Controls, Belief Updating, and Bias in Medical RCTs*

Gilles Chemla
Imperial College Business School, DRM/CNRS, and CEPR.

Christopher A. Hennessy
London Business School, CEPR, and ECGI.

June 20, 2019

Abstract

We develop a formal model of placebo effects. If subjects in seemingly-ideal single-stage RCTs update beliefs about breakthroughs based upon personal physiological responses, mental effects differ across medications received, treatment versus control. Consequently, the average cross-arm health difference becomes a biased estimator. Constructively, we show: bias can be altered through choice of control; higher-efficacy controls mitigate upward bias; and efficacy states can be revealed through controls of intermediate efficacy or controls that mimic a subset of efficacy states. Consistent with experimental evidence, our theory implies outcomes within-arm and cross-arm differences can be non-monotone in treatment probability. Finally, we develop novel differences-in-differences and triangle equality tests to detect RCT bias.

Keywords: Belief updating, medical RCTs, bias, control, treatment.

JEL: C10, C90, D04, I11, I10, K32, O3.

*We thank seminar participants at CEPR European Summer Symposium in Economic Theory (Experimental), Columbia, USC Price School, UCLA, Zurich, Stockholm, Imperial College, IESE, INSEAD and Baruch College. We also thank Bruce Carlin, John Rust, Jacob Sagi, Fanny Forssell Forsberg, and Jan Starmans for early feedback. Finally, we thank two anonymous referees and the editor. Corresponding author. Gilles Chemla, Imperial College Business School, London SW7 2AZ, United Kingdom; e-mail: g.chemla@imperial.ac.uk. This research was supported in part by a European Research Council Grant.

1 Introduction

A critical stated objective in medical trials is to produce an unbiased estimate of the non-placebo physiological effect of a treatment, also known in the medical literature as *characteristic effect* or *specific effect*. Since Fisher (1935), the double-blind randomized controlled trial (RCT below) has been viewed as the gold standard in eliminating placebo effects and isolating non-placebo physiological effects. In describing the rise of RCTs in medicine in *The Lancet*, Kaptchuk (1998) notes, "The greater the placebo's power, the more necessity there was for the masked RCT itself." In the U.S., E.U. and Japan, the gold standard status of RCTs is codified under the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH, 2000). The ICH writes, "Control groups have one major purpose: to allow discrimination of patient outcomes [...] caused by the test treatment from outcomes caused by other factors, such as the natural progression of the disease, observer or patient expectations, or other treatment." In fact, the perceived reliability of the RCT has caused the methodology to be emulated in other disciplines. For example, in their influential textbook, *Mostly Harmless Econometrics*, Angrist and Pischke (2009) hold up the RCT as the ideal for achieving unbiased estimates of causal effects.

Given its practical importance, as well as its contemporary methodological influence, it is worthwhile to revisit the logical foundations of the double-blind medical RCT. In the traditional statistical argument for the RCT, health quality is assumed to be the sum of a direct non-placebo physiological effect plus a brain-modulated physiological ("mental" or "placebo") effect.[1] Since subjects are randomly assigned to treatment and control groups, with blind assignment, the mental effects are assumed to be identically distributed random variables, having probability distributions independent of the assigned group. Under these assumptions, the difference between the average treatment and control group health quality yields an unbiased estimate of the expectation of the non-placebo physiological effect.

The medical literature's informal theoretical account of mental effects is known as *expectancy*

[1] Following the literature, we use these three terms interchangeably.

theory. Stewart-Williams and Podd (2004) state, "On the expectancy account, the effects of such factors come through their influence on the placebo recipient's expectancies." In this spirit, in perhaps the first known controlled medical trial, Haygarth (1801) wrote that his study, "clearly prove[d] what wonderful effects the passions of hope and faith, excited by mere imagination, can produce on disease."

Departing from purely statistical treatments, as well as informal articulations of expectancy theory found in the medical and psychiatry literatures, we posit that agents manifest better present-day health quality in response to expectations of better health quality in future periods arising from a higher probability of a medical breakthrough. For example, expectation of better future health can reduce anxiety, improving outcomes today for subjects suffering from ulcers or hypercholesterolemia. Similarly, expectation of higher future survival rates can alleviate the severe anxiety associated with life-threatening diseases, with relaxation, rest and sleep improving measured health quality today. Similarly, expectation of better future mental health can mitigate feelings of hopelessness and anxiety, thus reducing depression today.

Indeed, the psychiatry literature, with its *anxiety theory*, has informally articulated the mechanism central in our model, without apparently understanding the implications for RCTs. MacLeod, Williams and Bekerian (1991) argue that: "Worry is a cognitive phenomenon, it is concerned with future events where there is uncertainty about the outcome, the future being thought about is a negative one, and this is accompanied by feelings of anxiety." In turn, anxiety reduction may represent one source of mental effects. For example, in the context of pain treatment, Turner, Deyo, Loeser, von Korff, and Fordyce (1994) conjecture that "A patient's expectation that treatment will relieve symptoms may reduce anxiety and thus ameliorate symptoms."

In a stylized way, our model mimics a seemingly-ideal medical RCT. Specifically, we consider the testing of a new medication with an unobservable efficacy state that is either high or low. Test subjects enter the trial holding subjective prior probability assessments which may or may not be equivalent to objective probabilities.[2] Subjects are randomly assigned to take the new medication

[2] See Anscombe and Aumann (1963) for a definition of subjective probability.

or a control. After taking their assigned medication, but before health quality is measured, subjects privately observe their respective direct physiological state and then revise their beliefs regarding the efficacy state using either Bayesian or subjective non-Bayesian updating as in Epstein (2006) and Epstein, Noor and Sandroni (2008). Measured health quality at the end of the RCT is the sum of the direct physiological state plus a brain-modulated physiological effect which is a monotone increasing function of expected future health quality. In turn, expected future health quality is monotone increasing in the subject's posterior probability assessment of the high efficacy state.

