Journal of Econometrics 144 (2008) 341–351

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Patient enrollment in medical trials: Selection bias in a randomized experiment

Anup Malani [*]

University of Chicago, 1111 E. 60th Street, Chicago, IL 60637, United States

ARTICLE INFO

Article history:
Available online 8 May 2008

JEL classification:
I10
C90
C31

Keywords:
Selection
Randomization bias
Experiment
Treatment effect

ABSTRACT

Self-selection can bias estimates of treatment effects from randomized experiments if one is interested in extrapolating results to the general population. This paper notes that there is an isomorphism between the Roy model for the sorting of workers into sectors and the decision of subjects to participate in randomized experiments. The main implication is that, as the probability of receiving active treatment rises, patients who are less optimistic about new treatment will begin to enroll and estimates of treatment effects will fall. This, in turn, implies that selection bias is positive. These findings are confirmed with data from trials of ulcer medications.

Conventional wisdom holds that patient self-selection does not pose a problem for randomized medical trials. The reason is that random assignment eliminates unobservable differences between treatment groups. This is sufficient protection against bias if the researcher only cares about the effect of treatment on the treated — i.e. the effect of treatment on patients enrolled in the trial. Most trials, however, are concerned with predicting the effect of treatment on a patient population beyond the trial. In these cases, randomization does not prevent selection bias[1] which arises from systematic differences between the trial population and the general patient population.[2]

This paper proposes a model of patient self-selection into clinical trials in order to gain insight into the effect of this sorting on standard estimates of treatment effects. The model is based upon the Roy model of sorting of workers across sectors, — with patients playing the role of workers and treatment via a trial and treatment outside the trial as the two sectors. My analysis makes two critical assumptions. First, there is a new treatment that is being tested (against a placebo control) in a medical trial. The new treatment is only available through the trial. There may be a conventional treatment, however, that is available outside the trial. Second, the population of interest, i.e. the population to which the researcher would like to generalize her results, is those members of the general population who would take the new treatment if it were approved for sale. My main findings are as follows. First, in the absence of a conventional alternative therapy that is available outside the trial, clinical trials do not suffer from self-selection. The reason is that all members of the population of interest would enroll in the trial.

Second, even if there is a conventional therapy, the only patients who engage in self-selection are those who believe the new treatment superior to conventional treatment and conventional treatment superior to no treatment. For these individuals, my primary finding is that, as the probability of randomization into active treatment (as opposed to placebo control) rises, the average effect of the new treatment on the trial population falls. The reason is that the higher the probability of treatment, the less optimistic a patient has to be about the new treatment (relative to conventional treatment) in order for her to prefer enrolling in a trial – a lottery between the new treatment and no treatment – over the non-trial alternative – certain consumption of conventional treatment. I verify this prediction with a data set of outcomes from over 120 trials of anti-ulcer medications with different probabilities of treatment.

An important implication of this finding is that self-selection causes the standard estimate of treatment effects – the difference between average outcomes in the new-treatment and placebo-control groups – to have a positive bias. The reason is that medical trials always have a probability of treatment less than one, whereas outside the trial context a patient who takes the new treatment would do so with certainty, i.e., with probability equal one. If outcomes fall with higher probability of treatment, then outcomes in the real world must be smaller than outcomes in a trial. In my

[*] Tel.: +1 773 702 9602; fax: +1 773 702 0730.
E-mail address: amalani@uchicago.edu.

[1] This bias goes by the name "randomization bias" in the literature on social experiments (Heckman, 1992; Heckman and Smith, 1995).

[2] Another way to put this is that in randomized experiments, self-selection interferes with external validity but not internal validity. In non-randomized experiments or non-experiments, self-selection interferes with both external and internal validity.

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.04.002

sample of ulcer trials, I estimate that selection bias accounts for roughly half of treatment effects for two medications (H_2-blockers and proton-pump inhibitors) and can reverse the sign of treatment effects for a third medication (prostaglandins).

My third finding is that an alternative experimental design which permits patients to choose among treatment lotteries can eliminate bias from self-selection. In this design patients are offered a choice between (a) a lottery over new treatment and placebo or (b) a lottery over new treatment and conventional control. All patients in the population of interest, i.e. those who prefer new treatment, would enroll. Those that prefer conventional treatment to no treatment would simply enroll in the second lottery. The cost of this design is that one cannot identify the effect of new treatment relative to no treatment on these patients, but that information is not very valuable because these patients would take conventional therapy rather than no treatment if the new treatment were not available.

Although there is an extensive economics literature on self-selection across treatment and control groups in non-randomized studies, there is less discussion of self-selection across study participation and non-participation in the context of randomized experiments. That which exists notes that self-selection into a study is not a problem so long as one is interested only in the effect of treatment on the treated. If one is interested in treatment effects among a broader population, however, the participation decision may introduce what is called randomization bias (Heckman, 1992; Heckman and Smith, 1995; Heckman, 1996). The limited literature on randomization bias does not model the behavior that generates this bias, as this paper does, but merely attempts to demonstrate empirically that such bias exists (Heckman, 1992; Kamoinka and Lacroix, 2002).[3] The literature on self-selection in non-randomized studies does not exhaust the analytic insights possible with randomized experiments because the lottery aspect of the latter permits greater theoretical structure to be placed on models of self-selection behavior. This reduces the work that must be done by data to identify the effects of selection.

In the "statistics for medicine" literature there is little discussion of patient self-selection into medical trials. Rather, the focus is on the selection of patients *by doctors* into trials, either via recruitment (see, e.g., Senore et al. (1999)) or exclusion criteria (see, e.g., Ellenberg (1994) and Robinson et al. (1996)). Physician-selection might alter the impact patient self-selection has on estimates of treatment effects to the extent that it causes a deviation from random sampling from the population of interest. I leave that interesting question for future research. The few papers there are on the participation decision by patients merely attempt to survey the reasons why patients choose not to participate in a medical trial (see, e.g., Verheggen et al. (1998) and Britton et al. (1999)) or to document differences between the characteristics of patients who choose to enroll and those who do not (see, e.g., Olschewski et al. (1992) and Britton et al. (1999)), without drawing conclusion about the impact of the differences on estimates of treatment effects.

The paper may be outlined as follows. Section 1 sets forth my model of self-selection into clinical trials. Section 2 test the predictions of my model against a data set of outcomes from ulcer trials. Section 3 proposes and evaluates an alternative trial design to estimate treatment effects without bias due to self-selection.

1. Model of patient self-selection

1.1. Setup

Let us begin with some building blocks for the model. Assume an ill patient faces two possible future health states: continued illness $\underline{y}$ or recovery $\bar{y}$.[4] Utility in these states is $\underline{U} = u(\underline{y})$ and $\bar{U} = u(\bar{y})$, respectively. I will discuss three treatments for the patient's ailment, indexed by k: no treatment ($k = 0$), a new treatment ($k = 1$), and conventional treatment ($k = 2$). Let y_k be a random variable that gives the patient's health outcome following treatment k and define p_k as the probability of recovery with treatment k: $p_k = \Pr\{y_k = \bar{y}\}$. Let π_k be a patient's belief about p_k. In general, π_k would be a distribution function, but it will make no difference to the analysis if it is a point estimate and so I shall assume that. In order for self-selection to affect estimates of treatment effects, it must be that there is heterogeneity of treatment effects and of beliefs about treatment effects in the population. Let $g(\mathbf{p}, \boldsymbol{\pi})$ be the joint probability distribution function for chances of recovery $\mathbf{p} = (p_0, p_1, p_2)$ and for beliefs $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2)$ about these chances.

The assumption that patients have defined beliefs about treatment effects requires some clarification and perhaps justification, at least with regard to the new treatment. I do not assume that patients know their response to new treatment, only that they have beliefs about the value of that treatment. These beliefs could be based on experience with related compounds, experience that the patients have previously responded well or poorly to treatments generally, news reports about the new treatment, advice from personal doctors, or information about the average value of all new treatments. The beliefs could be wrong – though I will rule out a particular type of wrong in Section 1.3. While it would be useful to explore the source of beliefs about new treatment, it should not be controversial to assume that beliefs exist.

In order to formally model a medical trial, one needs to specify its design. My focus will be on blinded, randomized, placebo-controlled trials – the gold-standard according to the U.S. Food and Drug Administration. From the patient's perspective, the trial is a lottery over assignment to new treatment or the placebo. Let d be the probability of assignment to the new treatment arm. This probability is revealed to patients during the informed consent process.[5] The purpose of blinding is to prevent patients in the control arm from leaving or simultaneously seeking treatment outside the trial. In my analysis, blinding implies that researchers control the lottery payoffs that patients are offered. Without it, patients would, in effect, be able to pick the control arm payoff.

