meta-analysis corresponds closely to Clinicaltrial.gov variables.

approximate the missing trials.[27]

To benchmark the share of the sponsorship effect explained by publication bias, I assume that the sponsorship effect among the unpublished trials observed in the analysis sample has the same magnitude as among unobserved clinical trials. Second, I assume that the clinical trial registry encompasses the full universe of trials conducted after 2005. I also assume that the analysis sample contains all registered trials that will be published.

Appendix figure A6 presents counterfactual estimates of the sponsorship effect accounting for publication bias. Adding just one of each of the unpublished trials reduces the sponsorship effect by 20%. However, there likely exist many additional unobserved trials. Under the assumption that each observed unpublished trial is one of four trials conducted, the sponsorship effect would decrease by about 50%. Under the assumption that each observed trial in the whole analysis sample is one of four trials conducted, the sponsorship effect would fall by about 90%.[28] Without this publication bias, the reported efficacy of sponsored drugs would fall by 0.15 standard deviations (90% of the sponsorship effect of 0.17 standard deviations), which is almost half of the average difference in efficacy between arms in a trial. There are large standard errors on these estimates. They rely on assumptions about the selection of unobserved trials, the share of trials pre-registered, and the share of trials with reported results, but are consistent with publication bias explaining a substantial share of the sponsorship effect.

In comparison, the point estimate for the share of sponsorship effect explained by trial design from table 6, column (8) is just 4% (0.006 off a base of 0.17).[29] The remaining unexplained share of the sponsorship effect may be attributed to underestimating the described publication channels, mean reversion, noise in these estimates, unobserved aspects of trial design, data manipulation or reconciliation errors.

5 Replicability Policy and Theory

5.1 Mitigation, Pre-Registration, and Reporting

One major policy in regulating clinical trials is pre-registration, which requires investigators to register their trials as a condition of publication or funding. Requirements often include pre-specifying outcomes, reporting results, and pre-registration prior to patient enrollment. Arguably the most significant of these

[27] To build intuition, suppose each funder of each observed trial actually conducted that trial four times. One trial is found and included in the analysis sample and three were buried. Under the assumption that the unobserved trials are similar to the observed but unpublished trials, I can re-create counterfactual samples. The sponsorship effect in these counterfactual samples is an estimate of the sponsorship effect without publication bias. This requires strong assumptions outlined below and should be considered a back-of-the-envelope exercise.

[28] The missing trials are all drawn from the set of unpublished trials. In the last counterfactual, this means that each unpublished trial is included 19 times in order to have four times the number of trials as in the analysis sample.

[29] However, the 95% confidence interval ranges from -38% to 45% of the sponsorship effect.

requirements is the ICMJE's agreement to only publish clinical trials in affiliated journals that were registered before patient enrollment. This condition applied to trials starting on July 1st, 2005; trials that began earlier had to be registered before journal submission by September 13, 2005 (De Angelis et al., 2004).

The proportion of published trials that are pre-registered on ClinicalTrials.gov increases gradually over time, as shown in figure 4A. Appendix figure A7 compares pre-registered and non-registered trials on trial and patient characteristics. Within a drug pair, pre-registered trials are statistically indistinguishable from non-registered trials in the number of patients, length, dosage, baseline severity, dropout share, age, or share female. They do occur one standard deviation, or about ten years, later which fits with the policy's implementation.

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:

$$y_{ij} = \alpha + Sponsor_{ij} + \sum_{y} \beta_y Sponsor_{ij} * y(j) + \sum_{y} y(j) + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{6}$$

where the sponsorship effect is interacted with publication year bins $y(j)$. The controls X_{ij} are indicators for the measurement scale. All other terms are the same as in equation 1.

Figure 4B plots the coefficients β_y on the sponsorship effect over time. The coefficients decrease in magnitude gradually after the 2005 pre-registration requirements, which fits with the gradual implementation of the policy.[30] Table 8 column (2) presents the sponsorship effect as estimated in equation 1, but fully interacted with an indicator for after 2005. The effect of sponsorship on reported drug efficacy is statistically significant and positive before required pre-registration and decreases after required pre-registration. The difference in the effect of sponsorship before versus after required pre-registration is statistically significant.

Additionally, 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 8 column (3) presents the sponsorship effect interacted with an indicator for whether the trial was pre-registered. The difference in the effect of sponsorship for pre-registered versus non-registered trials is statistically significant at the 10% level. This evidence is suggestive that pre-registration may be effective at mitigating conflicts of interest and publication bias.

[30] Appendix table A10 shows that the sponsorship effect dropping pre-1991 trials is smaller than the main sponsorship effect, which includes trials in the 1970s and 1980s. While the sponsorship effect is smaller in more recent years, earlier trials remain relevant in the stock of existing drugs. For example, two of the most common antidepressants used currently are fluoxetine (brand name Prozac) which was approved in 1987 and sertraline (Zoloft) which was approved in 1991.