As shown, in such environments, *mental effects cannot be presumed to be equal across treatment and control groups even in ideal double-blind single-stage medical RCTs.* In particular, beliefs about the efficacy state of a novel medical treatment will vary systematically with the objective probability distributions governing the direct physiological responses induced by the treatment and control medications. Unless the objective probability distributions of direct physiological responses are equal across the treatment and control medications, beliefs regarding efficacy will generally differ across groups leading to differences in hope-based mental effects (expectancy). Therefore, the difference in average health outcomes across treatment and control groups generally delivers a biased estimate of the conditional expectation of the non-placebo physiological effect.[3]

We first characterize how bias varies with choice of control. We initially assume priors have the monotone likelihood-ratio property (MLRP) while objective probabilities satisfy first-order stochastic dominance (FOSD) conditions. It is shown that an unscrupulous drug manufacturer can create upward bias by using a stochastically dominated control. Conversely, a conservative regulator can ensure that bias is downward by using a stochastically dominating control. Controls of intermediate efficacy create upward (downward) bias in the high (low) efficacy state. Finally, despite existence of bias, a positive result obtained under these technical conditions is that the treatment-control difference is positive (negative) if the treatment generates higher (lower) mean health than the control. That is, here the treatment-control difference correctly ranks medications in terms of mean health outcomes.

[3] Conditioning here is on the efficacy state.

We then relax the MLRP and FOSD assumptions. It is shown that here the treatment-control difference need not rank medications correctly: a positive (negative) treatment-control difference can occur even though the treatment is no better (no worse) than the control in terms of mean health effects. Nevertheless, other constructive results emerge here. For example, the unconditional expectation of bias is zero if the objective probability density for the control mimics the objective unconditional density of health outcomes under the treatment. Further, bias can be eliminated in one of the two efficacy states by using a control mimicking that state's objective probability density. Finally, with such a mimicking control, a non-zero treatment-control difference identifies the true efficacy state as being the non-mimicked state.

We next develop a novel differences-in-differences test for RCT bias and for our posited control medication effect. In particular, if RCTs are unbiased, then the treatment-control difference should fall one-for-one with each increase in the mean health outcome associated with different controls. In contrast, we predict that when more effective controls are used in RCTs, the treatment-control difference will fall more than one-for-one with the control's mean outcome. That is, the difference between RCT differences should exceed the difference between control medication effects. If the objective effects of the two control medications are not known, a third RCT comparing the two controls must be conducted. In the absence of RCT bias, the following triangle equality will hold: The difference in RCT outcomes between the novel treatment and control 1 is equal to the difference between the novel treatment and control 2 plus the difference between control 2 and control 1.[4]

We also analyze the role played by treatment probability in altering the bias. To begin, we show bias approaches zero as treatment probability approaches zero.[5] Of course, concern over standard errors and eliciting participation would rule out infinitesimal treatment probabilities in practice. We therefore derive technical conditions under which bias is increasing in treatment probability. These monotonicity results can be seen as supporting the notion that smaller treatment groups are bias-reducing. However, the conditions for bias-monotonicity are restrictive, and we show that bias

[4] We thank an anonymous referee for suggesting this extension of our DiD test.

[5] Chassang, Padro i Miguel, and Snowberg (2012) present a similar limit result.

magnitude can locally decrease with treatment probability.

Finally, we derive testable implications regarding the effect of varying treatment probability. Here our model has the potential to explain experimental evidence at odds with the "canonical theory" of placebo effects. The canonical theory predicts better mental effects arise from beliefs that one is receiving the treatment rather than the control. Stewart-Williams and Podd (2004) describe the "archetypal placebo event" as follows: "A physician gives a patient a pill that, unbeknownst to the patient, is merely a sugar pill... Presently, the patient's health improves, apparently because of the belief that the pill was a pharmacological agent, effective for the condition." Consistent with the canonical theory, Malani (2006) writes, "patients in the higher-probability [of treatment] trial will expect better health outcomes from their clinical trial, all other things being equal."

We contrast testable implications. First, both theories predict outcomes within each trial arm should vary with treatment probability. Second, in the canonical placebo theory, health outcomes within each trial arm are predicted to increase with treatment probability. In this spirit, Chassang, Snowberg, Seymour and Bowles (2015) offer antidepressants as a motivating example, writing "participants treated with probability $p = 50\%$ (1/1 odds) will expect more social anxiety than participants treated with probability $p = 75\%$ (3/1 odds)."[6] In contrast, our model formally predicts negative responses to higher treatment probabilities if low-efficacy medications are administered, since subjects then assess a lower probability of a breakthrough. Finally, the canonical theory predicts health outcomes across arms should increase at equal rates, with cross-arm mental effects canceling, implying zero bias. In contrast, our proposed theory predicts unequal response rates, implying RCT bias.

In a pioneering paper, Malani (2006) tests these three hypotheses in a sample of medical RCTs with varying treatment probabilities for subjects suffering from ulcers or hypercholesterolemia. He finds that health outcomes do indeed vary with treatment probabilities. However, he finds some evidence of negative treatment-arm responses to increases in treatment probability, an observation inconsistent with the canonical theory. Further, Malani documents that treatment and control arms

[6] The authors also discuss potential negative (nocebo) effects.

exhibit unequal rates of response to changes in treatment probability, contradicting the canonical theory but supporting our central prediction of bias in medical RCTs.