I assume that patients decide to enroll in a clinical trial if and only if their expected utility from the trial is greater than their expected utility from treatment outside the trial. To ensure that this choice has bite, I assume that the new treatment is only available through the trial. Otherwise no patients would enroll in trials without monetary compensation, which is generally limited to out-of-pocket expenses.[6] I also assume that it is the utility

[3] One ostensible exception is Philipson (2000), which examines the effect of changes in the proportion of a population that is treated on the treatment effect in that population. The distinction between that work and the present paper is twofold. First, in Philipson's paper, differences in the treatment probability change the treatment effect in the population (because of externalities of treatment on those not treated). In this paper, however, differences in the treatment probability change the *sampling of* treatment effects, not the treatment effects themselves. Second, Philipson is concerned with bias in estimates of the effect of treatment on the treated. I am concerned with the bias in estimates for a broader population.

[4] Appendix A.2 extends the results of the paper to the case of continuous outcomes.

[5] Indeed, this is required by law in the U.S. See, 21 C.F.R. Section 50.25, 45 C.F.R. Section 46.116. See also National Cancer Institute's Simplification of Informed Consent Documents (NCI2006-ICTemplate, 2006, "Research Regimens") and its Informed Consent Template for Cancer Treatment Trials (NCI2006-ICRecommendations, 2006, "During the Study..."), and Section C.1 of the Belmont Report (for the Protection of Human Subjects of Biomedical and Research, April 18, 1979). For analogous international requirements, see the Declaration of Helsinki, Section I.9 (World Medical Association, 2000).

[6] Research institutions typically limit, by policy, compensation to levels not regarded as having an undue influence on the decision of patients to participate in research. See, e.g., University of North Carolina, Office of Clinical Trials (2004), Pfizer (2006). Grady et al. (2005) find that use of patient compensation is widespread but the amounts are modest. For more on the medical debate over compensation, see Dunn and Gordon (2005) and Resier (2005).

of health and not the disutility of foregone consumption due to the cost of treatment that drives enrollment in clinical trials. This assumption may be justified by the aforementioned limit on monetary compensation and the fact that patients typically have health insurance that covers the marginal cost of care.

In order to fully characterize the patient's choice problem, note initially that expected utility from consumption of treatment k is

$$EU_k = \pi_k \bar{U} + (1 - \pi_k) \underline{U}.$$

Expected utility outside the trial depends on whether the patient would choose no treatment or conventional treatment. Assuming that the patient chooses among these treatments on the same basis as she chooses whether to enroll in a trial, expected utility outside a trial is $\max\{EU_0, EU_2\}$.

Whereas outside the trial the patient has her choice of treatments, inside the trial she experiences a lottery over treatments. Because treatments themselves are lotteries over health states, the trial is actually a compound lottery over health states. Expected utility from the trial is

$$EU_T = \pi_T \bar{U} + (1 - \pi_T) \underline{U}$$

where $\pi_T = d\pi_1 + (1 - d)\pi_0$.[7]

The patient's choice between treatment within and without the trial will solve $\max\{EU_T, \max\{EU_0, EU_2\}\}$. This problem can be rewritten in an intuitive manner with a simple transformation of variables. Let $\tilde{p}_k = p_k - p_0$ be the health benefit of treatment k relative to no treatment, and $\tilde{\pi}_k = \pi_k - \pi_0$ be the patient's belief about this treatment k effect. Now the patient's problem becomes

$$\max\{d\tilde{\pi}_1, \max\{0, \tilde{\pi}_2\}\}. \tag{1}$$

In other words, the patient's choice is between (a) a chance d at a treatment effect of $\tilde{\pi}_1$ with the new treatment, and (b) a certain treatment effect of 0 with no treatment or of $\tilde{\pi}_2$ with conventional treatment, whichever is greater.

An immediate implication of (1) is that risk aversion does not affect patient selection into medical trials. The reason is that outcomes are assumed to be binary. Therefore, if two lotteries offer the same expected outcome, they must also offer the same variation in outcomes. If health outcomes were continuous and one treatment offered a tighter distribution over outcomes than another, however, risk aversion might affect a patient's decision to enroll in a trial. That said, risk aversion would not, on the margin, always discourage enrollment in a trial. If the new treatment had lower outcome variance than conventional treatment or no treatment, a trial would attract the risk averse patient, ceteris paribus.[8]

1.2. Baseline

In order to determine whether and how self-selection introduces bias into estimates of treatment effects, one must specify the population of patients in which the researcher is interested. If the population is simply those that enroll in the trial, i.e., one is interested in the effect of "treatment on the treated," then it is well established that randomization eliminates self-selection bias. In my analysis, however, the population of interest is those patients who would take the new treatment if it were available outside the medical trial, i.e. patients who believe treatment effects of new treatment are superior to either no treatment or conventional treatment[9]:

$$\tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}. \tag{2}$$

The rationale is that it is these patients who are most likely to consume the new treatment if approved following the medical trial.[10] Note that the trial population – those patients for whom $d\tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}$ – is a subset of my population of interest. Everyone who enrolls in a medical trial meets the conditions for membership in the population of interest because $\tilde{\pi}_2/d > \tilde{\pi}_2$.

Some researchers may be interested in treatment effects among the whole population of patients. To these researchers, the ideal trial would enroll randomly selected patients from the population. In order to analyze bias from self-selection in this case, I recommend separating the population of interest into two subgroups: those who prefer new treatment (my population of interest) and those who do not. Total bias is the population-weighted average of bias in each of these groups. I will address selection bias in the group that does not prefer new treatment at the tail end of the next subsection.

1.3. Implications

Selection bias is defined to be the difference between treatment effects in the trial population and in the population of interest:

$$\begin{aligned} &E_g\left(y_1 - y_0 | d\tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}\right) - E_g\left(y_1 - y_0 | \tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}\right) \\ &\quad = \left[E_g\left(\tilde{p}_1 | d\tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}\right) - E_g\left(\tilde{p}_1 | \tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}\right)\right] \Delta y \end{aligned}$$

where $\Delta y = \bar{y} - \underline{y}$. Bias arises from mismatches between the trial population and the population of interest (both of which are defined by patients' beliefs) that affect the sampling of treatment effects in a medical trial.

This description immediately highlights two cases where there is no bias from self-selection. The first is where there is no correlation between beliefs about treatment effects and actual treatment effects. Although there is self-selection based on beliefs, sampling of treatment effects remains random. The second case is where there is no conventional control. The condition for membership in the trial population and for membership in the population of interest converge to $\tilde{\pi}_1 > 0$. All patients who prefer the new treatment will enroll in a medical trial regardless of the probability of receiving it because the worst case outcome (placebo control) is as good as the only therapy available outside the trial (no treatment).[11]

In reality, these conditions are uncommon, which is why self-selection warrants medical researchers' attention. In order to get theoretical traction on the bias from self-selection, I make two substantive assumptions. First, there is a positive correlation between a patient's beliefs about treatment effects and her actual treatment effects. More formally, (A1) $\tilde{p}_{1i} = f(\tilde{\pi}_{1i}) + \tilde{\varepsilon}_{1i}$, where i indexes individual patients, $\tilde{\varepsilon}_{1i}$ can be thought of as prediction error by patients, $f' > 0$, and $\tilde{\varepsilon}_{1i}$ is mean-independent of $\tilde{\pi}_{1i}$. The assumption of positive correlation is reasonable. It is consistent

[7] This formulation is similar to that in Philipson and DeSimone (1997).

[8] Harrison et al. (2005) fail to consider this possibility, and not surprisingly find that risk aversion does not lead to significant randomization bias in a Danish field experiment.

[9] One objection to this definition is that the doctor and not the patient chooses the latter's treatment. But the same could be said of the study participation decision – the doctor determines whether the patient will enroll in a medical trial. In that case, one should simply substitute the treating physician's beliefs for the patient's belief in the analysis.

[10] I address the problem of learning in the conclusion.

[11] A related case immune to randomization bias is where there is a conventional treatment, it is the control treatment in the medical trial, and all patients prefer conventional treatment to no treatment. In this case, the worst outcome from the trial – conventional control – is at least as good as the perceived best treatment available outside the trial.

with patients under- or over-estimating the effects of treatment. The opposite case, a negative correlation between beliefs and outcomes, is implausible, implying that patients who respond well to treatment systematically believe treatment is worse than patients who respond poorly to treatment.[12]

Following Heckman and Honore (1990), let (μ_1, μ_2) and Σ be the mean of $(\ln\tilde{\pi}_{1i}, \ln\tilde{\pi}_{2i})$ for all $\tilde{\pi}_{ki} > 0$, $k = 1, 2$.[13] Define $U_{ki} = \ln(\tilde{\pi}_{ki}) - \mu_k$, for $k = 1, 2$, $D_i = U_{1i} - U_{2i}$, $\sigma = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$, $a_1 = (\sigma_{11} - \sigma_{12})/\sigma$, $a_2 = a_1 - 1$, and $V_i = a_1 U_{2i} - a_2 U_{1i}$. By construction $U_{ki} = a_k D_i + V_i$, where D_i and V_i are uncorrelated. My second assumption is that (A2) V_i is mean-independent of D_i. This assumption is satisfied if $(\ln\tilde{\pi}_{1i}, \ln\tilde{\pi}_{2i})$ are uncorrelated or if $(\ln\tilde{\pi}_{1i}, \ln\tilde{\pi}_{2i})$ is distributed bivariate normal. It is also satisfied for certain classes of Poisson, binomial and gamma distributions (Lancaster, 1959; Hunter, 1972). More general conditions can be found in Lancaster (1959) and Leipnik (1961).