At the same time that pre-registration was required, transparency and publication norms were also changing. Section 801 of the Food and Drugs Amendments Act, which requires results reporting for clinical trials was passed in 2007 and mandated compliance by April 18, 2017.[31] In figure 4B, the coefficient on sponsorship continues to drop with the passage of Section 801 after 2007. The share of pre-registered trials that are published also increased over time (Powell-Smith and Goldacre, 2016). Therefore, it is difficult to disentangle the effect of pre-registration from other norm changes and increased results reporting. To examine the role of pre-registration and increased publication rates, Appendix table A11 presents the sponsorship effect separately by whether the trial is pre-registered, published, both, or neither. The first row reports the baseline effect of sponsorship among trials that are not pre-registered and are published. Trials that are pre-registered have a lower sponsorship effect (P=0.112). Similarly, trials that are unpublished have a lower sponsorship effect (P<0.001). Finally, the additional effect of sponsorship among trials that are both pre-registered and unpublished also negative (P=0.103). This suggests that both improving publication rates (as in section 4.2.2) and pre-registration may reduce sponsorship bias.

5.2 Scaling Theory and Incentives

This paper finds that the treatment effects of clinical trials are substantially reduced when trials are not conducted by the drug's manufacturer. This is a version of the scaling problem, where treatment effects diminish in size when applied at a larger scale. The sponsorship effect is comparable to a scale-up drop. Theoretical results in the scaling literature have concluded that (1) increasing the reward for reporting a large treatment effect increases the magnitude of the scale-up drop, and (2) increasing the penalty for imperfect replicability decreases the magnitude of the scale-up drop (Al-Ubaydli et al., 2020).[32]

The potential reward for a large treatment effect in psychiatric clinical trials can be scientific or financial. In the first case, researchers might be particularly incentivized to find the first novel drug in a drug class. To test this theory, I plot the sponsorship effect for each drug relative to its novelty within a drug class. I compute the drug-specific sponsorship effect by estimating:

$$y_{ij} = \alpha + Sponsor_{ij} + \sum_d \eta_d Sponsor_{ij} * d(i) + X_{ij}\gamma + G_{d(i),p(j)} + \varepsilon_{ij} \tag{7}$$

where $d(i)$ is an indicator for each drug. Each term is the same as equation 1, but the $Sponsor_{ij}$ indicator

[31] Specifically, Section 801 stipulates that applicable clinical trials must register within 21 days after enrolling the first participant and report outcomes within one year after the primary completion date. However, compliance rates are estimated to be below 50%, and no fines have ever been imposed (Piller, 2020).

[32] There are four results in Al-Ubaydli et al. (2020), but two have ambiguous predictions for the scale-up drop.

is now interacted with each drug separately. Each antidepressant or antipsychotic drug belongs to a drug class: tricyclic, SSRIs, SNRIs, or atypical antidepressants and first or second generation antipsychotics (see appendix section A). For each drug, I compute the number of years between the first drug's approval in that class and the given drug's approval. The scientific novelty of a drug decreases with the number of years since the first approval in that class. Accordingly, Figure 5 shows that the sponsorship effect is negatively related to the year since the first drug approval in that class.

Turning to financial rewards, a measure of the financial reward for a large treatment effect is future prescriptions. If the potential market for a given drug is larger due to higher patient demand or fewer competitors, there might be additional incentives to obtain higher reported efficacy. Figure 6 plots the coefficients for each drug against a proxy for market size: the average number of MEPS prescriptions in the five years after FDA approval for that drug.[33] 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 and statistically significant correlation between the sponsorship effect and prescriptions shows the sponsorship effect is related to market factors and fits with theoretical results in scaling.

Al-Ubaydli et al. (2020) also show that increasing the penalty for imperfect replicability decreases the scale-up drop. Section 5.1 assessed the impact of a policy that increased the costs of not disclosing trials – required pre-registration. Consistent with this theory, I find the sponsorship effect decreased after the policy was enacted.

6 Conclusion

This paper demonstrates the impact of financial incentives on the reported outcomes of clinical trials. I find that a sponsored drug appears substantially more effective compared to the same drug tested in a trial with the same combination of drugs but without involvement from the drug manufacturer. Across a variety of specifications and outcomes, the sponsorship effect is large and consistently represents approximately half of the average difference in efficacy between trial arms. Publication bias explains most of this effect, while trial design and patient enrollment are less relevant. The remaining unexplained share of the sponsorship effect may be due to unobservable trial design characteristics, noise in the estimates, mean reversion, or data falsification.

The magnitude of the effect of funding on drug efficacy has substantial implications for drug approvals and prescriptions. The sample includes 23 FDA-approved drugs and seven non-approved drugs.

[33] The MEPS data begin in 1996. For drugs that were approved before 1996, I use the first five years of observed prescriptions, starting in 1996.

The relative efficacy of a drug in pre-approval trials strongly predicts FDA approval. If this relationship were causal and if drug efficacy decreased by the average sponsorship effect of 0.17 standard deviations, the approval rate would decline from 77% to 70%, resulting in two fewer approved psychiatric drugs. In terms of prescriptions, if the relationship between a drug's effectiveness and prescriptions in figure 6 were causal, then removing the average sponsorship effect from each drug would result in a 18% decrease in prescriptions. McKibbin (forthcoming) finds that after a statistically significant cancer trial is released, off-label prescriptions increase by 86%. This paper shows that sponsored arms are 10 percentage points more likely to report statistically significant improvements. Using McKibbin's estimate, this would translate to an 8.6% decrease in prescriptions without sponsorship. Fewer drug approvals and prescriptions may be good for welfare if consumers substitute to more effective drugs or alternative treatments.

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. However, even with current pre-registration requirements, only a quarter of all pre-registered trials report results. If trials without reported results were similarly selected to the observed unpublished trials, the estimated efficacy of these drugs would be lower than currently estimated, potentially influencing prescription decisions. Additionally, most existing antidepressant and antipsychotic drugs were approved prior to these requirements, so even with pre-registration requirements, there is a stock of existing drugs potentially based on biased evidence.