Chassang, Padro i Miguel, and Snowberg (2012) analyse RCTs in settings with hidden types or actions. They consider the normal form of an abstract signal structure. In contrast, we consider the extensive form of a specific signal technology and provide a detailed analytical treatment of how bias is affected by changes in control medication (and treatment probabilities). Relatedly, they show how mechanism-like extensions can improve on RCTs. In contrast, we attempt to improve on RCTs through bias-reducing or state-revealing control medications.

Philipson and Desimone (1997) also stress limitations on RCTs. In their model, bias arises via differential attrition in multi-stage trials, there is a single efficacy state for the novel drug, and the health signal is binary.[7] Our framework shows the problem of RCT bias is much more severe than their analysis suggests. First, as we show, with placebo effects, bias can even occur in single-stage trials where attrition is impossible. Second, if there is more than one efficacy state for the novel drug, as is the case in our model, and as must be the case in practice if an RCT is actually resolving uncertainty, the RCT must be biased in at least one efficacy state. That is, at best, bias elimination is state-contingent. Finally, with binary health signals, the sufficient condition for no bias is that treatment and control have *equal means*. In a more realistic environment in which health varies along a continuum, the analog sufficient condition for no-bias is much more-demanding: equality of treatment and control *probability density functions*. In addition, with continuous health signals, differences in density functions can change ordinal rankings of medications in a way that depends upon the shape of mental effect functions, in the same way that changes in security payoff functions change preferences over payoff densities.

The mental effect in our model operates through subject beliefs regarding the promise of *future* health, which is positively correlated with the true efficacy state of the novel therapy being tested. We posit that less anxiety about future health leads to better health quality during the experiment. In this sense, the underlying causal channel in our model is related to the anticipatory expected

[7] See Chan and Hamilton (2006) for a structural estimation and Deaton (2010) for a discussion of attrition.

utility framework formalized by Caplin and Leahy (2001). Specifically, the two models share the notion that "anxiety is anticipatory," with our model going on to consider concomitant health quality feedback effects.

Our paper contributes to the medical literature on RCTs. Existing critical examinations of medical RCTs have emphasized practical difficulties in their implementation. Bothwell and Podolsky (2016) provide an historical account of RCTs. Rothwell (2005, 2006) provides excellent critical surveys. There is also a large medical literature examining the role of expectations in influencing health outcomes. Regarding the potentially beneficial effects of greater hope, Blasi, Harkness, Ernst, Georgiou, and Kleijnen (2001) perform a meta-analysis of context effects in *The Lancet*. They note, "Three of these studies showed that enhancing patients' expectations through positive information about the treatment or the illness, while providing support or reassurance, significantly influenced health outcomes." Consistent with the notion that anxiety reduction leads to better health outcomes, Thomas (1987) offers empirical evidence that physicians offering to patients a more positive prognosis, holding fixed the nature of treatment, leads to reductions in reported symptoms. In addition, Shapiro and Shapiro (1984) offer evidence that placebo effects are more powerful in more anxious patients.

2 The Model

Unless stated otherwise, the details of the experimental setting are common knowledge. This is in the spirit of informed consent laws. There are two dates $d \in \{1, 2\}$. At $d = 1$, a double-blind randomized, parallel group, controlled trial (RCT) is conducted. The objective of the RCT is to assess the efficacy of a novel drug.[8] The assessed efficacy influences the probability of the tested drug being distributed at $d = 2$, as well as the probability of next-generation improvements to the drug. Depending on the setting, one can think of the control group as being given either an inert drug (placebo-controlled trial) or some traditionally-used drug (controlled trial).

[8] We consider drugs to fix ideas, but the analysis applies to medical treatments generally.

In the model, the RCT is ideal, with the test panel being representative of those currently afflicted with the disease, as well as future generations that will be afflicted with the disease, eliminating selection concerns.[9] In addition, the test panel $\mathcal{I}$ is a measure one continuum of agents, eliminating potential concern over small sample bias.[10] Current and future agents suffering from the disease live for two dates, with the dates $d \in \{1, 2\}$ corresponding to the lifetime of agents in the test panel.

Let $\mathcal{T}$ ($\mathcal{C}$) denote the set of agents randomly assigned to the treatment (control) group. The measure of the treatment group is $t \in (0, 1)$. Throughout, t is assumed to be common knowledge, consistent with treatment probability being an important element of informed consent.

We begin first with a description of objective probabilities. Just after taking her assigned drug at $d = 1$, agent $i \in \mathcal{I}$ observes her respective *direct physiological state* p_i^1, a random variable with support $\mathcal{P} \equiv \left[\underline{p}, \overline{p}\right]$, with $\mathcal{P}$ being common knowledge. The direct physiological state represents the health quality that would be experienced by the agent in the absence of any mental effect. If $i \in \mathcal{C}$, p_i^1 is an independent draw from an atomless twice continuously differentiable cumulative distribution F^C, with probability density f^C. If $i \in \mathcal{T}$, p_i^1 is an independent draw from an atomless twice continuously differentiable cumulative distribution F^S, with probability density f^S. Let S denote the *efficacy state* of the new drug. This efficacy state is unknown at the start of the RCT. The efficacy state $S \in \{L, H\}$, with L (H) denoting low (high) efficacy. The objective probability that $S = H$ is $\lambda \in (0, 1)$.

In the interest of generality, we do not assume the experimenter knows any element of $\{f^C, f^H, f^L, \lambda\}$. Common knowledge is subsumed as a special case of the model in which the experimenter and all $i \in \mathcal{I}$ know $\{f^C, f^H, f^L, \lambda\}$. As specified in the respective propositions below, the experimenter may only know that certain relationships hold. For example, the experimenter may know the control mimics the distribution of direct physiological states induced by the new drug in the low efficacy state, with $f^C = f^L$. Alternatively, the experimenter may know certain stochastic dominance relationships hold. For example, the control may be an inert pill and the experimenter may know

[9] See Malani (2008) for a detailed analysis of self-selection in RCTs.