These assumptions yield the central insight of this paper: so long as the correlation between the treatment effects of the new therapy and of the conventional therapy are not too great, an increase in the probability of assignment to new treatment will lower estimates of the treatment effect of the new therapy from the medical trial. More formally,

Proposition 1. *Under assumptions (A1)–(A2), if* $\operatorname{corr}(\ln\tilde{\pi}_{1i}, \ln\tilde{\pi}_{2i}) \le \tilde{\sigma}_1/\tilde{\sigma}_2$, *then*

$$\frac{\partial E_g\left(\tilde{p}_1 | d\tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}\right)}{\partial d} \le 0.$$

The proof can be found in Appendix A.1. The logic, however, is that the trial only attracts patients who believe that the benefit of the new treatment is so great than, even with a probability $1 - d$ of obtaining no treatment at all, the trial offers a better expected outcome than conventional treatment. As the probability of assignment to new treatment increases, the perceived benefit of the trial rises. Some patients who were not optimistic enough about the new treatment to have previously enrolled, will now do so. This will lower the average beliefs about the new treatment's effect among enrollees. Since expected treatment effects are monotonically increasing in beliefs, average treatment effects rise with average beliefs.[14]

Critical to the proposition is the condition that the correlation between beliefs about the efficacy of the new and conventional treatments is not too high. If that is violated, then a higher probability of obtaining the new treatment may attract not just patients with more pessimistic beliefs about the new treatment, but also those with more optimistic beliefs about the new treatment who nevertheless have not enrolled because they also have more optimistic beliefs about the conventional treatment. (Recall that patients enroll in the trial not merely because $\tilde{\pi}_{1i}$ is high, but because it is greater than $\tilde{\pi}_{2i}/d$.) Fortunately, the condition on the correlation of beliefs about new and conventional treatment effects is rather lax. It is satisfied whenever the variance of beliefs about the new treatment is greater than the variance of beliefs about the conventional treatment, i.e., when $\tilde{\sigma}_1^2 \ge \tilde{\sigma}_2^2$, which is likely simply because the new treatment is new.

The economics behind Proposition 1 are identical to the Roy model (Roy, 1951; Heckman and Honore, 1990) of sorting of workers across sectors. In that model, workers have sector-specific skills with associated skill prices. A worker chooses the sector that offers her the highest income, which is the product of skill and skill price. The effect of a change in one skill price on the distribution of workers across sectors, and thus the average skill level in each sector, depends on distribution of skills among workers and the correlation of skills in each sector. This maps neatly onto the problem of self-selection into clinical trials. Here there are two sectors: the trial and no trial. Skills are treatment effects. And the skill price in the trial is the probability of receiving the new treatment while the skill price outside the trial is one.

The practical import of Proposition 1 is that self-selection introduces positive bias into standard estimates of treatment effects. To see this, note that taking new treatment outside a trial is equivalent to a trial where the probability of receiving new treatment is one, and that the entire population of interest would enroll in such a trial. Because treatment effects fall as the probability of treatment rises, it must be that consumption of the new treatment outside the trial (again, equivalent to a trial where the probability of treatment is one) must produce smaller treatment effects than new treatment in a trial (which generally has a probability of treatment below one). Indeed, a researcher could estimate the magnitude of bias if she had a number of trials with different probabilities of treatment. She would simply have to estimate the marginal effect of increases in the probability of treatment on estimates of treatment effects from individual trials and predict the treatment effect in a trial where the probability was one.

There are two circumstances in which selection bias may not be positive. First, if the medical trial offers significant monetary incentives for participation, then patients who prefer no treatment or conventional treatment may enroll.[15] In the former case, the bias is still positive because patients who prefer no treatment will have lower treatment effects than those who prefer new treatment, given the presumed positive correlation of beliefs and treatment effects. If patients who prefer conventional treatment take the bait, one cannot be sure treatment effects would be lower because there may be a number of patients who believe that both new treatment and conventional treatment are really good, but that conventional treatment is just slightly better. These patients may

[12] Under certain conditions, assumption (A1) admits a rational expectations formulation of beliefs about treatment effects, i.e., $f(\tilde{\pi}_{ki}) = \tilde{\pi}_{ki}$ and $E(\tilde{\varepsilon}_{ki}) = 0$. The conditions are dictated by the assumption that $\tilde{\varepsilon}_{ki}$ is mean-independent of $\tilde{\pi}_{ki}$. An example of a sufficient condition for a rational expectations formulation is that $\tilde{p}_{ki}$ is bounded in $[a, b] \subset [-1, 1]$ and $\max\{\tilde{\varepsilon}_{ki}\} = \min\{1 - |a|, 1 - |b|\}$.

[13] To avoid the problem that one cannot take logs of non-positive number, add a constant to outcomes in the binary case such that they lie in, e.g., [1, 3] rather than [−1, 1]. The approach in the text is acceptable since only subjects for whom $\tilde{\pi}_{1i} > \tilde{\pi}_{2i} > 0$ will change their enrollment decisions in response to changes in the probability of treatment d.

[14] A medical researcher may also be interested in the effect of self-selection on the variance of outcomes because the variance affects sample size calculations and the statistical significance of treatment effects. In general, it is not possible to sign the effect of patient self-selection on the variance of outcomes. The reason is that there are no simple restrictions on f that translate the sign of a change in variance of $\tilde{\pi}_{1i}$ into the sign of change in variance of $\tilde{p}_{1i}$, which equals $f(\tilde{\pi}_{1i}) + \tilde{\varepsilon}_{1i}$. This is not a problem in the special case where f is linear, which includes the case where patients have rational expectations. The reason is that when f is linear, an increase (decrease) in the variance of $\tilde{\pi}_{1i}$ implies an increase (decrease) in the variance of $\tilde{p}_{1i}$. (Recall that by assumption $f' > 0$.) In the linear case, the sign of the effect of selection on variance will depend on whether the distribution of log beliefs is log-concave or log-convex. If the distribution of log beliefs is log-concave (log-convex), the variance of treatment effects in the sample will be greater (smaller) than the variance in the population of interest. (The proof would be similar to that of Proposition 1 but would employ Propositions 1 and 2 in Heckman and Honore (1990) rather than Theorems 1′ and 2′ in Hadar and Russell (1969). Note that this result will not depend on the correlation between $\ln\tilde{\pi}_{1i}$ and $\ln\tilde{\pi}_{2i}$ because the variance will be a function of $a_1^2 > 0$.) Verifying whether beliefs are distributed log-concave or log-convex is difficult because beliefs are unobservable. However, if $\tilde{\varepsilon}_{1i}$ is distributed such that $\tilde{p}_{1i}$ falls in the same class of distributions as $\tilde{\pi}_{1i}$, then the investigator can verify the distribution of $\tilde{\pi}_{1i}$ by examining the distribution of $\tilde{p}_{1i}$.

[15] Harrison et al. (2005) find that monetary compensation increases the average level of risk aversion among subjects in a field experiment.

raise the average treatment effect in the trial. In practice, however, the number of such patients is likely to be small, especially if the correlation between new treatment and conventional treatment effects is not very large. Therefore, the selection bias will likely remain positive.

A second circumstance in which bias may not be positive is when the researcher cares about treatment effects in the general patient population. Now the baseline includes patients in my population of interest (those who prefer new treatment and for which selection bias is positive), plus all patients who prefer no treatment or conventional treatment. For the latter, the logic is similar to that for a monetary incentive. Among, in particular, those who prefer conventional treatment, the sign of bias is unclear because preferring conventional treatment may not imply that new treatment is less effective. So long as the correlation between new treatment and conventional treat is not too high, however, selection bias will remain positive.[16]

2. Application to ulcer trials

This section examines a data set of outcomes from 121 medical trials of three ulcer medications in order to test my model of patient self-selection. Proposition 1 provides the prediction to be tested: as the probability of treatment rises, estimated treatment effects should fall. The ulcer data confirm this prediction. Moreover, they suggest that selection bias is roughly half the size of treatment effects for two of the medications studied, and reverses the sign of treatment effects for the last medication.