This paper focuses on financial incentives since these be quantified for a given drug and arm. Non-financial incentives may also be important in understanding drug efficacy. This paper also focuses on psychiatric medications. The difficulty in predicting treatment responses to these drugs could make sponsorship less significant in this setting. 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 with multiple substitutable drugs and variation in sponsorship.

My results are agnostic about the welfare consequences of different funding sources for clinical trials. The social beneficial of which parties conduct pharmaceutical research depends on how such restrictions might affect the amount of innovative research. Alternate funding schemes should also consider how sponsored clinical research is interpreted by physicians and patients, the availability of subsequent publications, and the external validity of clinical trials for different patients and settings. My findings on mechanisms show that sponsors affect the publication of trials and therefore the availability of knowledge. In terms of external validity, if funded trials targeted more effective populations or designed more effective trials, this could increase welfare. However, I find no evidence that sponsors target more effective populations or settings. Overall, this paper finds that the sponsor of a clinical trial significantly affects the reported efficacy of the drugs tested and restricts the availability of knowledge produced.

Data Availability

Code replicating the tables and figures in this article can be found in Oostrom (2024) in the Harvard Dataverse, https://doi.org/10.7910/DVN/R427A9.

References

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

Al-Ubaydli, Omar, John A. List, and Dana Suskind, "What Can We Learn From Experiments? Understanding the Threats to the Scalability of Experimental Results," *American Economic Review, Papers and Proceedings*, 2017, *107* (5), 282–86.

_, _, and _, "The Science of Using Science: Toward an Understanding of the Treats to Scalability," *International Economic Review*, 2020, *61* (4), 1387–1409.

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

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.

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.

Boulenger, Jean-Philippe, Henrik Loft, and Christina Olsen, "Efficacy and safety of vortioxetine (Lu AA21004), 15 and 20 mg/day," *International Clinical Psychopharmacology*, 2014, *29*, 138–149.

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.

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.

Camerer, Colin, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Júrgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikensten, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideo Nave, Thomas Pfeiffer, Michael Razen, and Hang Wu, “Evaluating Replicability of Laboratory Experiments in Economics,” *Science*, 2016, *351* (6280), 1433–1436.

Ching, Andrew T., Robert Clark, Ignatius Horstmann, and Hyunwoo Lim, “The Effects of Publicity on Demand: The Case of Anti-Cholesterol Drugs,” *Marketing Science*, 2016, *35* (1), 158–181.

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 H jgaard, 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,” *New England Journal of Medicine*, 2001, *345*, 825–827.

DiMasi, Joseph, Henry Grabowski, and Ronald Hansen, “Innovation in the Pharmaceutical Industry: New Estimates of R‘I&’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.

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.

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.

List, John A., *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*, Penguin Business, 2022.

McKibbin, Rebecca, "The Effect of RCTs on Demand for Off-Label Cancer Drugs," *Journal of Health Economics*, forthcoming.

Menkveld, Albert, Anna Dreber, Felix Holzmeisete, and et al., "Non-Standard Errors," *Journal of Finance*, forthcoming.

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.

Oostrom, Tamar, "Replication Data for: Funding of Clinical Trials and Reported Drug Efficacy," *Harvard Dataverse*, 2024. https://doi.org/10.7910/DVN/R427A9.

Open Science Collaboration, "Estimating the Reproducibility of Psychological Science," *Science*, 2015, *349* (6251), 1–8.

Østengaard, Lasse, Andreas Lundh, Tine Tjørnhøj-Thomsen, Suhayb Abdi, Mustafe H A Gelle, Lesley A Stewart, Isabelle Boutron, and Asbjørn Hróbjartsson, "Influence and Management of Conflicts of Interest in Randomised Clinical 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.

Powell-Smith, Anna and Ben Goldacre, "The TrialsTracker: Automated ongoing monitoring of failure to share clinical trial results by all major companies and research institutions [version 1; peer review: 2 approved]," *F1000Research*, 2016, *5* (2629).

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*, 2022, *14* (2), 439–477.

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.

Shelton, Richard, Kirsten Haman, Mark Rapaport, Ari Kiev, Ward Smith, Robert Hirschfeld, Bruce Lydiard, John Zajecka, and David Dunner, "A Randomized, Double-Blind, Active-Control Study of Sertraline Versus Venlafaxine XR in Major Depressive Disorder," *Journal of Clinical Psychiatry*, 2006, *67*, 1674–1681.

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

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.

Tables

Table 1: Types of Variation

Active vs. Placebo

Company A **Drug A** vs. Placebo	⟷	Unsponsored Drug A vs. Placebo

Active vs. Active

One Drug Never Sponsored

Company A **Drug A** vs. Drug B	⟷	Unsponsored Drug A vs. Drug B

One Drug Always Sponsored

Company A & Company B **Drug A** vs. **Drug B**	⟷	Company B Drug A vs. **Drug B**

Both Drugs Vary in Sponsorship

Company A **Drug A** vs. Drug B	⟷	Company B Drug A vs. **Drug B**

Notes: This table presents the different categories of variation in funding. 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 not bolded. Trials are only directly compared to the analogous trials in the same row.