[10] Deaton (2010) expresses concern over small sample biases in RCTs.

that, regardless of the efficacy state, the distribution of direct physiological states induced by the novel drug first-order stochastically dominates that of the control. One of our objectives will be to describe whether and how the experimenter can draw correct inferences about either the true efficacy state S or the expectation of the direct physiological state induced by the novel drug.

We also admit departures from common knowledge by allowing for the possibility that test subjects do not know any element of $\{f^C, f^H, f^L, \lambda\}$. Rather, test subjects are assumed to enter the RCT holding subjective probability assessments. Test subject i enters the trial holding the subjective probability assessment that the direct physiological state induced by the control medication is described by a twice continuously differentiable cumulative distribution F_i^C, with probability density f_i^C. Further, test subject i enters the trial holding the subjective probability assessment that, conditional upon the efficacy state being S, the direct physiological state induced by the novel medication is described by an atomless twice continuously differentiable cumulative distribution F_i^S, with probability density f_i^S. Finally, test subject i holds the subjective prior that $S = H$ with probability $\lambda_i \in (0, 1)$.

Anticipating, some propositions below assume subjective probabilities exhibit a degree of concordance with objective probabilities, with subsequent analysis relaxing the concordance assumption. Intuitively, one can understand informed consent forms and other information sources as providing test subjects with a rough understanding of their environment even though they fail to know any element of $\{f^C, f^H, f^L, \lambda\}$. For example, the control might be an inert pill that mimics the low efficacy state ($f^C = f^L$), and the test subject may be informed that this the case, so that $f_i^C = f_i^L$. Notice, here there is a degree of concordance between subjective beliefs and objective reality but this in no way rules out extremely large Kullback-Liebler divergence (relative entropy) between subjective and objective densities.

Let

$$\mu_{(i)}^J \equiv \int_{\mathcal{P}} p f_{(i)}^J(p) dp.$$

State H is superior to L in that

$$\mu^H > \mu^L.$$

The following technical assumption is adopted.

Assumption 1: *For all $i \in \mathcal{I}$ and $p \in \mathcal{P}$ there exists some $J \in \{C, L, H\}$ such that $f_i^J(p) > 0$ and for all $p \in (\underline{p}, \overline{p})$, $f_i^J(p) > 0$ for all $J \in \{C, L, H\}$.*

The first part of Assumption 1 ensures each test subject's Bayesian posterior probability assessment over S is well-defined on $\mathcal{P}$. The second part ensures the first derivative of the Bayesian posterior is well-defined on $(\underline{p}, \overline{p})$.

Health quality is measured without error. During the RCT ($d = 1$), agent $i \in \mathcal{I}$ experiences total health quality Q_i^1, where

$$Q_i^1 \equiv p_i^1 + m_i. \tag{1}$$

In the preceding equation, the first term, the direct physiological state, is assumed to be privately observed by the agent. The second term captures the mental effect. Since additivity of the mental effect plays an important role in the traditional proof for the unbiasedness of RCTs, we stress the adoption of this assumption.

Assumption 2: *The brain-modulated physiological effect (mental effect) enters health quality additively.*

Events at $d = 2$ are modeled in reduced-form with, say, the government mandating which medication must be taken by all agents suffering from the disease, including all $i \in \mathcal{I}$. Test subject i holds the subjective probability assessment that if the efficacy state is S, then at $d = 2$ the novel drug will be mandated with probability π_i^S.[11] Test subject i believes improvements to the novel drug will be made to it, conditional upon it becoming the mandated drug at $d = 2$, increasing the health quality of those who take it by a further incremental amount δ_i^S, with $\delta_i^H \geq \delta_i^L \geq 0$.

For simplicity, the effect of all medications is assumed to be non-cumulative. In particular, the terminal period health quality of agent $i \in \mathcal{I}$ at date $d = 2$, denoted p_i^2, is an independent draw of

[11] This subjective probability assessment may well reflect test subjects anticipating incorrect experimenter inferences.

the direct physiological state from the relevant distribution, specifically F^S if the novel drug is the mandated drug. If the novel drug is not mandated at $d = 2$, a *default medication* will be mandated, with the default medication being viewed by test subject i as generating mean health quality at $d = 2$ equal to μ_i^D.

After taking the assigned pill, but *before* the experimenter measures total health quality Q_i^1, test subjects observe their respective direct physiological state p_i^1 and then update their beliefs regarding the efficacy state. For example, a test subject may take a pill at the start of the RCT, feel better (worse) some hours (days) later, and then feel more (less) optimistic regarding the prospects that the novel drug constitutes a true medical breakthrough. Since this timing convention underpins the central causal mechanism in our model, we stress this assumption below.

