Identifying Placebo Effects with Data from Clinical Trials

Anup Malani

University of Virginia

A medical treatment is said to have placebo effects if patients who are optimistic about the treatment respond better to the treatment. This paper proposes a simple test for placebo effects. Instead of comparing the treatment and control arms of a single trial, one should compare the treatment arms of two trials with different probabilities of assignment to treatment. If there are placebo effects, patients in the higher-probability trial will experience better outcomes simply because they believe that there is a greater chance of receiving treatment. This paper finds evidence of placebo effects in trials of antiulcer and cholesterol-lowering drugs.

I. Introduction

A medical treatment is said to have a placebo effect if a patient's response to treatment depends positively on his or her expectations about the value of that treatment. In other words, placebo effects cause more optimistic patients to respond better to treatment than less optimistic patients.

There is naturally a great deal of interest in placebo effects in the

I thank Steven Levitt and two anonymous referees for their detailed feedback. I also thank Gary Becker, James Heckman, Dan Houser, Feifang Hu, Darius Lakdawalla, Willard Manning, Ron Michener, Charles Mullin, Tomas Philipson, and Andrew Sellgren; workshop participants at the University of Virginia, RAND, the University of Southern California, the University of California at Irvine, the University of Chicago, and George Mason University's Interdisciplinary Center for Economic Sciences; and conference participants at the 2004 North American summer meeting of the Econometric Society for useful comments. Finally, I am grateful to David Greene, Maulshree Solanki, Hema Srinivasan, and Jose Fernandez for their research assistance.

[*Journal of Political Economy*, 2006, vol. 114, no. 2]
© 2006 by The University of Chicago. All rights reserved. 0022-3808/2006/11402-0002$10.00

medical sciences. Placebo effects are thought to explain a good deal of premodern medicine (Shapiro and Shapiro 1997), and concerns about placebo effects drove the introduction of controlled medical trials (Kaptchuk 1998) and blinded medical trials (Vickers and de Craen 2000). Many in the medical community believe that complementary and alternative medications (Miller et al. 2004) and perhaps even antidepressants and certain analgesics (see, e.g., Kirsch et al. 2002; Zubieta et al. 2005) operate mainly through placebo effects.

Placebo effects are also relevant to the social sciences, especially economics. For one thing, placebo effects may be a behavioral rather than a physiological phenomenon. More optimistic patients may modify their behavior—think of the ulcer patient who reduces his or her consumption of spicy food or the cholesterol patient who exercises more often—in a manner that complements their medical treatment. If an investigator does not measure these behavioral changes (as is commonly the case), the more optimistic patient will appear to have a better outcome, that is, to have experienced placebo effects.

Even if placebo effects are a physiological phenomenon, for which there is some evidence (Wager et al. 2004; Zubieta et al. 2005), they raise a number of interesting economics questions. For one thing, if beliefs can be manipulated more cheaply than drugs can be produced, might they not reduce health care costs by allowing doctors to substitute from expensive medications to words of inspiration (Talbot 2000)? Second, beliefs that are responsible for placebo effects are endogenous because patients can regulate their research on a medication so as to modify their beliefs about the effects of that medication. The question is, Do patients invest in research so as to develop optimistic beliefs that maximize their returns from treatment, an admittedly novel form of self-production of health care?

Even if beliefs cannot be manipulated, placebo effects may cause a violation of the independence axiom. In the context of blinded clinical trials, different states of the world correspond to different treatments. Because the patients do not know their treatment assignment, their placebo effects are keyed to their beliefs about expected treatments and thus about expected states. As a result, patients' beliefs about states enter their utility in any given state; that is, utility is state-dependent.

Despite their import, it is uncertain whether placebo effects even exist. On one side of the debate is, for example, a prominent *New England Journal of Medicine* article (Hróbjartsson and Gøtzsche 2001) that examines 114 medical trials with both a blinded placebo control group and an unblinded no-treatment group and finds few systematic differences in outcomes between these groups (see also Hróbjartsson and Gøtzsche 2004). This result does not disprove the existence of placebo effects, however, because members of unblinded no-treatment groups

may have sought out alternative medication that elevated their health outcomes.

On the other side of the debate are, for example, studies (Kirsch and Sapirstein 1998; Kirsch et al. 2002) that highlight evidence that placebo control groups in trials often manifest substantially improved health outcomes. These findings are weak support for placebo effects because the improvements could be due to the natural progression of disease.[1] Better studies (Penick and Hinkle 1964; Penick and Fisher 1965; Marlatt and Rohsenow 1980; Kirsch and Weixel 1988) employ a balanced placebo design wherein patients are first randomized across treatments and then across instructions about the value of treatment or about whether they obtained active treatment or placebo. Unfortunately, not only are the results of these studies mixed, perhaps because of a small sample size, but their design is ethically questionable because it requires deception. Moreover, this design cannot be implemented on a large scale because subjects may rationally begin disbelieving instructions from clinical investigators.[2]

In this paper I present a simple insight about medical trials that suggests a novel test for placebo effects that overcomes the objections to prior tests for placebo effects. If a researcher has a single-blinded, randomized, controlled medical trial and she compares outcomes in the treatment arm with outcomes in the control arm, a "within-trial" comparison, it is well known that she will obtain the pharmacological effect of treatment. Suppose, however, that the researcher also has a second clinical trial of the same medication, with the same control and design. The only difference is that the second trial employs a higher probability of assignment to treatment. The main implication of this modification is that patients in the higher-probability trial will believe that they are more likely to receive treatment. It also implies that patients in the higher-probability trial will expect better health outcomes from their clinical trial, all other things being equal.

Consider what happens if the researcher, instead of comparing groups within a trial, compares matched groups across trials, for example, the treatment group of the first trial with the treatment group of the second

[1] Note the difference between "placebo control" and "placebo effects." Placebo control refers to an inert pill. Placebo effects refer, however, to a type of health response. A placebo control may trigger and thus capture placebo effects (due, e.g., to beliefs), but it also captures the effect of natural progression. Only when natural progression is known to be zero does positive response to placebo control demonstrate placebo effects.