Table 2: Effect of Sponsorship on Drug Efficacy

	Relative efficacy		Absolute efficacy	Significantly better at 0.05 level	Most effective in trial	% Decline
	(1)	(2)	(3)	(4)	(5)	(6)
$Sponsor_{ij}$	0.181***	0.171***	0.259**	0.104***	0.283***	0.019***
	(0.054)	(0.052)	(0.103)	(0.040)	(0.055)	(0.007)
Controls		X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X
Mean Outcome	0.35	0.35	0.06	0.24	0.39	0.05
N	1,990	1,990	1,990	1,741	1,990	1,816
$Weighted\,N$	1,215	1,215	1,215	1,087	1,215	1,085

Note: This table presents the coefficients from the estimation of equation 1, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. In columns (1) and (2), the dependent variable y_{ij} is the standardized efficacy measure, relative to the placebo arm in that drug pair if available or least effective arm otherwise. In column (3), the outcome is the standardized absolute efficacy measure. The outcome in column (4) is an indicator for whether arm i in trial j was found to be statistically significantly more effective than the other arms in that trial at the 0.05 level. In column (5), the outcome is an indicator for whether arm i was the most effective arm in trial j. The outcome in column (6) is the relative percent decline in the psychotic score. 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 3: Robustness of Sponsorship Effect

		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
	(1)	(2)	(3)	(4)	(5)	(6)
$Sponsor_{ij}$	0.171***	0.160***	0.135***	0.145***	0.147**	0.100**
	(0.052)	(0.052)	(0.052)	(0.054)	(0.057)	(0.041)
Controls	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X
Mean Outcome	0.35	0.35	0.35	0.43	0.35	0.30
Weighted N	1,215	1,215	1,215	795	1,215	1,215

Note: This table presents coefficients from the estimation of equation 1, where the fixed effects $G_{d(i),p(j)}$ control for each drug in each drug pair. Column (1) replicates the baseline estimate from table 2, column (2), where the outcome is relative efficacy. The dependent variable is the same in all subsequent columns. Column (2) includes controls for the order that the trial occurred within the drug pair, while column (3) includes controls for the year the trial was published relative to the drug approval year. Column (4) restricts the sample to exclude trials that were published before one of the drugs in the trial was approved by the FDA. Column (5) excludes trials for which the only sponsorship indication is a conflict of interest (COI) statement. Column (6) weights each trial's arm by the total enrollment in that arm. Standard errors are clustered at the trial level and reported in parentheses, with ${}^{*}p < 0.10$, ${}^{**}p < 0.05$ and ${}^{***}p < 0.01$.

Table 4: Heterogeneity of Sponsorship Effect

	Drug Pair Type		Drug Class		Variation Type	
	Active vs. Placebo	Active vs. Active	Anti-depressant	Anti-psychotic	Industry vs Non-Industry	Industry vs Industry
	(1)	(2)	(3)	(4)	(5)	(6)
$Sponsor_{ij}$	0.270***	0.124**	0.215***	0.092	0.159***	0.250**
	(0.103)	(0.058)	(0.068)	(0.061)	(0.059)	(0.107)
Controls	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X
Mean Outcome	0.49	0.25	0.36	0.31	0.41	0.30
Weighted N	520	695	900	315	541	674

Note: This table presents coefficients on $Sponsor_{ij}$ from the estimation of equation 1 for subsamples of the data. Column (1) restricts to drug pairs that compare one active drug to a placebo. Column (2) restricts to drug pairs that compare two active drugs. Each drug pair is in one of these two categories. Columns (3) and (4) split the sample by the drug type: antidepressant or antipsychotic. Column (5) restricts to drug pairs that compare industry-funded trials to at least one unsponsored trial. Column (6) restricts to drug pairs that only compare industry-funded trials. 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 5: Trial and Patient Characteristics

		Additional Controls			Within		
	Baseline	Trial Chars.	Patient Chars.	Trial and Patient Chars.	Dose	Dose, Age, Gender	Dose, Age, Gender, Baseline Severity
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$Sponsor_{ij}$	0.171***	0.178***	0.158***	0.163***	0.160**	0.163**	0.194**
	(0.052)	(0.052)	(0.052)	(0.051)	(0.073)	(0.076)	(0.095)
Controls	X	X	X	X	X	X	X
Drug by Drug Pair F.E.	X	X	X	X	X	X	X
Mean Outcome	0.35	0.35	0.35	0.35	0.35	0.35	0.35
Weighted N	1,215	1,215	1,215	1,215	1,215	1,215	1,215

Note: Column (1) replicates the main result from table 2, columns 2. Column 2 includes controls for trial characteristics: the length of the trial in weeks, number of patients, and 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. Columns (5)-(7) present the coefficients on $Sponsor_{ij}$ from the estimation of equation 2, where the fixed effects $G_{d(i),p(j),k(i)}$ control for each drug in each drug pair within each characteristic. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 6: 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	0.01	-0.02	0.01
	(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	
Mean Outcome	0.21	0.24	0.22	0.11	0.26	0.30	0.29	0.32
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 4, where the dependent variable is predicted drug efficacy. Each column predicts drug-specific efficacy using different trial characteristics, as shown in equation 3, 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 bootstrapped using 100 repetitions, drawing trials with replacement and are reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 7: Publication and Efficacy

	Published	
	Sponsored	Not Sponsored
	(1)	(2)
Relative Efficacy	0.149***	0.029
	(0.029)	(0.033)
Controls	X	X
Drug by Drug Pair F.E.	X	X
Mean Outcome	0.85	0.85
Weighted N	681	731

Note: Columns (1) and (2) present the coefficients from the estimation of equation 5, where the outcome is an indicator for whether the trial was published. Controls include the type of psychiatric score used. Column (1) restricts to the sample to sponsored arms, while column (2) restricts the sample to not sponsored arms. Standard errors are clustered at the trial level and reported in parentheses, with $^{*}p < 0.10$, $^{**}p < 0.05$ and $^{***}p < 0.01$.