Assumption 3: *Each test subject observes their respective draw of direct physiological state p_i^1 before the experimenter measures Q_i^1.*

Let $\widehat{\beta}_i$ denote the posterior probability assessment of agent i that $S = H$ given p_i^1. We allow test subjects to depart from Bayes' Rule, with posterior beliefs falling into a broader class consistent with subjective updating of subjective priors as in the axiomatic formulation of Epstein (2006) and Epstein, Noor and Sandroni (2008). We are agnostic as to whether the departures from Bayes' Rule constitute mistakes or whether, as argued by Epstein (2006), departures from Bayes' Rule can be understood as rational provided that priors are subjective objects rather than objective objects.[12]

Assumption 4: *For each test subject $i \in \mathcal{I}$ there exist weights $(\kappa_i^{\leq}, \kappa_i^{>}) \in (0, \infty) \times (0, \infty)$ such that posterior beliefs take the form*

$$\begin{aligned} \widehat{\beta}_i(p) &= \kappa_i^{\leq} \beta_i(p) + (1 - \kappa_i^{\leq})\lambda_i, \ \forall \ p \text{ s.t. } \beta_i(p) \leq \lambda_i \\ \widehat{\beta}_i(p) &= \kappa_i^{>} \beta_i(p) + (1 - \kappa_i^{>})\lambda_i, \ \forall \ p \text{ s.t. } \beta_i(p) > \lambda_i \end{aligned}$$

where β_i is the subject's probability assessment that $S = H$ derived from Bayes' Rule.

Assumption 4 ensures our model subsumes a broad class of behavioral assumptions as one considers possible combinations of weights on the pure Bayesian posterior. A Bayesian agent places

[12] In particular, see the discussion on page 415 in Epstein (2006).

weight 1 on β. Underreaction (overreaction) entails placing a weight less (greater) than 1 on β. Finally, one can think of an optimistic (pessimistic) agent as overreacting to positive (negative) signals and underreacting to negative (positive) signals. Summarizing, such subjective updating rules are subsumed within our model by setting parameters as follows.

$$\begin{aligned} \text{Bayesian:} \quad & \kappa_i^{\leq} = \kappa_i^{>} = 1 \\ \text{Underreaction:} \quad & (\kappa_i^{\leq}, \kappa_i^{>}) \in (0,1) \times (0,1) \\ \text{Overreaction:} \quad & (\kappa_i^{\leq}, \kappa_i^{>}) \in (1,\infty) \times (1,\infty) \\ \text{Optimism:} \quad & \kappa_i^{\leq} \in (0,1],\ \kappa_i^{>} > 1 \\ \text{Pessimism:} \quad & \kappa_i^{\leq} > 1,\ \kappa_i^{>} \in (0,1]. \end{aligned} \tag{2}$$

Let X_i denote the expectation held by agent i regarding their terminal period ($d = 2$) health quality conditional upon the direct physiological state they experienced during the RCT:

$$X_i(p) \equiv \mathbb{E}_i[Q_i^2 | p_i^1 = p].$$

It follows:

$$\begin{aligned} X_i(p) &= \overline{Q}_i^{2L} + \widehat{\beta}_i(p) \left(\Delta \overline{Q}_i^2 \right) \\ \Delta \overline{Q}_i^2 &\equiv \overline{Q}^{2H} - \overline{Q}^{2L} \\ \overline{Q}_i^{2H} &\equiv \mathbb{E}_i[Q_i^2 | S = H] = \pi_i^H(\mu_i^H + \delta_i^H) + (1 - \pi_i^H)\mu_i^D \\ \overline{Q}_i^{2L} &\equiv \mathbb{E}_i[Q_i^2 | S = L] = \pi_i^L(\mu_i^L + \delta_i^L) + (1 - \pi_i^L)\mu_i^D. \end{aligned} \tag{3}$$

Three points are worth noting in the preceding equation. First, expected future health quality (X_i) varies with the state-contingent continuation probabilities (π_i^L, π_i^H), which can be viewed as being influenced by test subject beliefs regarding company or country-specific factors, e.g. financial and legal constraints, rather than representing the sort of biological constants of interest to physicians. Second, expected future health capitalizes not only the effects of the novel drug being tested but also anticipated improvements to this new drug, as captured by the improvement parameters (δ_i^L, δ_i^H).

Thus, the model reveals that the mental effect enjoyed by the current drug capitalizes the anticipated efficacy of next-generation drugs. Third, expected future health quality capitalizes efficacy states that may well be disproved by the RCT. It is apparent then that mental effects will generally be time-varying since information sets generally vary over time. All three factors show that, as a general matter, it is invalid to extrapolate across time and space the placebo effects from a given RCT.

Test subjects expect higher future health quality if $S = H$. In particular, we make the following assumption.

Assumption 5: *For all test subjects* $i \in \mathcal{I}$

$$\pi_i^H(\mu_i^H + \delta_i^H) + (1 - \pi_i^H)\mu_i^D > \pi_i^L(\mu_i^L + \delta_i^L) + (1 - \pi_i^L)\mu_i^D.$$

Although we are agnostic regarding the exact parameter configuration underpinning Assumption 5, perhaps the most natural interpretation is that test subjects assign relatively high values to π_i^H and μ_i^H. That is, subjects may believe that if the novel drug is of high efficacy, there is a high probability of its being distributed and, in this state, the drug generates high mean health outcomes.

Expectancy theory posits that better expected future health quality maps to better present-day health quality. We capture such a mapping with the following functional form assumption.

Assumption 6: *Brain-modulated physiological effects (mental effects) are equal to* $\Psi_i(X_i)$, *where* Ψ_i *is continuously differentiable and strictly increasing for all* $i \in \mathcal{I}$.

With Assumption 6 in-hand, the mental component of measured health quality at $d = 1$ (equation (1)) can be expressed as follows:

$$\begin{aligned} m_i &= M_i(p_i^1) \\ M_i(\cdot) &\equiv \Psi_i\left[X_i(\cdot)\right] = \Psi_i\left[\overline{Q}_i^{2L} + \widehat{\beta}_i(\cdot)\left(\Delta\overline{Q}_i^2\right)\right]. \end{aligned} \tag{4}$$

As reflected in the preceding equation, in this formalization of expectancy theory, beneficial brain-modulated physiological effects are driven by optimism about the efficacy of future medication, i.e.

hope, not by beliefs regarding whether one is in the treatment group or the control group. The key probability assessment is $\widehat{\beta}_i$, the subjective posterior probability assessment that the efficacy state is H.