[2] There are studies (Skovlund 1991; Amanzio and Benedetti 1999; Benedetti, Arduino, and Amanzio 1999; Pollo et al. 2001; Wager et al. 2004; Zubieta et al. 2005) on pain medications that are more compelling. Unfortunately, with two exceptions (Wager et al. 2004; Zubieta et al. 2005), outcomes are measured by subjective self-reports of pain by patients. Moreover, the sample sizes of all the trials are small, and all but one of the studies (Skovlund 1991) require deception.

TABLE 1
DIFFERENCE IN OUTCOMES IN TREATMENT ARMS, BY PROBABILITY OF TREATMENT: PROBABILITY 0.5 AND ONE TRIALS ONLY

Treatment	Control	Outcome	Probability of Treatment: 0.5	Probability of Treatment: One	Difference
		Nongastric Ulcer			
H_2 blockers	Placebo and/or palliative	Healed (0/1)	.697	.807	.110
			(.004)	(.001)	(.004)
			2,604	10,490	
PPIs	Conventional therapy	Healed (0/1)	.797	.812	.016
			(.004)	(.005)	(.006)
			2,127	1,115	
		High Cholesterol			
Statins	Placebo	Reduction in LDL level (mg/dl)	51.425	66.027	14.601
			(.346)	(.317)	(.469)
			1,564	5,305	
		Any side effects (0/1)	.667	.990	.323
			(.012)	(.001)	(.012)
			1,540	5,105	
		Usual side effects (0/1)	.066	.316	.251
			(.006)	(.007)	(.009)
			1,540	5,105	

NOTE.—Each cell contains the mean outcome, the standard error of the mean (in parentheses), and the number of patients over which the mean is calculated. All statistics are calculated from arm-level means, standard deviations, and sample sizes. For binary outcomes, only arm-level means and sample sizes are necessary.

trial. The "across-trial" comparison employs the same pharmacological intervention but varies the probability of treatment and thus the patients' expectations about outcomes. If the medication being studied has placebo effects, patient outcomes would be greater in the higher-probability trial than in the lower-probability one, holding constant the pharmacological intervention. In other words, the treatment (control) group of the higher-probability trial should perform better than the treatment (control) group of the lower-probability trial. One can predict the magnitude of placebo effects outside of medical trials by dividing the change in outcomes across two trials by the difference in probabilities of treatment across the two trials. This test can easily be generalized to the case of more than two trials with more than two probabilities of treatment.

I apply this test to a sample of over 200 medical trials of ulcer medications (H_2 blockers and proton pump inhibitors [PPIs]) and over 30 medical trials of cholesterol medications (statins). The advantage of ulcer and cholesterol trials is that outcomes are objectively measured by endoscopy and blood tests, respectively. Table 1 provides a preview of the results. It presents mean outcomes in trials in which half of the patients are randomized to treatment and in trials in which all patients

are randomized to treatment. Ulcer patients in the latter trials had a 0.11 higher probability of healing following treatment with H_2 blockers and a 0.02 higher probability of healing following treatment with PPIs. Cholesterol patients in all-treated trials experienced both a 14 milligram per deciliter (mg/dl) additional reduction in low-density lipoprotein (LDL) levels—the "bad" cholesterol—and a 0.25–0.32 higher probability of side effects in all-treated trials than patients in half-treated trials. Each of these differences is statistically significant. Regression analysis confirms most of these findings. It reveals that a higher probability of treatment is associated with a significant increase in the rate of ulcer healing in the treatment arms of H_2 blocker trials and in the amount of LDL reduction and side effects in the treatment arms of trials of statins. These correlations imply that placebo effects from, for example, the dominant H_2 blockers (Zantac and Tagamet) and the dominant statins (Lipitor and Zocor) are roughly one-third of the size of the pharmacological effects of these drugs.

This paper is organized as follows. Section II presents the theoretical foundation for my empirical test for placebo effects. Section III applies the test to trials of antiulcer and cholesterol-lowering medications.

II. Theory

This section presents a simplified model of placebo effects and patients' beliefs in medical trials in order to motivate my empirical test for placebo effects. Consider a binary health outcome, such as a healed or an unhealed ulcer. A general formulation of the relationship between consumption of some treatment, k, and the outcome, y, is $y = f(p_k, \pi_k)$, where p_k is the pharmacological effects of treatment k and π_k is the patient's belief about the pharmacological effects of k. Although both p_k and π_k might themselves be distribution functions, for simplicity I assume that p_k and π_k are scalar probabilities: the former is the probability of healing due to the pharmacological effects of k and the latter is the patient's point estimate of the probability of healing following consumption of k.

If treatment k does not have any placebo effects, then the patient's beliefs about its efficacy do not affect his health outcome, that is, $\partial y/\partial \pi_k = 0$. If treatment k does have placebo effects, then $\partial y/\partial \pi_k > 0$. Because it will facilitate a clearer exposition of the core intuition of my analysis, I adopt a linear, additively separable specification of the health outcome function:

$$y = (1 - \alpha)p_k + \alpha\pi_k, \tag{1}$$

where $\alpha \in [0, 1]$ describes the relative importance of beliefs in deter-

TABLE 2
PHARMACOLOGICAL EFFECTS AND BELIEFS IN BLINDED, RANDOMIZED, CONTROLLED TRIALS

Treatment Group	Probability of Treatment in a Trial		Across-Trial Comparisons
	d	$d' > d$	
k	p_k, $d\pi_k + (1-d)\pi_0$	p_k, $d'\pi_k + (1-d')\pi_0$	0, $\Delta d \Delta \pi$
0	p_0, $d\pi_k + (1-d)\pi_0$	p_0, $d'\pi_k + (1-d')\pi_0$	0, $\Delta d \Delta \pi$
Within-trial comparisons	Δp, 0	Δp, 0	

NOTE.—The first term in a cell gives the pharmacological effect experienced by that group. The second term gives the beliefs of patients in that group.

mining health outcomes. In this formulation, treatment k has placebo effects if and only if $\alpha > 0$.