2.1. Data

The data set includes the published results from clinical trials studying treatments for non-gastric ulcers. (The data set and citations for the specific studies included in the data set are available from the author.) Ulcers are the erosion of the mucous lining in the stomach or small intestine and are judged healed—a binary outcome—via endoscopy by the researcher. Three classes of medication are considered. The first class, H_2-blockers, was introduced in 1977, and is thought to prevent the production of acid in the stomach. The most popular brands are Tagamet (cimetidine), Zantac (ranitidine), and Pepcid (famotidine). The second class of medication, prostaglandins, was introduced in the mid-1980s and is thought to build up the mucous lining of the stomach and intestine. The most common prostaglandins are enprostil and misoprostil. The third class, proton-pump inhibitors (PPIs), were introduced after most prostaglandins, starting in 1989. Like H_2-blockers, these medications prevent the production of acid in the stomach. The most popular brands are Prilosec (omeprazole), Nexium (esomeprazole) and Prevacid (lansoprazole).[17]

Table 1 describes the design of the trials in the data set. There are three things to note. First, each of the trials is randomized, double-blind, and parallel-armed, which means each patient is observed in only one treatment state. Trials employ either a placebo, palliative (antacid or bismuth subcitrate), or conventional control. I will group all the placebo and palliative controlled trials and simply call them placebo controlled. H_2-blocker trials all employ placebo controls. Prostaglandin trials may employ either a placebo control or an H_2-blocker control. PPI trials only employ H_2-blockers as controls.[18] Second, the probability of treatment is calculated by taking the number of arms with the most advanced class of treatment and dividing by the total number of arms in the trial, on the theory that researchers intend to assign the same number of patients to each arm to maximize the power of comparisons across arms. (Where the numbers actually randomized to two different arms of a trial differ from this by a factor of greater than 1.5, the probability of treatment is calculated by inferring the number of subjects intended to be randomized to each arm and dividing the number that were to be randomized to treatment arms and dividing by the total number that were to be enrolled in the trial.) What introduces variation in the probability of treatment is that some trials examine multiple dosages or multiple brands of the same treatment.[19] Because each trial may have more than one treatment arm, there are more than two-times as many arms as trials reported in Table 1. Third, trials in my sample vary in the dates on which they measure treatment outcomes. Many also measure outcomes in an arm multiple times. As a result, there are more arm-date observations than there are arms.

The data set has two important limitations. First, the data contain observations only at the arm-date level, not at the patient-date level. For example, the data report the fraction of patients in an arm whose ulcer healed, but not whether an individual patient's ulcer healed. I ignore the problems raised by this sort of aggregation in the analysis below. Second, the trials in the data suffer from the attrition of patients. Average patient characteristics for an arm, however, are measured only on the date that patients are randomized to that arm. These averages are not updated as patients attrite out of an arm. The data also fail to measure outcomes for patients that attrite out of a trial. My analysis also ignores the problem of attrition. I assume that all patients that attrite out would have healed at the same rate as patients that remained in the trial, i.e. that attrition is random.[20]

Table 2 presents summary statistics for health outcomes and the characteristics of trials, treatments and patients. Each observation is on a specific treatment arm at a specific measurement

[16] There is a literature that attempts to identify the full distribution of treatment effects, though in its parlance the goal is to identify the Roy model. The data requirements, however, are typically beyond the capabilities of medical researchers. For example, they would need knowledge about either the distribution of conventional treatment effects among those who do not enroll in the trial or about the proportion of the general patient population that enrolls in the medical trial (Heckman and Honore, 1990). They do not, however, even know locally the size of this population.

[17] A distinguishing feature of all three classes of medication is that they offer a much higher chance of healing an ulcer than do antacids or bismuth subcitrate, which are mainly palliatives. That being said, it is now recognized that 90% of non-gastric ulcers are caused by the bacteria helicobacter pylori. These infections are usually treated with a combination of antibiotics and H_2-blockers or PPIs. This paper examines trials where H_2-blockers, prostaglandins or PPIs are used in isolation. These trials typically predate the switch to antibiotic-based treatments.

[18] There was one H_2-blocker trial with conventional control and one PPI trial without a conventional control. I removed these because, without a second trial, there is no possibility of variation in probability of treatment within these groups.

[19] While individuals may have different beliefs about efficacy of each class of treatment relative to the control, it is assumed that individuals do not have refined beliefs about the relative efficacy of different dosages of the same treatment or of different treatments in the same class of medication, e.g., cimetidine v. ranitidine among H_2-blockers.

[20] A more complete model would endogenize both selection into trials and attrition from trials. A natural starting point for a model of attrition is Philipson and DeSimone (1997). In that model patients remain in the trial in order "self-sample" and refine their estimates of treatment effects. Since arm assignment is blinded, one might extend the model to allow that patients self-sample in order to identify the arm to which they have been randomized. Thus a model of endogenous attrition would resemble a model of job search where the worker did not know the market distribution of wages. Because an increase in the probability of treatment above one-half reduces uncertainty about arm assignment, one might hypothesize that, holding sampled treatment effects constant, patients would attrite more quickly the higher the probability of treatment is above one-half. The effect of an increase in the probability of treatment on estimated treatment effects, however, is ambiguous. On the one hand, a patient who has a bad draw on outcomes is less likely to attrite because she can be more confident that the bad draw does not indicate she was randomized to the control group. On the other hand, she is more likely to use her bad draw to revise downward her beliefs about the value of the new treatment and thus attrite because she would prefer the conventional treatment.

Table 1
Frequencies of trials and measurements on treatment arms by design characteristic

	Treatment			Total
Frequency of trials	H_2blocker	Prostaglandin	PPI	
Total	69	24	28	121
Type of control				
Placebo, antacid or bis. subcitrate	69	14	0	83
Conv. treatment	0	10	28	38
Number of treatment arms				
1	61	20	21	102
2	6	4	6	16
3	2	0	1	3
Number of control arms				
1	66	24	27	117
2	0	0	1	1
3	3	0	0	3
Number of measurements				
1	39	6	0	45
2	25	14	20	59
3	5	4	7	16
4	0	0	1	1

	Treatment			Total
Frequency of measurements on treatment arms	H2 blocker	Prostaglandin	PPI	
Total	119	52	86	257
Type of control				
Placebo, antacid or bis. subcitrate	119	29	0	148
Conv. treatment	0	23	86	109
Probability of receiving treatment				
0.25	3	0	0	3
0.33	0	0	2	2
0.43	0	0	4	4
0.50	90	40	44	174
0.67	12	12	30	54
0.75	14	0	6	20
Date of measurement (weeks)				
1	4	0	2	6
2	21	16	36	73
3	2	0	2	4
4	63	25	36	124
6	19	8	3	30
8	7	1	7	15
10	1	0	0	1
12	2	2	0	4

date. The data are weighted such that each patient has identical weight, regardless of how many times his outcome is measured. The results are broken down by treatment. A number of the statistics warrant explanation. First, the table appears to suggest that H_2-blockers and PPIs have unconditional treatment effect, i.e., incremental probability of healing an ulcer after some weeks, of 22% and 14%, respectively. Trials in my sample, however, compare H_2-blockers to placebo, while they compare PPIs to H_2-blockers. Thus the proper interpretation is that PPIs are 14% better than H_2-blockers and 36 percent better than placebo. (One cannot infer the unconditional treatment effect of prostaglandins from the table since prostaglandin trials may have either placebo or H_2-blocker controls.) Second, exclusion criteria are rules established by researchers to exclude certain types of patients from the trial. In industry-sponsored studies (at least 47% of the sample), the purpose is to eliminate treatment non-responders and elevate estimates of treatment effects. The specific exclusion variable I measure is whether the trial excluded patients with any serious problem other than ulcers. Third, most ulcer trials without antacid as control still permit patients to consume antacids. The antacid-permitted variable is coded from 1 to 5. One indicates that subjects were prohibited from taking antacids, two that subjects were discouraged from taking antacids, three that subjects were permitted to take antacids (or the study did not counsel subjects on antacids), four that antacids were provided, and five that antacids were required.

2.2. Empirical strategy

Because the trials in the ulcer data take multiple measurements on the same patients, I estimate treatment effects using a duration model. Because the trial, treatment and patient characteristics variables for a given arm are time-invariant, I employ an exponential model for survival following assignment to treatment arm k of trial j and a constant baseline hazard: $S_{jkt} = \exp(-\lambda(\mathbf{I}_{jk}, \mathbf{Z}_j, \mathbf{X}_{jk})t)$, where survival in the aggregated ulcer data is the fraction of patients in an arm who still have an unhealed ulcer after t days of treatment. I adopt a simple linear parametrization for $\lambda\left(\mathbf{I}_{jk}, \mathbf{Z}_j, \mathbf{X}_{jk}\right) = \boldsymbol{\alpha}\mathbf{I}_{jk} + \boldsymbol{\beta}\mathbf{Z}_j + \boldsymbol{\gamma}\mathbf{X}_{jk} + \varepsilon_{jk}$, where $\mathbf{I}_{jk}$ is a vector of indicator variables indicating the medication that patients receive in that arm; $\mathbf{Z}_j$ is a vector of trial and treatment characteristics common to treatment and their matched control arms; $\mathbf{X}_{jk}$ is a vector of treatment and patient characteristics unique to the treatment arm and measured at the start of the trial; and the error term captures variation due to unobserved average characteristics of patients. I assume the errors are independent of $\left(\mathbf{I}_{jk}, \mathbf{Z}_j, \mathbf{X}_{jk}\right)$ and are independent across arms with mean zero and variance σ_{jk}.