The objective of the RCT is to estimate the expectation of the direct non-placebo effect of the drug on health quality. Let:

$$\text{Direct Effect} \equiv \mathbb{E}[p_i^1 | i \in \mathcal{T}, S] - \mathbb{E}[p_i^1 | i \in \mathcal{C}, S] = \mu^S - \mu^C. \tag{5}$$

The expected treatment-control health quality difference is:

$$\mathbb{E}[Q_i^1 | i \in \mathcal{T}, S] - \mathbb{E}[Q_i^1 | i \in \mathcal{C}, S] = \mu^S - \mu^C + \{\mathbb{E}[m_i | i \in \mathcal{T}, S] - \mathbb{E}[m_i | i \in \mathcal{C}, S]\}. \tag{6}$$

From equation (4) it follows:

$$\mathbb{E}[Q_i^1 | i \in \mathcal{T}, S] - \mathbb{E}[Q_i^1 | i \in \mathcal{C}, S] = \underbrace{\mu^S - \mu^C}_{\text{Direct Effect}} + \underbrace{\int_{\mathcal{I}} \left[\int_{\mathcal{P}} \Psi_i \left[\widehat{\beta}_i(p) \overline{Q}_i^{2H} + (1 - \widehat{\beta}_i(p)) \overline{Q}_i^{2L} \right] \right] di.}_{\text{Bias}} \tag{7}$$
$$[f^S(p) - f^C(p)]dp$$

Notice, bias depends upon the objective probability densities $\{f^C, f^H, f^L\}$, as well as the subjective posteriors $(\widehat{\beta}_i)$ which, in turn, depend upon the subjective priors $\{f_i^C, f_i^H, f_i^L\}$

Under the traditional interpretation, the bias term in equation (6) is equal to zero, implying the mean treatment-control health quality difference yields an unbiased estimate of the mean of the direct non-placebo physiological effect of the treatment relative to the control. Thus, the absence of bias in RCTs can be understood as being predicated upon two assumptions: additivity and i.i.d. mental effects.

3 Analysis of Posterior Beliefs

This section characterizes how test subjects update beliefs in reaction to the interim signal they receive, their respective direct physiological state (p_i^1).

In order to illustrate the central role played by updating, we begin by presenting a sufficient condition for absence of bias.

Proposition 1 *A sufficient condition for absence of bias, regardless of the true efficacy state S, is that for all $i \in \mathcal{I}$ posterior beliefs are equal to prior beliefs ($\widehat{\beta}_i = \lambda_i$), a condition satisfied if subjects do not observe any interim signal, in violation of Assumption 3, or subjects place zero weight on the Bayesian posterior, in violation of Assumption 4.*

Proof. If $\widehat{\beta}_i(p) = \lambda_i \; \forall \; p \in \mathcal{P}$, the bias term in equation (7) is 0.■

The intuition for the preceding proposition is as follows. Under the stated conditions, test subjects cling to their prior beliefs regardless of the direct physiological state they experience during the RCT. In this case, the fact that the treatment and control groups are drawing from different objective probability distributions is inconsequential in terms of hope-based mental effects.

3.1 Bayesian Posteriors

Since the Bayesian posterior always enters a test subject's subjective posterior with positive weight (Assumption 4), it is useful to consider some of its properties. From Bayes' Rule it follows:

$$\beta_i(p) = \frac{\lambda_i \left[t f_i^H(p) + (1-t) f_i^C(p)\right]}{t\left[\lambda_i f_i^H(p) + (1-\lambda_i) f_i^L(p)\right] + (1-t) f_i^C(p)}. \tag{8}$$

Notice, the Bayesian posterior β_i regarding the efficacy state is distinct from the Bayesian posterior regarding assignment to the treatment group. To see this, let b_i denote the posterior probability assessment of having been in the treatment group. From Bayes' Rule it follows:

$$b_i(p) = \frac{t \left[\lambda_i f_i^H(p) + (1-\lambda_i) f_i^L(p)\right]}{t\left[\lambda_i f_i^H(p) + (1-\lambda_i) f_i^L(p)\right] + (1-t) f_i^C(p)}. \tag{9}$$

From the preceding equation it follows that a Bayesian agent will view her direct physiological state as being uninformative regarding her assignment category if the control medication is viewed as satisfying the following *ex post blinding condition:*

$$f_i^C = \lambda_i f_i^H + (1-\lambda_i) f_i^L \Rightarrow b_i(p) = t \; \forall \; p \in \mathcal{P}. \tag{10}$$

The preceding two equations imply Bayesian posteriors regarding the efficacy state can expressed

in terms of Bayesian posteriors regarding the assignment category as follows:

$$\begin{aligned}\beta_i(p) &= b_i(p)\left[\frac{\lambda_i f_i^H(p)}{\lambda_i f_i^H(p)+(1-\lambda_i)f_i^L(p)}\right]+[1-b_i(p)]\,\lambda_i \\ &= \lambda_i\left[1+b_i(p)\left(\frac{f_i^H(p)}{\lambda_i f_i^H(p)+(1-\lambda_i)f_i^L(p)}-1\right)\right].\end{aligned} \tag{11}$$

From the second line in the preceding equation it follows that

$$f_i^H(p) \geq f_i^L(p) \Leftrightarrow \beta_i(p) \geq \lambda_i. \tag{12}$$

That is, Bayesian posteriors will exceed priors if the direct physiological state (p) experienced during the RCT is viewed as being more likely in state H than in state L.