In order to identify whether treatment k generates placebo effects, a researcher would ideally like an intervention that manipulated an individual's beliefs π_k while holding constant the pharmacological effects p_k. Note that a single-blinded, randomized, controlled medical trial—the ideal randomized experiment—does exactly the opposite. It manipulates a patient's treatment but holds constant the patient's beliefs about his likely health outcome. The patient is randomized to either treatment k or a control, with index 0, causing him to experience a pharmacological effect p_k or p_0, respectively. In either case, because the patient is blinded to his treatment assignment, he has identical beliefs about his likely outcome:

$$d\pi_k + (1-d)\pi_0, \tag{2}$$

where d is the probability of being assigned to active treatment and π_0 is the patient's belief about the efficacy of a placebo.[3]

The central insight of this paper is that if the researcher has two medical trials with different probabilities of treatment, d and $d' > d$, she can introduce the sort of variation in beliefs required to identify placebo effects. For illustration, table 2 presents the pharmacological effects and patient beliefs in each arm of d and d' trials. If the researcher, instead of comparing treatment and control groups "within" a trial, compares either treatment groups "across" d and d' trials or control groups across the trials, she will observe the effects of manipulating beliefs while holding pharmacological effects constant. This would allow her to identify the existence of local average placebo effects,

$$\int_d^{d'} \left[\frac{\partial f(p_j,\ \pi(s))}{\partial \pi}\right] ds$$

[3] Patients learn the probability of treatment in a trial during informed consent. This is required by law (see, e.g., U.S. Code of Federal Regulations 21, §§7.3(f), 20.25).

for $j = \{0, k\}$. By plugging blinded-trial beliefs (2) into the linear health specification (1), one can easily see that the sign of α is identical to the sign of the coefficient $(\alpha[\pi_k - \pi_0])$ on a regression of outcome (by arm) on probability of treatment d. If this term is positive, treatment k is subject to placebo effects.[4] The magnitude of placebo effects outside the trial is given by $(y_{kd'} - y_{kd})/(d' - d)$. This equation provides the linear prediction of the change in outcomes in the treatment group as the probability of treatment goes from zero (known consumption of no treatment) to one (known consumption of treatment k).[5]

A natural concern with this test is whether it is robust to self-selection of patients into clinical trials.[6] In other words, do trials with different probabilities of treatment attract the same patients, and if not, does a comparison of treatment groups properly identify placebo effects? A problem would arise if the higher-probability trial attracts patients who would have a better response to both treatment and placebo. In that case, one could not determine whether the test I propose identified self-selection or placebo effects. To explore this problem, I assume that patients choose to enroll in a clinical trial only if it offers to improve their expected health outcomes relative to therapies available outside the trial, and that the treatment k studied in the d and d' trials is not available outside the trial, as is usually the case.[7] An immediate and important implication of these assumptions is that I can ignore patients who believe that conventional therapy (if one is available) or no treatment is better than treatment, that is, patients for whom $\pi_k < \pi_0$. These patients will never enroll in a trial, regardless of the probability of treatment.

On the basis of my assumptions about selection, there are two possible

[4] This conclusion assumes $\pi_k - \pi_0 > 0$. I shall demonstrate this when discussing patient self-selection into clinical trials. This basic test for placebo effects—searching for a positive relationship between the probability of treatment and outcomes conditional on treatment assignment—generalizes to the case of continuous outcomes (including side effects) and to the case of nondegenerate distributions of pharmacological effects and beliefs about pharmacological effects. It even generalizes to the case in which placebo effects are driven not by beliefs about the pharmacological effects of treatment, but by the total effects—including placebo effects—of treatment. Readers can find these results in the dissertation on which this paper is based (Malani 2003).

[5] Note that the linear projection should not in general be based on outcomes in the control group. A linear projection based on control group outcomes would give the effect of a change in expectations about treatment given consumption of the control. The placebo effect that is of interest, however, is the effect of a change in expectation about treatment given consumption of treatment. That will be different unless, of course, placebo effects are independent of pharmacological effects. The results in the next section suggest otherwise.

[6] When one is estimating pharmacological effects with within-trial comparisons, bias from self-selection into the randomized trial is called randomization bias (Heckman 1992).

[7] If the treatment were available outside the trial, enrollment could be secured only by payment or a promise of free health care. But then there would be no self-selection unless wealth was correlated with beliefs about the effects of treatment.

situations that could arise. In the first situation, there is no conventional or alternative therapy available outside the trial or the control group of the trial receives this conventional therapy. (The ulcer and cholesterol trials I examine in Sec. III fall into this category.) In this situation, all patients who believe that the treatment is better than the control will enroll in the clinical trial regardless of the probability of treatment.[8] The reason is that the only outside option—no treatment—is no better than the "bad" outcome in the trial—treatment with placebo. Because both high- and low-probability trials attract the same population, there is no complication from self-selection.[9]

In the second situation, there is a conventional therapy available outside the trial and the control group receives placebo therapy. Because the bad outcome in the trial is now worse than the outside option (i.e., placebo is worse than conventional therapy), a patient's decision to enroll will depend on beliefs about the relative value of the treatment and conventional therapy and the probability of receiving treatment in the trial. In particular, the worse a patient believes the treatment is, the higher the probability of treatment must be in order to induce him to enroll. In the extreme, if the patient believes that the treatment is infinitely (epsilon) better than conventional therapy, a probability of treatment equal to epsilon (one) is required to induce enrollment. This suggests that the high-probability trial will attract all the patients the low-probability trial would and, in addition, some patients who are less optimistic about treatment.[10] If patients are even remotely competent at predicting their own pharmacological response to treatment, the additional patients will experience lower pharmacological effects from treatment. Therefore, although a higher probability of treatment elevates expectations about treatment for *any given patient*, it also attracts a *sample* of patients with lower expectations, which in turn suppresses both pharmacological effects and placebo effects. Nonetheless, if one observes better health outcomes in the high-probability trial than in the

[8] Without loss of generality, I assume that all patients believe that the conventional therapy is better than placebo control, i.e., no therapy. I can ignore patients who believe that conventional therapy is better than treatment; none of these individuals will enroll at any probability of receiving treatment.