Table 8: Sponsorship Effect After Pre-Registration

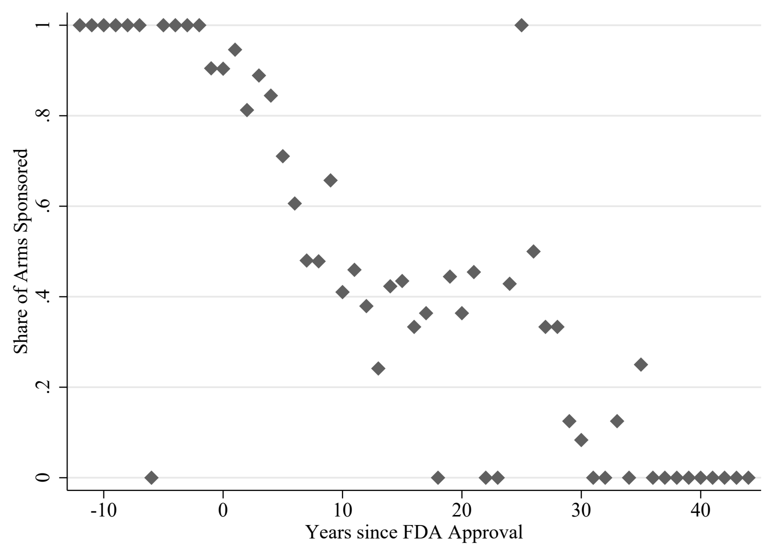
	Relative Efficacy		
	(1)	(2)	(3)
$Sponsor_{ij}$	0.171***	0.221***	0.190***
	(0.052)	(0.059)	(0.053)
Post 2005		-0.084	
		(0.178)	
$Sponsor_{ij}$ x Post 2005		-0.155**	
		(0.068)	
Pre-Registered			0.053
			(0.045)
$Sponsor_{ij}$ x Pre-Registered			-0.190*
			(0.103)
Controls	X	X	X
Drug by Drug Pair F.E.	X	X	X
Mean Outcome	0.33	0.33	0.33
Weighted N	1,215	1,215	1,215

Note: Table presents the coefficients from the estimation of equation 1 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 $^{***}p < 0.01$.