3.2 Subjective Posteriors

Intuition suggests a higher realization of the direct physiological state p_i^1 will tend to be associated with a higher posterior probability assessment that the novel medication has high efficacy. This subsection derives a set of technical conditions on subjective priors $\{f_i^C, f_i^H, f_i^L\}$ under which $\widehat{\beta}_i$ is indeed strictly monotone increasing.

To begin, it is useful to note some basic properties of subjective posteriors. We have the following lemma.

Lemma 1 *If $\kappa_i^{\leq} = \kappa_i^{>}$, then $\widehat{\beta}_i$ is continuously differentiable on $(\underline{p}, \overline{p})$. If $\kappa_i^{\leq} \neq \kappa_i^{>}$, then $\widehat{\beta}_i$ is continuously differentiable at $p \in (\underline{p}, \overline{p})$ if and only if $f_i^L(p) \neq f_i^H(p)$.*

Proof. Continuous differentiability of β_i follows from twice continuous differentiability of the cumulative distributions. If $\kappa_i^{\leq} = \kappa_i^{>} \equiv \kappa_i$ continuous differentiability follows from $\widehat{\beta}_i = \kappa_i \beta_i$. If $\kappa_i^{\leq} \neq \kappa_i^{>}$, $\widehat{\beta}_i$ is continuously differentiable if $\beta_i(p) \neq \lambda_i$. From equation (11) this holds if and only if $f_i^L(p) \neq f_i^H(p)$.■

At points where $\widehat{\beta}_i$ is indeed differentiable

$$\beta_i'(p) > 0 \Leftrightarrow \widehat{\beta}'(p) > 0. \tag{13}$$

With this in mind, define the following state-contingent likelihood ratio function:

$$R_i^S(p) \equiv \frac{f_i^S(p)}{f_i^C(p)} \ \forall \ S \in \{L, H\} \text{ and } p \in (\underline{p}, \overline{p}).$$

Differentiating β_i and rearranging terms one finds:

$$\frac{[tR_i^H(p) + (1-t)]'}{tR_i^H(p) + (1-t)} > \frac{[tR_i^L(p) + (1-t)]'}{tR_i^L(p) + (1-t)} \Rightarrow \beta_i'(p) > 0. \tag{14}$$

The preceding equation will be used repeatedly below to establish sufficient conditions for monotone posteriors.

4 Bias: Role of Control Medication

As stated in the introduction, existing work (e.g. Malani (2006)) has focused on the role played by the treatment probability parameter (t) in determining the magnitude of placebo effects. This issue is analyzed in the next section. In this section, we consider the distinct issue of the role played by the statistical characteristics of the control. The first subsection signs bias arising from mental effects. The second subsection relaxes some of the technical assumptions relied upon in the first subsection. The final subsection formulates an empirical test for the control medication effect implied by our model.

4.1 Signing Bias

Consider first a conservative regulator whose objective is to avoid upward bias. We have the following result.

Proposition 2 *Suppose objective probabilities are such that $F^C = F^H$, with F^H first-order stochastically dominating F^L. Suppose further subjective probabilities for all $i \in \mathcal{I}$ are such that $F_i^C = F_i^H$, with f_i^H / f_i^L strictly increasing. Then the expected treatment-control health quality difference is equal to (less than) the expectation of the direct physiological effect in efficacy state H (L).*

Proof. Zero bias in state H follows from equation (7). From equation (14) it follows each $\beta_i' > 0$ implying each $\widehat{\beta}_i$ is strictly increasing. Strictly increasing posteriors imply each $\Psi_i\left[X_i(\cdot)\right] \equiv M_i(\cdot)$ is strictly increasing. Integrating by parts, we have

$$\int_{\mathcal{P}} M_i(p)[f^S(p) - f^C(p)]dp = \int_{\mathcal{P}} M_i'(p)[F^C(p) - F^S(p)]dp. \tag{15}$$

Negative bias in state L follows from F^C first-order stochastically dominating F^L.■

It is worth stressing the preceding proposition does not assume common knowledge. In particular, the proposition places no direct limit on the statistical distance between subjective and objective probability densities. Rather, the proposition assumes a degree of concordance between subjective and objective probabilities. Conversely, if f^H/f^L were to have the monotone likelihood ratio property (MLRP below), then the stated conclusions would hold under common knowledge of the objective probabilities.

The preceding result has the potential to provide useful guidance in settings where an existing drug is known to be highly effective, but features, say, extremely high costs of production, so that one is forced to evaluate lower cost alternatives, say generic drugs. From the proposition it follows that using the existing expensive medication as the control has an attractive feature from the perspective of a conservative regulator: the use of a control mimicking F^H precludes upward bias.

It is also worth stressing the elimination of bias in state H is predicated upon the *objective* probabilities having a particular property, $F^C = F^H$, rather than being predicated upon properties of the subjective priors. Intuitively, equating hope-based mental effects across treatment and control groups in a given state is achieved by having the two groups make draws of direct health effects from identical distributions in that state.

Another point worthy of note is that the preceding proposition could be utilized by a regulator with limited information. For example, suppose the regulator knows the stated conditions are satisfied but does not know $\{f^C, f^H, f^L\}$. Here, the observation of a zero treatment-control difference would reveal that the novel treatment has the same direct health effect as the high-efficacy control, and a negative difference would reveal the novel treatment to be inferior to the control. Phrased

differently, here a zero difference would reveal state H and a negative difference would reveal state L.

In many settings, no existing medication achieves the high efficacy standard, so that the preceding proposition cannot be relied upon. The following proposition illustrates that, nevertheless, the use of a relatively effective control, say an active pill rather than an inert pill, can serve the dual purpose of reducing the probability of upward bias, as well as helping to reveal the efficacy state.