[9] I assume that, if there are more subjects who want to enroll in a trial than there are spots available in the trial, the researcher will select patients by random draw. In other words, the problem I address is that of self-selection by patients, not selection by the researcher. There is no suggestion in the medical statistics literature that researchers select patients with different beliefs depending on the probability of treatment in a trial.

[10] For a systematic analysis of this case, see Malani (2005). An important, but less than obvious, assumption behind this intuition is that patients' beliefs about treatment are not too highly correlated with their beliefs about conventional therapy. This condition is satisfied, however, whenever the variance of beliefs about the treatment is greater than those about conventional therapy. This is likely to be the case because the treatment is newer than the conventional therapy.

low-probability trial, it must be that patients experienced placebo effects. Without placebo effects, sample selection would cause the higher-probability trial to produce, if anything, worse outcomes, not better. In other words, self-selection introduces type 1 error (false negatives) into my proposed test for placebo effects but does not sap all its power.

III. Empirical Analysis

This section applies my test for placebo effects to data from ulcer and cholesterol trials. The ulcer data set comprises published results from over 200 trials studying the effect of H_2 blockers and PPIs on nongastric ulcers. These ulcers are the erosion of the mucous lining in the lower stomach or small intestine and are judged to be healed—a binary outcome—via endoscopy by the researcher. The first class of ulcer medication, H_2 blockers, was introduced in 1977; the most popular brands are Tagamet (cimetidine), Zantac (ranitidine), and Pepcid (famotidine). The second class, PPIs, was introduced in 1989; the most popular brands are Prilosec (omeprazole), Nexium (esomeprazole), and Prevacid (lansoprazole). Both classes of medication are thought to prevent the production of acid in the stomach.[11]

The cholesterol data set comprises published results from over 30 trials examining the effect of statins on LDL—the "bad" cholesterol. High levels of LDL (> 160 mg/dl) increase the risk of developing atherosclerosis, the hardening and narrowing of arteries, which may cause stroke, heart failure, and loss of limbs. Statins, which were first introduced in 1987, block the formation of an enzyme in the liver that is needed to make cholesterol. The main types in my data are Lipitor (atorvastatin), Zocor (simvastatin), Pravachol (pravastatin), Mevacor (lovastatin), and Lescol (fluvastatin).

A valuable feature of both data sets is that they are unlikely to suffer bias from self-selection. Although the H_2 blocker and statin trials all have placebo (or palliative) controls, they were each the first meaningful treatments for ulcers and high cholesterol. Moreover, while PPIs were introduced after H_2 blockers, all the PPI trials in the sample employ H_2 blockers as controls. Therefore, a strong argument can be made that the entire population of patients would have wanted to enroll in each

[11] A distinguishing feature of both classes of medication is that they offer a much higher chance of healing an ulcer than antacids, bismuth subcitrate, or carbenoxolone, which are mainly palliatives. That being said, it is now recognized—most recently by the 2005 Nobel Prize in medicine—that 90 percent of nongastric ulcers are caused by the bacterium *Helicobacter pylori*. These infections are usually treated with a combination of antibiotics and H_2 blockers or PPIs. This paper tests for placebo effects in trials in which H_2 blockers or PPIs are used in isolation. These trials typically predate the change to antibiotic-based treatments.

of the trials in my data regardless of the probability of treatment in each trial.

Table 3 describes the design of the trials in my data. Each of the trials is randomized, double-blind, and parallel-armed,[12] which means that each patient is observed in only one treatment state. There is substantial variation in the probability of treatment across trials,[13] though much of the variation is generated by probability 0.5 and one trials. For simplicity, only the latter trials are used to generate the results presented below (and in table 1). (Employing all trials affirms the results from the limited sample.) The trials in my data also vary in their length, and a number measure outcomes on the same patient at multiple dates.

A limitation of my data is that each observation occurs at the arm-date level rather than the patient-date level. For example, I have data on the fraction of patients whose ulcer healed or the average (and standard deviation of) change in LDL among patients in an arm at a given date, but I do not have data on whether individual patients' ulcers healed or how much their LDL fell at that date. Because each trial may have more than one treatment arm, there are more than two times as many arms as trials reported in table 3. Because some arms are measured multiple times during a trial, there are more arm-date observations than there are arms per trial. Finally, because each arm has multiple patients, there are fewer arms than patients (and fewer arm-date observations than patient-date observations). Therefore, although the sample sizes for some of the regressions reported below may appear small, they represent a much larger number of patients. Indeed, table 1 reveals that the probability 0.5 and one trials alone include over 13,000 patients treated with H_2 blockers, over 3,000 treated with PPIs, and nearly 7,000 treated with statins.

Table 4 present summary statistics for the arm-date level data. The data are weighted such that each patient has the identical weight, regardless of how many times his outcome is measured. The results, which are limited to treatment arms of probability 0.5 and one trials, are broken down by class of treatment. A number of the statistics warrant

[12] There is one exception. One of the statin trials has a crossover design.

[13] Much of the variation arises because the trials often test several different dosages of a treatment against a control. Each dosage group has roughly an equal numbers of subjects. Therefore, the probability of receiving treatment is typically $n/(n-1)$, where n is the number of different dosages being examined. It might appear odd that trials can both be blinded and have a probability of treatment equal to one. This makes sense, however, given my method for calculating the probability of treatment. If there is a trial that compares two treatments from the same class of medication (H_2 blocker, PPI, or statin) or a trial that examines multiple dosages of the same medication, that trial is recorded as having a probability of treatment equal to one because there is no chance of getting a placebo, palliative, or older-class drug. The trial will also be blinded, however, because investigators may not want doctors or patients to know the exact treatment and dosage given.