Table 2
Summary statistics

Variable	Treatment								
	H2-blockers			Prostaglandins			PPI		
	Obs	Mean	SD	Obs	Mean	SD	Obs	Mean	SD
Share healed:									
Treatment arm	119	0.659	0.210	52	0.640	0.198	86	0.791	0.182
Matched control arm	118	0.438	0.250	52	0.589	0.300	86	0.654	0.221
Probability of treatment	119	0.570	0.117	52	0.546	0.075	86	0.536	0.099
Trial characteristics									
Exclusion criteria (no other serious problems) (0/1)?	119	0.730	0.446	52	0.610	0.492	86	0.713	0.455
Treatment characteristics for treatment arms									
Antacid permitted in trial (1–5)?	119	3.46	0.83	52	3.64	0.57	86	3.24	1.09
Frequency of dosage (times/day)	118	2.48	1.03	52	2.64	0.94	86	1.18	0.66
Total daily dosage (mg/1000)	118	0.5078	0.4407	52	0.0854	0.2258	86	0.0272	0.0180
Patient characteristics in treatment arm									
Male (0/1)?									
Treatment group	110	0.760	0.089	47	0.728	0.095	84	0.701	0.082
Control group	110	0.762	0.089	47	0.738	0.107	84	0.686	0.081
Ever smoke (0/1)?									
Treatment group	84	0.611	0.111	49	0.572	0.135	80	0.482	0.116
Control group	84	0.624	0.097	49	0.558	0.121	80	0.483	0.150
Age (log years)									
Treatment group	102	3.83	0.11	47	3.81	0.15	76	3.81	0.11
Control group	102	3.83	0.10	47	3.80	0.15	76	3.81	0.12

Notes. Each observation is a measurement on a treatment arm in a trial. Observations are weighted such that each arm makes a contribution to estimates in proportion to the number of patients in the arm, regardless of the number of measurements made on each patient.

Taking logs and dividing by $-t$, the model can be written:

$$-\ln S_{jkt}/t = \boldsymbol{\alpha}\mathbf{I}_{jk} + \boldsymbol{\beta}\mathbf{Z}_j + \boldsymbol{\gamma}\mathbf{X}_{jk} + \varepsilon_{jk}. \tag{3}$$

I posit a similar model for survival in control arms k', though that model omits common trial and treatment characteristics.

Treatment effects are estimated by comparing outcomes in treatment arms to those in matched control arms. Therefore, I estimate a single-difference version of the (3) that compares each treatment arm k to its corresponding control arm k' at date t:

$$\begin{aligned} -\left[\ln S_{jkt} - \ln S_{jk't}\right]/t &= \boldsymbol{\alpha}\mathbf{I}_{jkk'} + \boldsymbol{\beta}\mathbf{Z}_j + \boldsymbol{\gamma}\left(\mathbf{X}_{jk} - \mathbf{X}_{jk'}\right) \\ &\quad + \boldsymbol{\theta}\left[\mathbf{I}_{jkk'} \times d_j\right] + \left(\varepsilon_{jk} - \varepsilon_{jk'}\right) \end{aligned} \tag{4}$$

where, for convenience, $\mathbf{I}_{jkk'}$ are written as indicators for different treatment k/control k' combinations. Proposition 1 suggests that self-selection of patients is driven by the probability of treatment in a trial. To test this hypothesis I include as regressors the interactions between treatment/control indicators and d_j, the probability of treatment in trial j. Proposition 1 predicts specifically predicts that θ should be negative.

I estimate (4) by feasible GLS.[21] Each observation is a measurement on a treatment arm/matched control pair in a trial. Observations are weighted such that each arm makes a contribution to estimates in proportion to the number of patients in the arm, regardless of the number of measurements taken on each patient. This weighting does not materially alter the results. Standard errors are calculated assuming group-wise heteroskedasticity at the trial-level. Four specifications of the right hand side of (4) are employed. Specification (1) includes indicators for different treatment/control pairs, the interactions between these indicators and the probability of treatment d_j, and trial characteristics (exclusion criteria and exclusion criteria interacted with d_j); (2) includes (1) plus treatment characteristics (antacid role, daily frequency of medication, total daily dosage interacted with class of ulcer medication) and control characteristics (total daily dosage interacted with whether the control is an H_2-blocker); (3) includes (1) plus patient characteristics (male, smoker, log age) in the treatment arm interacted with class of treatment and patient characteristics in the control arm interacted with whether the control is an H_2-blocker; and (4) includes all of the above. The results are robust to common sense modifications of these specifications, including further interactions with d_j.

2.3. Results

Table 3 presents the core results from estimation of Eq. (4). The main finding is that coefficient estimates for the interaction of trial indicators and the probability of treatment – reported in the second section of panel C – are uniformly negative. Nearly all are statistically significant. This strongly supports Proposition 1.

In order to interpret the coefficient estimates, keep in mind, first, that the dependent variable can be interpreted as the difference in the average daily healing rate among patients in the treatment arm and patients in the control arm. So if there are 100 patients in both the treatment and control arms and the dependent variable is, say, .05, the proper interpretation is that 5 more patients healed in the treatment arm than in the control arm. Second, the median medical trial assigns 50% of patients to the treatment arm. Therefore, my estimate of treatment effects in the typical H_2-blocker trials (under specification 1) is 0.017 (0.03, the coefficient on the H_2-blocker trial indicator, plus 0.5 times -0.026, the coefficient on the interaction between probability of treatment and the H_2-blocker trial indicator). Estimates of treatment effects in 50% trials for other medications and covariate specifications are presented in Panel A of Table 4. Third, predictions of treatment effects without selection bias are identical to predictions of treatment effects in a trial where 100% of patients are assigned to new treatment. These estimates are reported in Panel B of Table 4.

What we see is that PPIs tend to perform better than H_2-blockers and H_2-blockers sometimes better, sometimes worse than prostaglandins in 50% clinical trials without correction for selection bias. (Recall that PPI's are always being tested

[21] Although the right hand side is the relative hazard rate into healing and this rate is in $[-1, 1]$, it is entirely appropriate to estimate the model using a simple linear regression. The reason is that the dependent variable, as a theoretical matter, can range from $[-\infty, \infty]$. Moreover, because individuals in an arm are aggregated, the error term is more likely to resemble a normal distribution.

Table 3
Coefficient estimates

	(1)	(2)	(3)	(4)
A. Covariate specification				
Trial characteristics (exclusion criteria)	×	×	×	×
Treatment characteristics (antacid use, frequency, dosage)		×		×
Individual characteristics (sex, smoking, age)			×	×
B. Sample size				
Number of arms	145	144	110	110
Number of arms × measurements (= total observations)	252	251	200	200
C. Coefficients				
H2-blocker v. placebo trial indicator	0.030	0.028	0.193	0.182
(p-value)	(0.0000)	(0.0060)	(0.0010)	(0.0030)
Prostaglandin v. placebo trial indicator	0.063	0.058	0.125	0.035
	(0.0000)	(0.0000)	(0.1600)	(0.7400)
Prostaglandin v. conventional control trial indicator	0.144	0.127	0.279	0.374
	(0.0000)	(0.0000)	(0.0100)	(0.0000)
PPI v. conventional control trial indicator	0.052	0.035	0.102	0.231
	(0.0000)	(0.0070)	(0.3310)	(0.0000)
Probability × H2-blocker/placebo trial	−0.026	−0.026	−0.063	−0.037
	(0.0460)	(0.1160)	(0.0000)	(0.0120)
Probability × Prostaglandin/placebo trial	−0.085	−0.075	−0.144	−0.131
	(0.0000)	(0.0000)	(0.0000)	(0.0030)
Probability × Prostaglandin/conv. treatment trial	−0.317	−0.296	−0.439	−0.484
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Probability × PPI/conv. treatment trial	−0.037	−0.021	−0.085	−0.015
	(0.1040)	(0.3570)	(0.0010)	(0.4800)

Notes. This table reports results from estimation of Eq. (4) by feasible GLS. Each observation is a measurement on a treatment arm in a trial. Observations are weighted such that each arm makes a contribution to estimates in proportion to the number of patients in the arm, regardless of the number of measurements made on each patient. P-values are based on standard errors calculated assuming group-wise heteroskedasticity at the trial-level. The dependent variable is $-[\ln S_{jk}(t) - \ln S_{jk'}(t)]/t$, where $S_{jk}(t)(S_{jk'}(t))$ is the fraction of patients who heal in trial j of treatment k (control k') at time t, measured in days. The dependent variable is calculated assuming patients who attrite out heal at the same rate as those who are evaluated. Four specifications of $(\mathbf{Z}_j, \mathbf{X}_{jk}, \mathbf{X}_{jk'})$ are employed. Specification (1) includes indicators for different treatment/control pairs, the interactions between these indicators and the probability of treatment d_j, and trial characteristics (exclusion criteria and exclusion criteria interacted with d_j); (2) includes (1) plus treatment characteristics (antacid role, daily frequency of medication, total daily dosage interacted with class of ulcer medication) and control characteristics (total daily dosage interacted with whether the control is an H_2-blocker); (3) includes (1) plus patient characteristics (male, smoker, log age) in the treatment arm interacted with class of treatment and patient characteristics in the control arm interacted with whether the control is an H_2-blocker; and (4) includes all of the above. The results are robust to common sense modifications of these specifications, including further interactions with d_j.