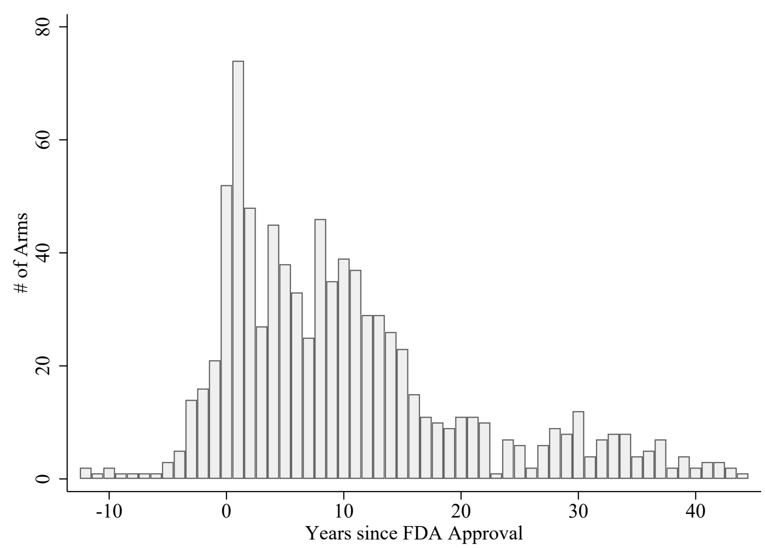
Figures

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

(A) Sponsorship



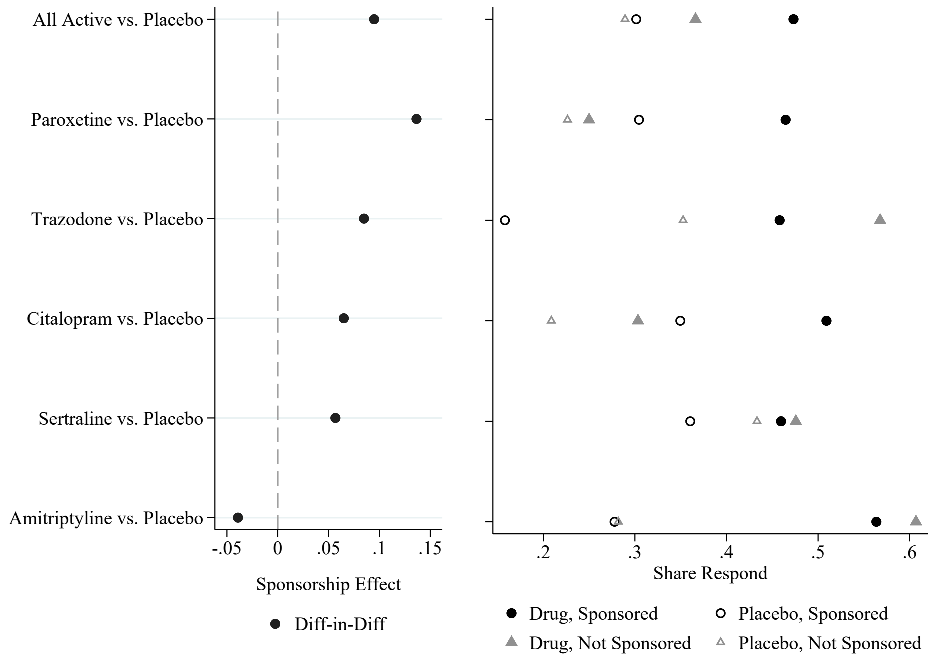
(B) Sample Size



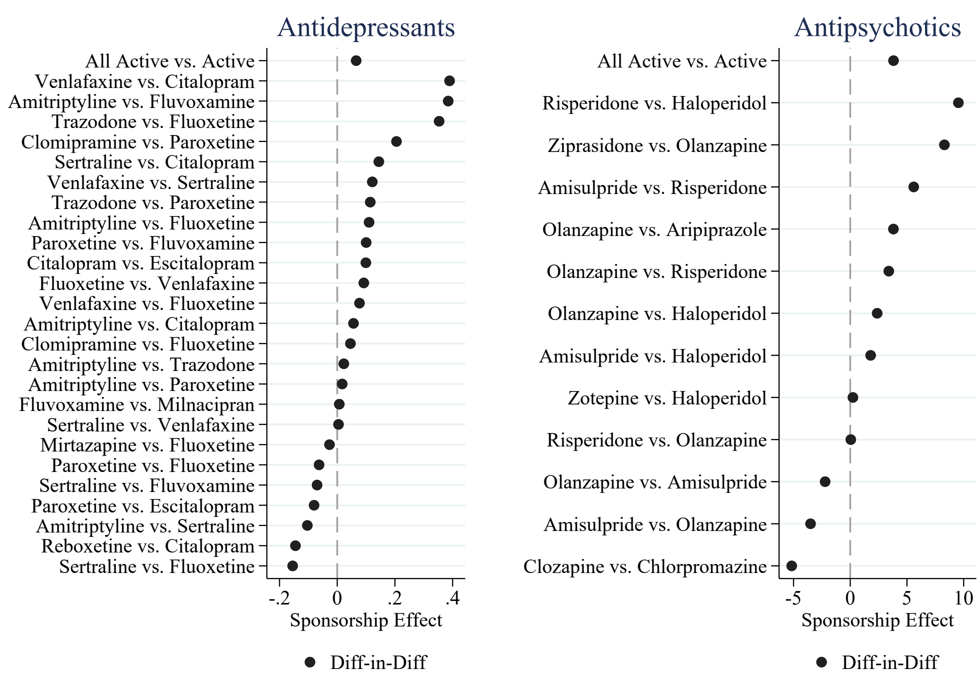
Notes: Panel A plots information on sponsorship over time. The x-axis plots the number of years since FDA approval for a given drug. The y-axis plots the share of those arms that are sponsored. This figure excludes placebo arms and drugs that are not approved by the FDA (agomelatine, amisulpride, milnacipran,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

Figure 2: Difference in Difference Framework

(A) Active vs Placebo

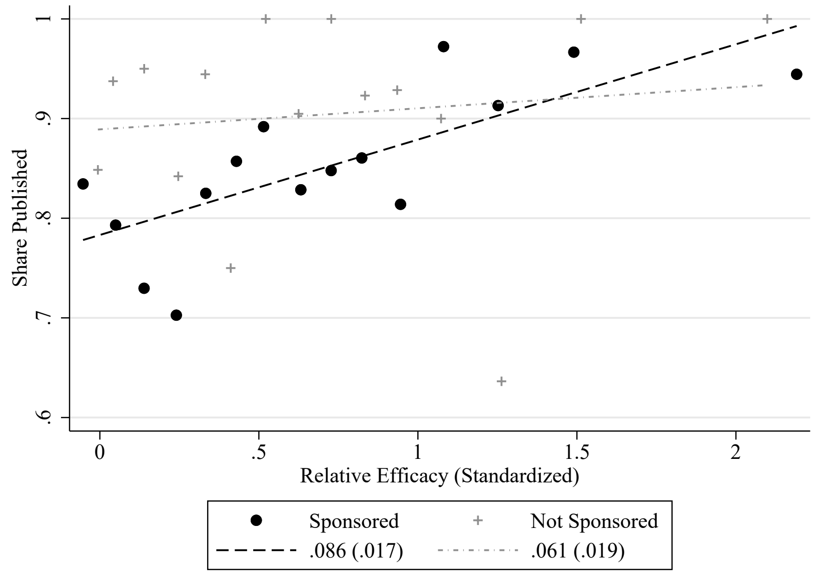


(B) Active vs Active



Notes: This figure presents the difference-in-difference estimate of the sponsorship effect within drug sets. Each row represents a drug set, where the first listed drug varies in sponsorship across trials and the second listed drug has the same sponsorship status in all trials. Panel A presents estimates for the active versus placebo drug sets, which are all antidepressants. The dots represent the average efficacy of the first listed drug when it is sponsored (solid black circle) versus not sponsored (solid gray triangle), versus the placebo in trials where the first drug is sponsored (hollow black circle) or not sponsored (hollow gray triangle). The black dots represent the difference-in-difference estimate computed from the maroon and gray points. Panel B presents these estimates for the active versus active drug sets. Efficacy for antidepressants and antipsychotics are measured on different scales and therefore vary in magnitude.