TABLE 3
Summary Statistics Concerning Trial Design: All Trials (Frequency of Trials)

	Treatment		
	H_2 Blockers	PPIs	Statins
	By Type of Control		
Placebo or palliative	96	0	44
H_2 blockers	100	29	
PPIs	0	10	
Statins			14
Total	196	39	58
	By Type of Control: Only Probability 0.5 and One Trials		
Placebo	79	0	20
H_2 blockers	100	21	
PPIs	0	10	
Statins			14
Total	179	31	34
	By Probability of Treatment		
.25 (1/4)	3	0	0
.33 (1/3)	0	1	0
.40 (2/5)	1	0	0
.50 (1/2)	79	21	20
.60 (3/5)	0	0	1
.67 (2/3)	7	6	8
.75 (3/4)	6	1	6
.80 (4/5)	0	0	6
.83 (5/6)	0	0	1
.86 (6/7)	0	0	2
1.00 (1)	100	10	14
	By Length of Trial (Weeks)		
2	3	0	0
3	1	0	1
4	95	32	6
5	0	0	1
6	47	3	10
8	45	4	9
10	1	0	0
12	4	0	11
16	0	0	6
18	0	0	1
24	0	0	4
26	0	0	1
48	0	0	2
104	0	0	2
156	0	0	3
208	0	0	1
	By Number of Measurements per Patient		
1	92	1	33
2	93	30	8
3	11	7	6
4	0	1	9

TABLE 3
(*Continued*)

	Treatment		
	H_2 Blockers	PPIs	Statins
5	0	0	1
6	0	0	1

NOTE.—Included in the category of placebo-controlled H_2 blocker trials are trials in which the control is a palliative such as an antacid or bismuth subcitrate.

further explanation. The high probability of treatment in H_2 blocker and statin trials reflects the large number of trials in which the probability of treatment is one. Most ulcer trials, regardless of control therapy, permit patients to consume antacids, a palliative. (This fact further reduces the risk of self-selection.) Most cholesterol trials required patients to restrict their diet during the experiment. The number of patients who are evaluated—that is, whose outcomes are measured—in an arm is often lower than the number of patients enrolled in that arm. The reason is attrition. I handle this by calculating outcomes and estimating placebo effects under four alternative assumptions: that all patients who exit the trial heal, not all patients heal, all heal at the same rate as those in the control arm, and all heal at the same rate as those who do not exit from the treatment arm. Although I report only the results under the last assumption, my results are robust to the other alternatives. Although the same proportion of patients—roughly 80 percent—heal with H_2 blockers and PPIs, the rate of healing is much faster with PPIs, as indicated by the daily healing rate, which is 5 percent and 9 percent per day for H_2 blockers and PPIs, respectively. (The daily healing rate is calculated assuming a constant hazard rate into healing.) The average patient in the treatment arm of a statin trial experiences a 32 percent reduction in LDL levels. Ninety-two percent experience some adverse events or side effects and 26 percent experience the usual side effects associated with statins (i.e., nausea, diarrhea, constipation, or muscle aching).

The empirical model I estimate takes the form

$$y_{it} = f(\beta_0 + \beta_1 d_i + \boldsymbol{\beta}_2 \boldsymbol{A}_{it} + \beta_3 \log(t) + \boldsymbol{\beta}_4 \boldsymbol{P}_{it}) + \epsilon_{it}, \tag{3}$$

where y_{it} is the average outcome in treatment arm i at time t (weeks); d_i is the probability of treatment (either 0.5 or one); $\boldsymbol{A}_{it}$ are trial and arm characteristic variables, such as whether antacids are permitted and dosage; and $\boldsymbol{P}_{it}$ are patient characteristic data, such as baseline average LDL levels and average age. I account for the time-series aspects of the data rather crudely by including the log of weeks of treatment as a regressor and weighting observations so that each patient (not patient-date) has the identical weight. Ordinary least squares estimation was

TABLE 4

Summary Statistics for Treatment Arms: Probability 0.5 and One Trials Only

Variable	H_2 Blocker Trials			PPI Trials			Statin Trials		
	Arms	Mean	Standard Deviation	Arms	Mean	Standard Deviation	Arms	Mean	Standard Deviation
Date results published (year)	334	1987	5	76	1992	4	83	1997	3
Probability of treatment (0–1)	334	.90	.20	76	.67	.24	83	.89	.21
Antacid permitted? (1–5)	334	3.56	.69	76	2.67	1.30			
Diet restricted? (0/1)							80	.71	.46
Dosage (grams)	331	.47	.35	76	.03	.01	83	.04	.02
Frequency of dosing (times/day)	329	1.85	1.08	76	1.05	.21	83	1.08	.27
Patients enrolled (#)	334	179	142	76	114	42	79	265	182
Patients evaluated (#)	334	161	127	76	105	40	76	232	178
Patients healed (#)	334	130	111	76	83	34			
Patients healed (share)	334	.78	.16	76	.80	.17			
Constant daily healing rate (share)	334	.05	.02	76	.09	.03			
Baseline LDL (mg/dl)							83	195	25
Baseline HDL (mg/dl)							83	48	5
Baseline triglycerides (mg/dl)							82	160	38
Reduction in LDL (mg/dl)							83	63	22
Reduction in LDL (fraction)							83	.32	.08
Reduction in LDL (mg/dl·week)							83	5.02	4.06
Any adverse events (0/1)							81	.92	.28
Usual adverse events (0/1)							81	.26	.44
Average age (years)	296	46	4	66	46	4	66	56	3
Male (share)	302	.71	.07	70	.72	.07	71	.60	.18

Note.—Each observation is an arm. Each arm is weighted by the number of patients in the arm, regardless of the number of measurements on the patient. The antacid permitted variable is coded 1 if patients are prohibited from taking antacids, 2 if antacids were discouraged, 3 if antacids were permitted or no instruction was given, 4 if antacids were provided, and 5 if antacids were required. The probability of healing and daily healing rate for ulcer trials were calculated assuming patients who exited healed at the same rate as those who remained.

used, with heteroskedasticity permitted across arms. The results for three specifications of covariates for each of five outcomes are presented in table 5. One specification includes only a constant and the probability of treatment; the second adds trial and arm characteristics, as well as the log of the measurement date; and the final specification adds patient characteristics. The five outcomes are ulcer healing in H_2 blocker trials, ulcer healing in PPI trials, reduction of LDL in statin trials, and any and usual adverse events in statin trials.