Table 4
Estimates of corrected treatment effects and selection bias

	(1)	(2)	(3)	(4)
A. Treatment effect in 50% trial				
H2-blocker v. placebo	0.017	0.014	0.162	0.163
	(0.0000)	(0.0000)	(0.0030)	(0.0070)
Prostaglandin v. placebo	0.020	0.021	0.053	−0.030
	(0.0000)	(0.0000)	(0.5350)	(0.7730)
Prostaglandin v. conventional treatment	−0.014	−0.021	0.060	0.132
	(0.0000)	(0.0000)	(0.4590)	(0.0360)
PPI v. conventional treatment	0.034	0.024	0.060	0.224
	(0.0000)	(0.0000)	(0.5560)	(0.0000)
B. Corrected treatment effect (predicted in 100% trial)				
H2-blocker v. placebo	0.004	0.001	0.130	0.145
	(0.5220)	(0.8790)	(0.0170)	(0.0160)
Prostaglandin v. placebo	−0.022	−0.016	−0.019	−0.095
	(0.0030)	(0.1270)	(0.8190)	(0.3620)
Prostaglandin v. conventional treatment	−0.172	−0.170	−0.160	−0.110
	(0.0000)	(0.0000)	(0.0180)	(0.0730)
PPI v. conventional treatment	0.015	0.013	0.017	0.216
	(0.1640)	(0.2530)	(0.8610)	(0.0010)
C. Difference (due to selection bias)				
H2-blocker v. placebo	−0.013	−0.013	−0.031	−0.018
	(0.0460)	(0.1160)	(0.0000)	(0.0120)
Prostaglandin v. placebo	−0.042	−0.037	−0.072	−0.065
	(0.0000)	(0.0000)	(0.0000)	(0.0030)
Prostaglandin v. conventional treatment	−0.158	−0.148	−0.219	−0.242
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
PPI v. conventional treatment	−0.018	−0.011	−0.042	−0.007
	(0.1040)	(0.3570)	(0.0010)	(0.4800)

against conventional control.) When we compare treatments after correcting for selection bias, however, PPIs emerge as superior to H_2-blockers and H_2-blockers clearly better than prostaglandins. The reason is that selection bias was stronger in prostaglandin trials than in H_2-blocker or PPI trials. There are two explanations. First, because H_2-blockers were introduced before other classes of ulcer medications, there was no conventional alternative for patients outside of trials. Therefore, the entire population of

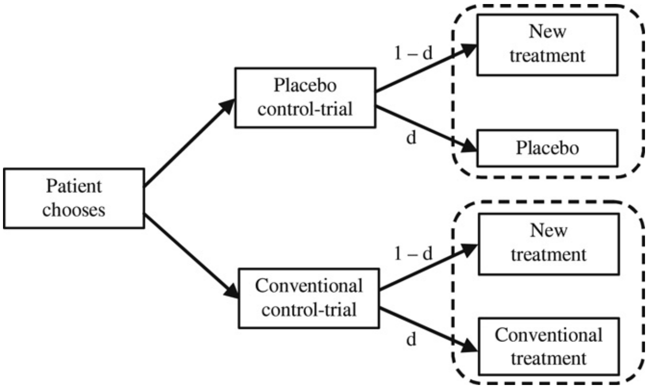

Fig. 1. Alternative "choose your control" design.

interest probably sought to enroll. Second, all the PPI trials are conventional treatment (typically H_2-blocker) controlled. Therefore, there is a smaller downside with the trial. This remains true in the case of H_2-blocker controls because even selection-biased estimates of treatment effects suggested H_2-blockers were better than prostaglandins.

How large are my estimates of selection bias? Panel C of Table 4 reports that the difference between biased and unbiased estimates of treatment effects is roughly half the size of treatment effects in 50% trials of H_2-blockers and PPIs. In trials of prostaglandins, however, the bias is roughly 2 to 6 times the size of treatment effects. In other words, in prostaglandin trials, self-selection appears to change the sign of treatment effects. A treatment that is worse than control actually appears to perform better than control without a correction for selection bias!

3. Alternative design

There is an alternative trial design that would permit the estimation of relevant treatment effects without bias from patient self-selection. Instead of offering patients a lottery over new treatment and *the investigator's choice* of control, placebo or conventional therapy, the design would offer a lottery over new treatment and *the patient's choice* of control, placebo or conventional therapy. (The alternative trial would offer the same probability of assignment to new treatment as the standard trial.) The two designs are illustrated in Fig. 1.[22]

The intuition behind the modified design is that the population of interest – defined to include only those who prefer the new treatment – includes two subgroups: those who prefer no treatment to conventional treatment and those who prefer conventional treatment to no treatment. The former will choose as their control placebo and the latter will choose conventional treatment.[23] Both subgroups, however, will enroll in the trial regardless of the probability of receiving new treatment, and estimates of treatment effects will not depend on this probability.

There are two problems with this design. First, the researcher will only be able to estimate the effect of new treatment relative to conventional treatment for the subgroup that prefers a conventional control. That information, however, is more valuable to this subgroup than an estimate of the effect of treatment relative to placebo. Second, the alternative design requires a larger sample size, and is therefore more expensive, than the standard placebo-controlled trial. The required sample size is not double, however, because the same outcomes in the new treatment arm can be used to calculate the effect of new treatment relative both to placebo and to conventional control. For example, in a parallel-armed trial, the sample size increment is only 50%.[24] The relevant question for the researcher, then, is whether selection bias is significant enough to warrant the extra cost of a larger sample.

4. Conclusion

This paper not only suggests that medical trials are subject to selection bias, but that this bias causes standard estimators to overstate treatment effects. The result is intuitive: experiments on a new treatment attract patients that are optimistic about the new treatment, and these patients probably respond better to that treatment. There is no reason to think the finding does not generalize to social experiments.

The analysis that led to this result also raises some more fundamental questions about evaluation of treatment effects. First, how do individuals form their beliefs about a new treatment, beliefs that are assumed to guide their decision whether to participate in a research study? Do they predict that new treatment has the same incremental benefit as the average medical innovation? Or do they simply experience Knightian uncertainty? If so, does decisionmaking under such uncertainty play out – with respect to study participation – like decisionmaking under risk? Second, how can one identify treatment effects in the relevant population in the presence of learning? The concern is that, if a study finds that new treatment has, say, a higher treatment effect than conventional treatment, people who previously preferred conventional treatment may now prefer the new treatment. The average treatment effect may be different in this larger population of interest than in the pre-study population of interest. This problem is an example of the principle that observation changes the system being observed.

Acknowledgments

The author thanks Gary Becker, Albert Choi, John Geweke, Bart Hamilton, James Heckman, Feifang Hu, Steve Levitt, Willard Mannng, Ron Michener, Tomas Philipson, a co-editor and an associate editor at the Journal, and an anonymous referee, as well as workshop participants at the University of Virginia and the University of Chicago, for helpful comments. He also thanks David Greene, Maulshree Solanki, and Hema Srinivasan for their excellent research assistance.

[22] The alternative design is identical to offering both a placebo-control and a conventional-control trial to the same general patient population. The design is similar, but not identical, to an unblinded trial. There we assume the patient assigned to placebo control will seek conventional treatment outside the trial. But because the patient may not report this to the researcher, estimates of treatment effects relative to placebo will be polluted with estimates of treatment effects relative to conventional control.

[23] In more formal terms, whereas the selection equation for a placebo-controlled trial is given by $d\tilde{\pi}_1 > \max\{0, \tilde{\pi}_2\}$, the selection equation for the alternative trial would be $\max\{d\tilde{\pi}_1, d\tilde{\pi}_1 + (1-d)\tilde{\pi}_2\} > \max\{0, \tilde{\pi}_2\}$, which collapses to the definition of the population of interest (2).