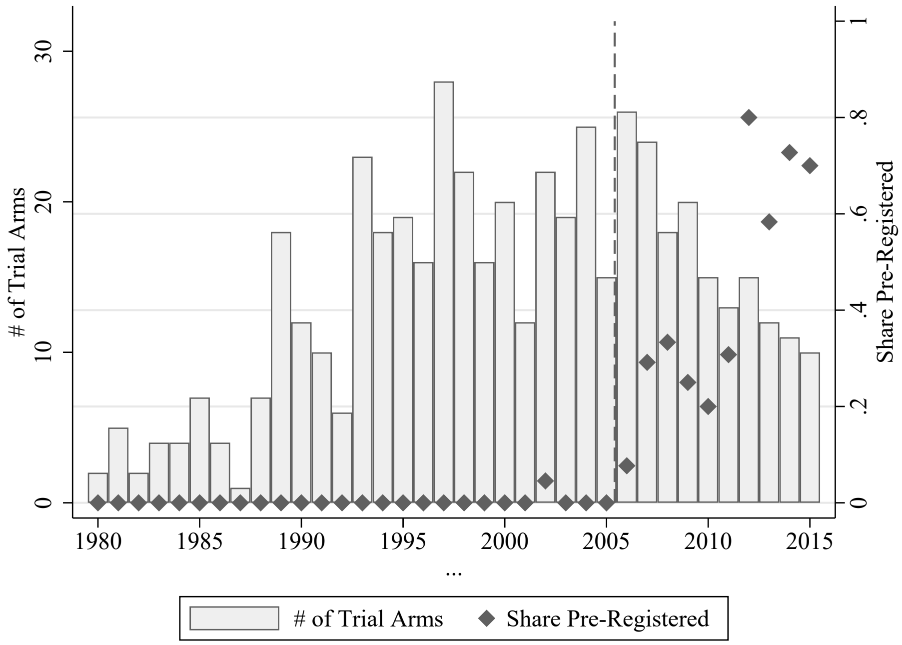
Figure 3: The Relationship Between Efficacy and Publication



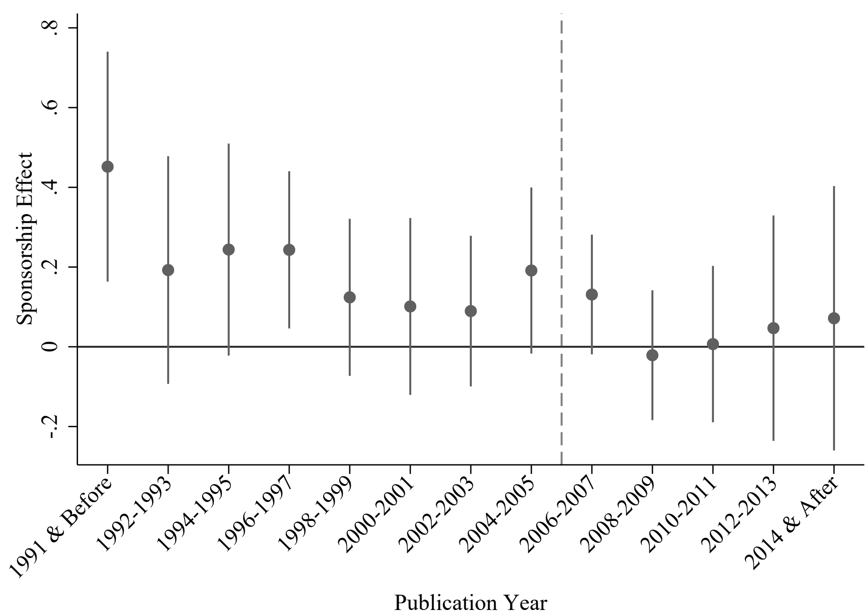
Notes: This figure presents the relationship between effectiveness and publication status. The x-axis plots the standardized relative efficacy. Efficacy is binned based on the whole sample. Bins are equally sized, when possible, though observations with the same x-value must be in the same bin. The y-axis presents the probability that arms in the given efficacy bin are published. The dashed lines represent the best fit lines. I report the coefficient on relative efficacy from the regression of an indicator for published on relative efficacy separately for sponsored and non-sponsored arms. This regression is at the arm level. No controls are included.