The central finding is that a higher probability of treatment is associated with a significantly higher healing rate in H_2 blocker trials and with a significantly greater reduction in LDL levels and side effects in statin trials. The coefficient estimates suggest that going from a probability zero to a probability one trial, which generates the full placebo effect, increases the probability of healing an ulcer with H_2 blockers by 0.06–0.22. It increases the reduction in LDL levels following statin treatment by roughly 45–92 percent. It also causes an increase in the probability of any side effects and usual side effects from statin therapy by at least 0.6 and 0.5, respectively. All these results but two are significant at conventional levels; the effect for H_2 blockers and usual side effects in the third specification are each significant at the 10 percent level (one-sided test) or 15 percent level (two-sided test). The evidence on healing in PPI trials suggests that there are no placebo effects of the sort I hypothesize.

The finding with respect to side effects is particularly revealing. Patients in higher-probability statin trials should expect not merely a higher probability of a reduction in LDL but also a higher probability of side effects. The data suggest that not only do expectations about good outcomes affect patient response, but expectations about bad outcomes do as well. In other words, there are placebo effects for both positive outcomes (reduction in LDL levels) and negative outcomes (side effects).

The regression results I present are robust to specification. In addition to the variations I have already mentioned, the results persist when I employ a logit model for ulcer healing and statin side effects, interact all variables with the probability of treatment, or account for the time-series aspects of the data to allow unstructured correlation of errors within arms across dates. The results are also robust to inclusion of proxies for self-selection and attrition, such as the date of the trial or the fraction of subjects who exit. Estimates become more precise, and there emerges evidence for placebo effects even with PPIs when I analyze the ulcer data using survival models, whether I assume a constant or proportional hazard. Finally, the results are robust to alternative weighting schemes (none or weighting in proportion to patients × measure-

TABLE 5

Regression Results: Treatment Arms of Probability 0.5 and One Trials Only

							Statins								
	H_2 Blockers: Dependent Variable: Probability of Healing			PPIs: Dependent Variable: Probability of Healing			Dependent Variable: Log Lower LDL			Dependent Variable: Any Adverse Events			Dependent Variable: Usual Adverse Events		
Treatment Specification	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Constant	.586	.278	.657	.781	.638	1.587	3.646	2.990	−3.943	.344	.508	11.549	−.185	−.193	−15.82
	(.052)	(.054)	(.308)	(.050)	(.076)	(.448)	(.148)	(.434)	(6.889)	(.345)	(.263)	(6.202)	(.125)	(.433)	(4.08)
Probability of treatment (0–1)	.220	.071	.063	.032	−.050	−.103	.481	.822	.622	.646	.742	.853	.501	.613	.539
	(.056)	(.036)	(.041)	(.071)	(.078)	(.107)	(.213)	(.335)	(.356)	(.346)	(.261)	(.350)	(.203)	(.332)	(.367)
Dosage (g)		−.005	−.029		2.065	2.365		4.400	.4980		−1.509	.008		−6.084	−5.949
		(.030)	(.038)		(.755)	(.857)		(2.338)	(2.306)		(2.227)	(1.461)		(3.136)	(3.373)
Frequency of dosing (times/day)		.006	.015		−.190	−.190		.171	.188		.196	.115		−.145	.087
		(.010)	(.013)		(.027)	(.026)		(.161)	(.117)		(.124)	(.085)		(.152)	(.133)
Antacid permitted (1–5)		−.007	.001		−.016	−.027		NA	NA		NA	NA		NA	NA
		(.010)	(.011)		(.013)	(.020)									
Diet restricted (0/1)		NA	NA		NA	NA		−.278	−.175		−.295	−.291		.177	−.215
								(.117)	(.077)		(.140)	(.097)		(.135)	(.141)
Measurement date (log weeks)		.293	.299		.356	.361		.071	.110		−.071	.018		.057	.011
		(.017)	(.018)		(.031)	(.034)		(.059)	(.085)		(.054)	(.048)		(.078)	(.119)
Baseline LDL (mg/dl)									1.860			−.961			−.942
									(.465)			(.481)			(.866)
Baseline HDL (mg/dl)									−.006			−1.055			1.550
									(.402)			(.766)			(.733)
Baseline triglycerides (mg/dl)									.215			−.089			.350
									(.192)			(.115)			(.204)
Age (log years)			−.075			−.262			−.982			−.298			3.200
			(.079)			(.117)			(1.526)			(.918)			(.767)
Male (0/1)			−.173			.139			.208			−.944			.208
			(.113)			(.142)			(.517)			(.553)			(.717)
Arms	334	329	287	76	76	66	83	80	60	81	78	59	81	78	59

Note.—Each observation is an arm-date pair. Each observation is weighted by the number of patients in the arm divided by the number of dates on which the arm is measured. Standard errors are in parentheses below coefficient estimates. They are calculated assuming heteroskedasticity across treatment groups. See also the note for table 4.

ments) and modification of distributional assumptions with a gamma regression.

An important footnote to the findings reported in this paper is that regression analysis also reveals a positive correlation between the probability of treatment and outcomes *in the control group*, but only when the control therapy is active treatment, namely H_2 blockers in PPI trials. It does not find a positive correlation when the control is an inert pill.[14] This result suggests that placebo effects differ in the treatment and control arms. As a result, within-trial comparisons may not properly identify relative pharmacological effects. The outcome in the treatment arm is the pharmacological effect of treatment plus a fraction of the placebo effect of the treatment; the outcome in the control arm is the pharmacological effect of control plus a fraction of the placebo effect of the control. Because the placebo effect of the treatment and the placebo effect of the control do not cancel out, the within-trial comparison identifies the relative pharmacological effect of treatment plus a fraction of the difference in placebo effects across arms. Indeed, since the fraction of the placebo effect manifested in each arm equals the probability of treatment, the within-trial comparison will also vary with the probability of treatment.[15]