[24] Schouten (1999) derives the following approximation for sample sizes for a two-group trial where one group (n_2) is γ times as large as the other (n_1):

$$n_1 \geq \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \times \frac{(1+\gamma)\,\sigma^2}{\gamma\delta^2} + \frac{z^2_{1-\alpha/2}}{2\,(1+\gamma)}$$

where the groups are assumed to have the same variances σ^2, α is the desired significance level, β is the desired power, and δ is the difference the researcher would like to detect. This equation implies that the total number of patients required with unequal groups is $(1+\gamma)^2/4\gamma$ times the number required with equal group sizes. If the treatment group will be used for comparison to both controls, then $\gamma = 2$. Therefore, the total number of patients required is 9/8 that with equal group sizes. This means, instead of two trials with equal groups, the researcher will require one trial with 9/8 the size of an equal group trial and a second trial with just the smaller group, i.e., $9/8 \times 1/3$. Summing the two yields a factor of 3/2.

Appendix

This appendix provides the proof to Proposition 1, extends it to the case of continuous outcomes, and provides a list of the studies that comprise the ulcer data set.

A.1. Proof of Proposition 1

All expectations are taken with respect to g. An individual will choose to enroll if and only if $\tilde{\pi}_{1i} \geq \max\{0, \tilde{\pi}_{2i}/d\}$. Pick any two arbitrary values of d, d_A and $d_B > d_A$. Because $E(\tilde{p}_{1i}|\tilde{\pi}_{1i} \geq \max\{0, \tilde{\pi}_{2i}/d\}) = \Pr\{\tilde{\pi}_{2i} \leq 0\}E(\tilde{p}_{1i}|\tilde{\pi}_{1i} \geq 0) + \Pr\{\tilde{\pi}_{2i} > 0\}E(\tilde{p}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d)$, it is the case that

$$\begin{aligned}&\operatorname{sign}\left\{E\left(\tilde{p}_{1i}|\tilde{\pi}_{1i} > \max\{0, \tilde{\pi}_{2i}/d_A\}\right)\right.\\ &\quad \left.- E\left(\tilde{p}_{1i}|\tilde{\pi}_{1i} > \max\{0, \tilde{\pi}_{2i}/d_B\}\right)\right\}\\ &= \operatorname{sign}\left\{E\left(\tilde{p}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_A\right) - E\left(\tilde{p}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_B\right)\right\}. \qquad (5)\end{aligned}$$

Because by assumption (A1) $\tilde{\varepsilon}_{1i}$ is mean-independent of $\tilde{\pi}_{1i}$, (5) equals

$$\operatorname{sign}\{E(f(\tilde{\pi}_{1i})|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_A) - E(f(\tilde{\pi}_{1i})|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_B)\}. \qquad (6)$$

Because by assumption (A1) $f' > 0$, (6) in turn equals

$$\operatorname{sign}\{E(\tilde{\pi}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_A) - E(\tilde{\pi}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_B)\}. \qquad (7)$$

To see this, observe that the left-truncation at $\tilde{\pi}_{2i}/d_A$ for any given $\tilde{\pi}_{2i} = s$ will either be to the right of, at, or to the left of left-truncation at $\tilde{\pi}_{2i}/d_B$ (depending, as we shall see, on the correlation between $\tilde{\pi}_{1i}$ and $\tilde{\pi}_{2i}$) for $\tilde{\pi}_{2i} = s$. Assume for the moment and without loss of generality that it is truncation to the right of that at s/d_B. Because the cumulative distribution function (CDF) of the truncated-at-s/d_A distribution is identical to the CDF of a distribution with the same support as the truncated-at-s/d_B distribution but with zero mass in $[s/d_B, s/d_A)$, it is apparent that the truncated-at-s/d_A distribution first-order stochastically dominates the truncated-at-s/d_B distribution appropriately normalized. Relax the assumption that truncation at s/d_A is truncation is to the right of truncation-at-s/d_B. This same logic implies that if left-truncation at s/d_A is at or to the left of left-truncation at s/d_B, the truncated-at-s/d_A distribution will either be identical to or first-order stochastically dominated by the truncated-at-s/d_B distribution, respectively. By Theorems 1′ and 2′ in Hadar and Russell (1969), because $f' > 0$, $E(\tilde{\pi}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_A, \tilde{\pi}_{2i}) > (<)\, E(\tilde{\pi}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_B, \tilde{\pi}_{2i})$ if and only if $E(f(\tilde{\pi}_{1i})|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_A, \tilde{\pi}_{2i}) > (<)\, E(f(\tilde{\pi}_{1i})|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_B, \tilde{\pi}_{2i})$. Because the truncation points d_A and d_B are the same for all $\tilde{\pi}_{21}$, the sign of

$$E(\tilde{\pi}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_A, \tilde{\pi}_{21}) - E(\tilde{\pi}_{1i}|\tilde{\pi}_{1i} > \tilde{\pi}_{2i}/d_B, \tilde{\pi}_{21})$$

is the same for all $\tilde{\pi}_{2i}$. Therefore, the sign is preserved when integrating over $\tilde{\pi}_{2i}$.

Because we are looking at the subpopulation for whom $\tilde{\pi}_{2i} > 0$, the enrollment condition in turn implies $\tilde{\pi}_{1i} > 0$. This, along with the fact that $\ln\tilde{\pi}_{1i}$ is monotone increasing in π_{1i}, means (7) equals

$$\begin{aligned}&\operatorname{sign}\{E(\ln\tilde{\pi}_{1i}|\ln\tilde{\pi}_{1i} - \ln\tilde{\pi}_{2i} > -\ln d_A)\\ &\quad - E(\ln\tilde{\pi}_{1i}|\ln\tilde{\pi}_{1i} - \ln\tilde{\pi}_{2i} \geq -\ln d_B)\}. \qquad (8)\end{aligned}$$

Given the definitions of D and V, we can write $\ln\tilde{\pi}_{1i} = \mu_1 + a_1 D_i + V_i$ and $\ln\tilde{\pi}_{1i} - \ln\tilde{\pi}_{2i} = \mu_1 - \mu_2 + D_i$. Combined with assumption (A2) that V is mean-independent of D, this implies that (8) equals

$$\operatorname{sign}\{a_1[E(D_i|D_i > -\Delta\mu - \ln d_A) - E(D_i|D_i > -\Delta\mu - \ln d_B)]\} \qquad (9)$$

where $\Delta\mu = (\mu_1 - \mu_2)$. Since left-truncation under d_A is to the right of that under d_B, (9) equals the sign of $a_1 = \left(\tilde{\sigma}_1^2 - \tilde{\sigma}_{12}\right)/\tilde{\sigma}^2$. This is positive so long as $\tilde{\sigma}_1^2 > \tilde{\sigma}_{12}$, which is the same as the condition that $\operatorname{corr}(\ln\tilde{\pi}_{1i}, \tilde{\pi}_{2i}) < \tilde{\sigma}_1/\tilde{\sigma}_2$.

A.2. Continuous health outcomes

Suppose health is a continuous, univariate outcome, $y \in Y$, and $u(y)$ gives the utility from health for all individuals, where $u > 0$, $u' > 0$. Define $p_{ki}(y)$ as the actual distribution of outcomes for patient i given treatment k and $\pi_{ki}(y)$ as patient i's beliefs about the probability of each outcome. Let g be the population distribution of $\tilde{\mathbf{p}}_i = (\tilde{p}_{1i}, \tilde{p}_{2i})$ and $\tilde{\boldsymbol{\pi}}_i = (\tilde{\pi}_{1i}, \tilde{\pi}_{2i})$, where $\tilde{p}_{ki} = p_{ki} - p_{0i}$ and $\tilde{\pi}_{ki} = \pi_{ki} - \pi_{0i}$, for $k = 1, 2$.

The population of interest includes all individuals who satisfy

$$E_{\tilde{\pi}_1}[u(y)] \geq \max\left\{0, E_{\tilde{\pi}_2}[u(y)]\right\}$$

where I have suppressed i-subscripts for convenience. The trial population, however, includes only those for whom

$$E_{\tilde{\pi}_1}[u(y)] \geq \max\left\{0, E_{\tilde{\pi}_2}[u(y)]/d\right\}.$$

These conditions are analogous to the binary case.

Consider the following more substantive assumptions, which are slight modifications of (A1)–(A2).

(A1′) $E_{\tilde{p}_1}[u(y)] = E_{\tilde{\pi}_1}[f(u(y))] + \tilde{\varepsilon}_{1i}$, where $f' > 0$ and $\tilde{\varepsilon}_{1i}$ is mean-independent of $E_{\tilde{\pi}_1}[f(u(y))]$.

(A2′) Following Heckman and Honore (1990), assume $\left(\ln E_{\tilde{\pi}_1}[u(y)], \ln E_{\tilde{\pi}_2}[u(y)]\right)$ has finite mean (μ_1, μ_2) and variance Σ. Define $U_{ki} = \ln\left(E_{\tilde{\pi}_k}[u(y)]\right) - \mu_k$, for $k = 1, 2$, $D_i = U_{1i} - U_{2i}$, $\sigma = \sigma_{11} + \sigma_{22} - 2\sigma_{12}$, $a_1 = (\sigma_{11} - \sigma_{12})/\sigma$, $a_2 = a_1 - 1$, and $V_i = a_1 U_{2i} - a_2 U_{1i}$. By construction $U_{ki} = a_k D_i + V_i$, where D_i and V_i are uncorrelated. I shall further assume that V_i is mean-independent of D_i.