Figure 4: Introduction of Clinical Trial Pre-registration

(A) Pre-Registration by Calendar Year

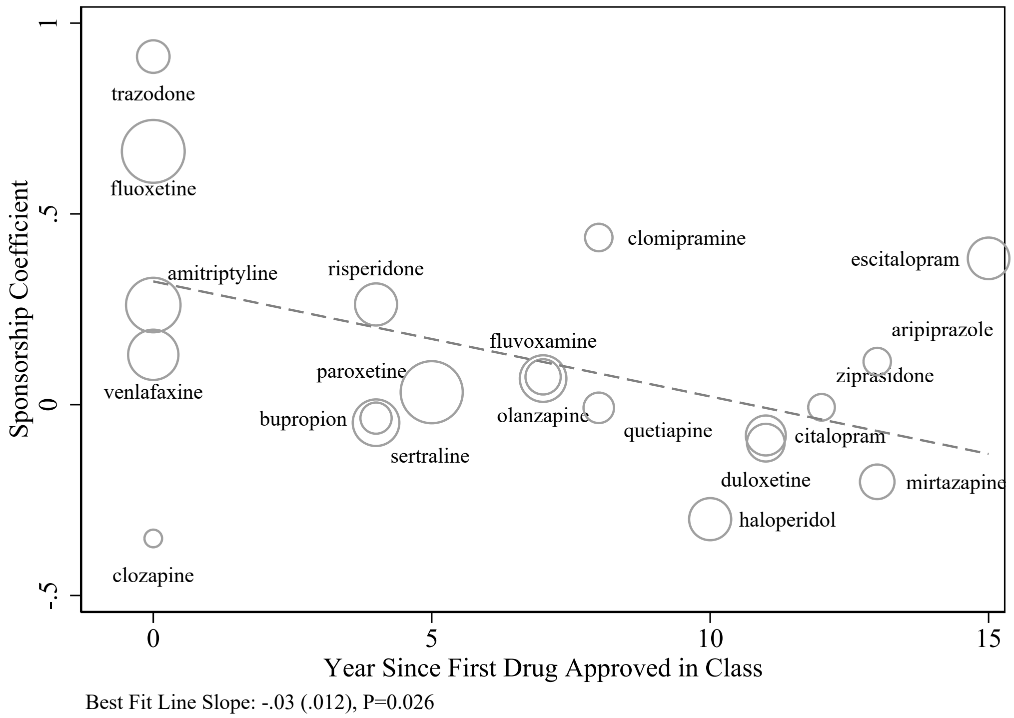


(B) Sponsorship Effect by Calendar Year



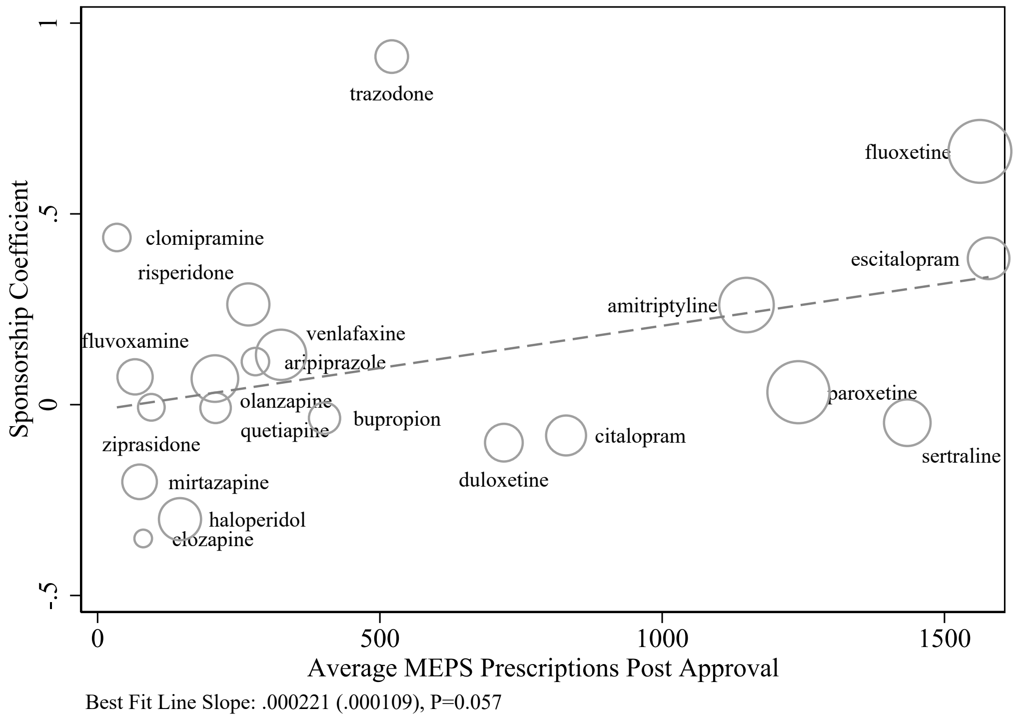
Notes: Panel A plots the share of antidepressant trials in the analysis sample that were pre-registered in ClinicialTrials.gov. The gray bars plot the sample size of treatment arms by publication year. The vertical dashed line midway between 2005 and 2006 represents July 1st, 2005, when the International Committee on Medical Editors agreed to only publish clinical trials that had been registered before patient enrollment. In 2007, Section 801 of the Food and Drugs Amendments Act was passed, which nominally required results reporting. Panel B presents the coefficients β_y from the estimation of equation 6. Standard errors are clustered at the trial level.

Figure 5: Year Since Drug First Approved in Class



Notes: The x-axis plots the number of years between FDA approval and the year the first drug in that class was approved. The y-axis plots the sponsorship coefficient for each drug from the estimation of equation 7. The best-fit line is plotted in gray. Each point is weighted according to the number of arms for that drug.

Figure 6: Sponsorship Effect and Drug Sales



Notes: This figure plots the coefficient on sponsorship for each drug from the estimation of equation 7 against the average annual number of MEPS prescriptions in the five years post-approval for that drug. The best-fit line is plotted in gray. Each point is weighted according to the number of arms for that drug.