Table 6 breaks down estimates of placebo effects by individual drugs within each class of medication. To keep things simple, I employ unconditional mean outcomes in probability 0.5 and one trials to calculate pharmacological effects (difference between mean outcomes in the treatment and placebo-controlled arms of probability 0.5 trials) and placebo effects (twice the difference between mean outcomes in the treatment arms of probability 0.5 and one trials). The primary finding is that placebo effects vary across drugs within a medication class. Among H_2 blockers, for example, they range between 31 percent and over 200 percent the size of the pharmacological effect of these drugs, differences that are statistically significant. This is important because skeptics might argue that one can simply ignore placebo effects when deciding whether to approve or prescribe a drug on the assumption that placebo effects are identical across medications for a given ailment. This assumption is rejected by the data. Indeed, if one were to rank drugs on the basis of total effects—pharmacological plus placebo effects—rather than on solely pharmacological effects, columns 5 and 6 of table 6 demonstrate

[14] The results are presented in Malani (2004). These regressions include trials with probabilities other than 0.5 and one to generate variation in the probability of treatment because probability one trials have no control arm. The correlation between outcomes and probability of treatment in arms in which H_2 blockers are the control is similar in magnitude to the correlation in arms in which H_2 blockers act as the treatment.

[15] Another implication may be that a prerequisite for placebo effects is that a drug has positive pharmacological effects. This would rule out the use of pure sugar pills for therapy.

TABLE 6
Comparing Pharmacological and Placebo Effects

Therapy	Pharmacological Effect (1)	Placebo Effect (2)	Total Effect (3)	Placebo Effect/ Pharmacological Effect (4)	Rank: Without Placebo Effects (5)	Rank: With Placebo Effects (6)
	Ulcer (Outcome: Probability of Healing)					
Tagamet (cimetidine)	.38	.14	.52	.38	2	3
Axid (nizatidine)	.23	.50	.73	2.13	3	1
Gastrosed (pirenzepine)	.20	.28	.48	1.42	4	4
Zantac (ranitidine)	.43	.13	.56	.31	1	2
Prilosec (omeprazole)	.48	.03	.51	.06	2	2
Protonix (pantoprazole)*	.52	−.0001	.52	−.0002	1	1
	High Cholesterol (Outcome: Reduction in LDL [mg/dl])					
Lipitor (atorvastatin)	64.50	19.25	83.75	.30	1	1
Lescol (fluvastatin)	32.18	22.37	54.56	.70	5	4
Mevacor (lovastatin)	61.27	−5.49	55.78	−.09	2	3
Pravachol (pravastatin)	49.27	−1.59	47.68	−.03	4	5
Zocor (simvastatin)	59.10	19.96	79.05	.34	3	2

Note.—Pharmacological effects are relative to placebo control. Placebo effects are two times the difference in outcome between probability 0.5 and one arms. All effects are calculated from unconditional means in arm-level data weighted such that each patient has the identical weight regardless of the number of measurements on that patient. Placebo effects pertain to drugs for which there are not both probability 0.5 and one trials.

* All effects are statistically significant, except for the placebo effect of pantoprazole.

that the ordering of drugs would change. For example, traditionally Zantac was thought to be the most effective H_2 blocker, but my analysis suggests that Axid is in fact the better H_2 blocker. Although Lipitor remains the most effective statin, the second-best statin is not Mevacor, but rather Zocor.

A second finding is that expectation effects may be positive for some drugs and negative for other drugs within a class. Although the top two statins by market size, Lipitor and Zocor, generate positive placebo effects roughly 30 percent the size of pharmacological effects of these drugs, other statins, Pravachol and Mevacor, generate negative expectation effects between 3 percent and 9 percent the size of pharmacological effects.[16] At first blush the result appears similar to the side effects result in that expectations about receiving a drug correlate with adverse outcomes. The two results, however, are distinct. With side effects, patients expected more and got more. Because the outcome was undesirable, this relationship is called the "nocebo" or negative placebo effect in the medical literature. In contrast, with Pravachol and Mevacor, patients expected a greater reduction in LDL in probability one trials relative to probability 0.5 trials, but they got a lower reduction in LDL. There is no term for this phenomenon in the medical literature, but it is related to the idea of moral hazard in the economics literature or risk compensation in the psychology literature.[17]

IV. Conclusion

This paper highlights a neglected feature of clinical trials—that they manipulate patient expectations—and suggests that this feature can be used to identify the effects of expectations on outcomes, that is, placebo effects. The primary result is that there are statistically and medically significant placebo effects from ulcer and cholesterol drugs. Importantly, these are drugs that have serious pharmacological effects[18] and had or have significant markets. Indeed, Lipitor and Zocor were the number 1 and 2 best-selling pharmaceuticals in 2003, with combined sales of over $16 billion (IMS Health 2004). Nexium and Prevacid—though now used mainly for heartburn—were the number 6 and 7 best-selling pharmaceuticals, with combined sales of over $7 billion (IMS Health 2004).

[16] A caveat to this conclusion is that if one suspects self-selection, it could be responsible for this finding.

[17] In this regard, expectation effects might be related to the trade-off that Peltzman (1975) suggested might cause mandatory vehicle safety regulation to raise vehicle crash rates or that Goldman, Lakdawalla, and Sood (2004) suggest cause innovations in HIV treatment to raise sexual risk-taking behavior.

[18] Many believe that placebo effects are limited to placebo or complementary and alternative therapies.

There are several limitations of the analysis in this paper. For example, I employ arm-level data from parallel-armed trials. To generate results that would inspire greater confidence, one would like patient-level data (which are not subject to aggregation bias) and data from crossover trials (which should generate more precise estimates).[19] Moreover, my test examines the effects of change in beliefs about the probability of treatment rather than about the value of treatment on outcomes. The two effects are equal only when outcomes are a function of the expected value of treatment. The gap is analogous to that between expected income and expected utility of income.

There are a number of questions that naturally follow from my findings. First, to what extent are placebo effects a behavioral phenomenon? The negative placebo effects from consumption of Pravachol and Mevacor suggest moral hazard, which is a behavioral phenomenon. But not enough is known about the physiological effects of expectation to rule out that modality even in these cases. Moreover, it is odd that behavioral effects would vary across drugs within the same class of medication. Nonetheless, it could be that all statins have counterproductive behavioral effects but that some also experience physiological placebo effects that more than offset those effects.