These yield the continuous analogue of Proposition 1.

Proposition 2. *Under assumptions* A1′ *and* A2′, *if* $\operatorname{corr}(\ln E_{\tilde{\pi}_1}[u(y)], \ln E_{\tilde{\pi}_2}[u(y)]) \leq \tilde{\sigma}_1/\tilde{\sigma}_2$, *then*

$$\frac{\partial E_g\left(E_{\tilde{p}_1}[u(y)]\,|E_{\tilde{\pi}_1}[u(y)] \geq \max\left\{0, E_{\tilde{\pi}_2}[u(y)]/d\right\}\right)}{\partial d} \leq 0.$$

Proof. The proof is very similar to that of Proposition 1 except for the step that shows

$$\operatorname{sign}\left\{\begin{array}{l} E_g\left[f\left(E_{\tilde{\pi}_1}(y)\right)|E_{\tilde{\pi}_1}(u) > E_{\tilde{\pi}_2}(u)/d_A\right] \\ -E_g\left[f\left(E_{\tilde{\pi}_1}(y)\right)|E_{\tilde{\pi}_1}(u) > E_{\tilde{\pi}_2}(u)/d_B\right]\end{array}\right\} \qquad (10)$$

is equivalent to

$$\operatorname{sign}\left\{\begin{array}{l} E_g\left[E_{\tilde{\pi}_1}(u)\,|E_{\tilde{\pi}_1}(u) > E_{\tilde{\pi}_2}(u)/d_A\right] \\ -E_g\left[E_{\tilde{\pi}_1}(u)\,|E_{\tilde{\pi}_1}(u) > E_{\tilde{\pi}_2}(u)/d_B\right]\end{array}\right\}. \qquad (11)$$

To demonstrate this step, perform a change of variable from y to $w = u(y)$ so that $\hat{\pi}_{ki}$ is the distribution of w when y is distributed $\tilde{\pi}_{ki}$ for $k = 1, 2$ and $\hat{p}_{1i}$ is the distribution of w when y is distributed p_{1i}. Perform a second change of variable so that $\hat{g}_j$ is the distribution of $W(j) = E_j(w)$ among the population, for $j = \left(p_{1i}, \hat{\pi}_{1i}, \hat{\pi}_{2i}\right)$. Now (11) can be written more simply as

$$\operatorname{sign}\left\{\begin{array}{l} E_{\hat{g}}\left[W\left(\hat{\pi}_1\right)|W\left(\hat{\pi}_1\right) > W\left(\hat{\pi}_2\right)/d_A\right] \\ -E_{\hat{g}}\left[W\left(\hat{\pi}_1\right)|W\left(\hat{\pi}_1\right) > W\left(\hat{\pi}_2\right)/d_B\right]\end{array}\right\}. \qquad (12)$$

Because every distribution is first-order stochastically dominated by a left-truncated version of that same distribution and because first-order stochastic dominance implies third-order stochastic dominance, (12) is greater than, equal, or less than zero depending on whether left-truncation at $W\left(\hat{\pi}_2\right)/d_A$ is to right of, equal to, or to left-truncation at $W\left(\hat{\pi}_2\right)/d_B$. Whatever the case, consider a function h that transforms $E_{\hat{\pi}_1}(w)$ into $E_{\hat{\pi}_1}\left(f\left(u^{-1}(w)\right)\right)$. Because f (by assumption (A1)) and u^{-1} are monotonically increasing,

$W\left(\hat{\pi}_1\right) = E_{\hat{\pi}_1}(w)$ does not change with w, and h does not change $\hat{\pi}_1$, it is the case that h is weakly monotonically increasing. This fact and first-order stochastic dominance implies (12) is equivalent to

$$\text{sign}\left\{\begin{array}{l} E_{\hat{g}}\left[h\left(W\left(\hat{\pi}_1\right)\right)|W\left(\hat{\pi}_1\right) > W\left(\hat{\pi}_2\right)/d_A\right] \\ -E_{\hat{g}}\left[h\left(W\left(\hat{\pi}_1\right)\right)|W\left(\hat{\pi}_1\right) > W\left(\hat{\pi}_2\right)/d_B\right] \end{array}\right\}.$$

Because the change of variable alters how an integral is written but not the value of the integral, (11) can be written as (10). ■

References

Britton, A., McKee, M., Black, N., McPherson, K., Sanderson, C., Bain, C., 1999. Threats to applicability of randomised trials: Exclusions and selective participation. Journal of Health Services Research and Policy 4 (2), 112–121.

Dunn, L.B., Gordon, N.E., 2005. Improving informed consent and enhancing recruitment for research by understanding economic behavior. Journal of the American Medical Association 293 (5), 609–612.

Ellenberg, J.H., 1994. Selection bias in observational and experimental studies. Statistics in Medicine 13 (5–7), 557–567.

For the Protection of Human Subjectsof Biomedical. T. N.C., Research, B., April 18, 1979. The belmont report.

Grady, C., Dickert, N., Jawetz, T., Gensler, G., Emanuel, E., 2005. An analysis of u.s. practices of paying research subjects. Contemporary Clinical Trials 26, 365–375.

Hadar, J., Russell, W.R., 1969. Rules for ordering uncertain prospects. American Economic Review 59 (1), 25–34.

Harrison, G.W., Lau, M.I., Rutstrom, E.E., 2005. Risk attitudes, randomization to treatment, and self-sleection into experiments. Centre for Economic and Business Research Discussion Paper.

Heckman, J., Honore, B.E., 1990. The empirical content of the roy model. Econometrica 58 (5), 1121–1149.

Heckman, J.J., 1992. Randomization and social policy evaluation. In: Manski, C., Garfinkel, I. (Eds.), Evaluating Welfare and Training Programs. Harvard University Press, pp. 201–230.

Heckman, J.J., 1996. Randomization as an instrumental variable. Review of Economics and Statistics 78 (2), 336–341.

Heckman, J.J., Smith, J.A., 1995. Assessing the case for social experiments. Journal of Economic Perspectives 9 (2), 85–110.

Hunter, J.J., 1972. Independence, conditional expectation, and zero covariance. American Statistician 26 (5), 22–24.

Kamoinka, T., Lacroix, G., 2002. Assessing the extent of randomization bias in the canadian self-sufficiency experiment. Working paper.

Lancaster, H., 1959. Zero correlation and independence. Australian Journal of Statistics 1 (3), 53–56.

Leipnik, R., 1961. When does zero correlation imply independence?. American Mathematical Monthly 68 (6), 563–565.

National Cancer Institute. 2006a. Informed consent template for cancer treatment trials.

National Cancer Institute. 2006b. Simplification of informed consent documents.

Olschewski, M., Schumacher, M., Davis, K.B., 1992. Analysis of randomized and nonrandomized patients in clinical trials using the comprehensive cohort follow-up study design. Controlled Clinical Trials 13, 226–239.

Pfizer,, 2006. Policy on compensation to human research subjects.

Philipson, T., 2000. External treatment effects and program implementation bias.

Philipson, T., DeSimone, J., 1997. Experiments and subject sampling. Biometrika 84, 221–234.

Resier, S.J., 2005. Research compensation and monetarization of medicine. Journal of the American Medical Association 293 (5), 613–614.

Robinson, D., Woerner, M.G., Pollack, S., Lerner, G., 1996. Subject selection biases in clinical trials: Data from a multicenter schizophrenia treatment study. Journal of Clinical Psychopharmacology 16 (2), 170–176.

Roy, A., 1951. Some thoughts on the distribution of earnings. Oxford Economic Papers (New Series) 3, 135–146.

Schouten, H.J.A., 1999. Sample size formula with a continuous outcome for unequal group sizes and unequal variances. Statistics in Medicine 18, 87–91.

Senore, C., Battista, R.N., Ponti, A., Segnan, N., Shapiro, S.H., Rosso, S., Aimar, D., 1999. Comparing participants and nonparticipants in a smoking cessation trial: Selection factors associated with general practitioner recruitment activity. Journal of Climincal Epidemiology 52 (1), 83–89.

University of North Carolina, Office of Clinical Trials, 2004. Developing good recruitment practices. Office of Clinical Trials Newsletter 2(4), 1, 3.

Verheggen, F.W., Nieman, F., Jonkers, R., 1998. Determinants of patient participation in clinical studies requiring informed consent: Why patients enter a clinical trial. Patient Education and Counseling 35 (2), 111–125.

World Medical Association. 2000. Declaration of helsinki: Ethical principles for medical research involving human subjects.