Second, are placebo effects keyed to patient beliefs about pharmacological effects or to patient beliefs about pharmacological and placebo effects, and how are these beliefs generated? The answer to the latter question, in particular, is central to determining whether beliefs can be employed to reduce health care costs and whether placebo effects are a serious challenge to the independence axiom.

References

Amanzio, Martina, and Fabrizio Benedetti. 1999. "Neuropharmacological Dissection of Placebo Analgesia: Expectation-Activated Opioid Systems versus Conditioning-Activated Specific Subsystems." *J. Neuroscience* 19 (January 1): 484–94.

Benedetti, Fabrizio, Claudia Arduino, and Martina Amanzio. 1999. "Somatotopic Activation of Opioid Systems by Target-Directed Expectations of Analgesia." *J. Neuroscience* 19 (May 1): 3639–48.

Goldman, Dana, Darius Lakdawalla, and Neeraj Sood. 2004. "HIV Breakthroughs and Risky Sexual Behavior." Working Paper no. 10516 (May), NBER, Cambridge, MA.

Heckman, James J. 1992. "Randomization and Social Policy Evaluation." In *Evaluating Welfare and Training Programs*, edited by Charles F. Manski and Irwin Garfinkel. Cambridge, MA: Harvard Univ. Press.

[19] Such a trial is now under way at the University of Virginia. The trial exposes patients to different probabilities of treatment and then randomizes patients to treatment according to those probabilities. The treatment is caffeine, and the outcome is blood pressure and heart rate. See Malani and Houser (2006).

Hróbjartsson, Asbjørn, and Peter C. Gøtzsche. 2001. "Is the Placebo Powerless? An Analysis of Clinical Trials Comparing Placebo with No Treatment." *New England J. Medicine* 344 (May 24): 1594–1602.

———. 2004. "Is the Placebo Powerless? Update of a Review with 52 New Randomized Trials Comparing Placebo with No Treatment." *J. Internal Medicine* 256 (2): 91–100.

IMS Health. 2004. *IMS World Review*. London: IMS Health.

Kaptchuk, Ted J. 1998. "Powerful Placebo: The Dark Side of the Randomised Controlled Trial." *Lancet* 351 (June 6): 1722–25.

Kirsch, Irving, Thomas J. Moore, Alan Scoboria, and Sarah S. Nicholls. 2002. "The Emperor's New Drugs: An Analysis of Antidepressant Medication Data Submitted to the U.S. Food and Drug Administration." *Prevention & Treatment* 5 (0023a). http://www.journals.apa.org/prevention/volume5/pre0050023a.html.

Kirsch, Irving, and G. Sapirstein. 1998. "Listening to Prozac but Hearing Placebo: A Meta-analysis of Antidepressant Medication." *Prevention & Treatment* 1 (0002a). http://www.journals.apa.org/prevention/volume1/pre0010002a.html.

Kirsch, Irving, and Lynne J. Weixel. 1988. "Double-Blind versus Deceptive Administration of a Placebo." *Behavioral Neuroscience* 102 (April): 319–23.

Malani, Anup. 2003. "Placebo Effects, Self-Selection and the External Validity of Clinical Trials." PhD diss., Dept. Econ., Univ. Chicago.

———. 2004. "Testing for Placebo Effects Using Data from Blinded, Clinical Trials." Working paper, Univ. Virginia.

———. 2005. "Patient Enrollment in Medical Trials: Selection Bias in a Randomized Experiment." Working paper, Univ. Chicago.

Malani, Anup, and Daniel Houser. 2006. "Expectations Mediate Objective Physiological Placebo Effects." Working paper, Univ. Chicago.

Marlatt, G. A., and D. J. Rohsenow. 1980. "Cognitive Processes in Alcohol Use: Expectancy and the Balanced Placebo Design." In *Advances in Substance Abuse: Behavioral and Biological Research*, vol. 1, edited by Nancy K. Mello. Greenwich, CT: JAI.

Miller, Franklin G., Ezekiel J. Emanuel, Donald L. Rosenstein, and Stephen E. Straus. 2004. "Ethical Issues Concerning Research in Complementary and Alternative Medicine." *J. American Medical Assoc.* 291 (February 4): 599–604.

Peltzman, Sam. 1975. "The Effects of Automobile Safety Regulation." *J.P.E.* 83 (August): 677–725.

Penick, S. B., and S. Fisher. 1965. "Drug-Set Interaction: Psychological and Physiological Effects of Epinephrine under Differential Expectations." *Psychosomatic Medicine* 27 (2): 177–82.

Penick, S. B., and L. E. Hinkle. 1964. "The Effect of Expectation on Response to Phenmetrazine." *Psychosomatic Medicine* 26 (4): 369–74.

Pollo, Antonella, et al. 2001. "Response Expectancies in Placebo Analgesia and Their Clinical Relevance." *Pain* 93 (July): 77–84.

Shapiro, Arthur K., and Elaine Shapiro. 1997. *The Powerful Placebo: From Ancient Priest to Modern Physician*. Baltimore: Johns Hopkins Univ. Press.

Skovlund, Eva. 1991. "Should We Tell Trial Patients That They Might Receive Placebo?" *Lancet* 33 (April 27): 1041.

Talbot, Margaret. 2000. "The Placebo Prescription." *New York Times Mag.* (January 9).

Vickers, Andrew J., and Anton J. M. de Craen. 2000. "Why Use Placebos in

Clinical Trials? A Narrative Review of the Methodological Literature." *J. Clinical Epidemiology* 53 (February): 157–61.

Wager, Tor D., et al. 2004. "Placebo-Induced Changes in fMRI in the Anticipation and Experience of Pain." *Science* 303 (February 20): 1162–67.

Zubieta, Jon-Kar, et al. 2005. "Placebo Effects Mediated by Endogenous Opioid Activity on μ-Opioid Receptors." *J. Neuroscience* 25 (August 24): 7754–62.