Proposition 3 *Suppose objective probabilities are such that F^H first-order stochastically dominates F^C which first-order stochastically dominates F^L. Suppose further subjective probabilities for all $i \in \mathcal{I}$ are such that f_i^H/f_i^C and f_i^C/f_i^L are strictly increasing. Then the expected treatment-control health quality difference is greater (less) than the expectation of the direct physiological effect in efficacy state H (L).*

Proof. From equation (14) it follows that each $\beta_i' > 0$ implying each $\widehat{\beta}_i$ is strictly increasing. Strictly increasing posteriors imply each $\Psi_i[X_i(\cdot)] \equiv M_i(\cdot)$ is strictly increasing. The bias signs follow from equation (15) and the assumed FOSD properties of objective probabilities.■

The preceding proposition does not assume common knowledge and does not place a direct limit on the statistical distance between subjective and objective probability densities. However, the proposition does assume a degree of concordance between subjective and objective probabilities. Conversely, if f^H/f^C and f^C/f^L were to have the MLRP property, the stated conclusions would hold under common knowledge of the objective probabilities.

The preceding proposition could also be utilized by a regulator with limited information. For example, suppose the regulator knows the stated assumptions are satisfied but is otherwise ignorant regarding $\{f^C, f^H, f^L\}$. Under the stated conditions, the observation of a positive (negative) treatment-control difference would be sufficient to conclude $S = H$ $(S = L)$ implying $\mu^S > \mu^C$ $(\mu^S < \mu^C)$. That is, the sign of the treatment-control difference would here serve as a reliable indicator in ranking medications according to their mean health outcomes. However, a couple of caveats are in order. First, as we show in the next subsection, if the FOSD assumption is violated,

the reliability of the treatment-control difference as a ranking criterion breaks down. Second, even when the preceding proposition's assumptions are satisfied, the magnitude of the incremental health impact would still remain unknown. After all, here a positive (negative) treatment-control difference overstates the direct health benefit (loss).

The preceding proposition allows us to demonstrate a useful result en passant. Consider that the traditional story for placebo effects is that test subjects manifest better outcomes if they think they are receiving the real treatment. In such case, mental effects would be equated across treatment and control groups if the control could be fine-tuned so that the ex post blinding condition (10) is satisfied. The following result shows ex post blinding is no cure-all.

Remark 1 *Ex post blinding of test subjects regarding their treatment status is not sufficient to eliminate mental effect biases.*

Proof. Assume Bayesian subjects holding common knowledge that $\mathcal{P} = [0,1]$, $f^H = 2p$, $f^L = 2(1-p)$, $f^C = 1$ and $\lambda = 1/2$. Then $\beta(p) = t$ for all $p \in \mathcal{P}$. The assumptions of Proposition 3 are satisfied implying bias in both states.■

Consider next an unscrupulous pharmaceutical manufacturer hoping to achieve upward bias. Symmetry suggests the manufacturer will want to choose an ineffective control. Indeed, consistent with this intuition, the next proposition presents technical conditions under which there will be upward bias in both states.

Proposition 4 *Suppose objective probabilities are such that both F^H and F^L first-order stochastically dominate F^C. Suppose further that subjective probabilities are such that f_i^H/f_i^C is strictly increasing and*

$$\left(\frac{f_i^H}{f_i^C}\right)'(p) > \left(\frac{f_i^L}{f_i^C}\right)'(p) > 0 \text{ for all } p \in (\underline{p}, \overline{p}).$$

Then the expected treatment-control health quality difference is greater than the expectation of the direct physiological effect in efficacy state L and efficacy state H.

Proof. Given equation (15) and the FOSD assumptions, to establish the claim it is sufficient to establish each $\beta_i' > 0$ implying each $\widehat{\beta}_i$ is strictly increasing implying each $\Psi_i[X_i(\cdot)] \equiv M_i(\cdot)$ is strictly increasing. To this end, we suppress the identifier i and rewrite equation (14) as follows:

$$\begin{aligned} 0 &< R_H'(p)\left[tR_L(p)+(1-t)\right] - R_L'(p)\left[tR_H(p)+(1-t)\right] \\ &\Leftrightarrow \; 0 < t\left[R_L(p)R_H'(p) - R_L'(p)R_H(p)\right] + (1-t)\left[R_H'(p) - R_L'(p)\right] \\ &\Leftrightarrow \; 0 < t[R_L(p)]^2\left[\frac{f_H(p)}{f_L(p)}\right]' + (1-t)\left[R_H'(p) - R_L'(p)\right]. \blacksquare \end{aligned}$$

In addition to the problem of upward bias, comparison of the two preceding propositions reveals another weakness associated with utilizing low efficacy control medications rather than control medications of intermediate efficacy, potential inability to infer the true efficacy state. We have the following remark.

Remark 2 *If the conditions of Proposition 3 are satisfied, the expected treatment-control health quality difference is positive in efficacy state H and negative in efficacy state L. If instead the conditions of Proposition 4 are satisfied, the expected treatment-control health quality difference is positive in both efficacy states.*

Notice, if the conditions of Proposition 4 are satisfied, the experimenter cannot rely upon the sign of the treatment-control difference to distinguish between the two efficacy states. In this situation, the experimenter would need to rely upon magnitudes of the treatment-control difference to determine the state. However, interpreting treatment-control magnitudes is more difficult since magnitudes depend upon unobservables such as the respective mental effect functions (Ψ_i) as illustrated by equation (7).

To better illustrate the role of the control, consider running two separate RCTs where the same novel drug is given to the respective RCT treatment groups, but with the two control groups receiving different medications C_1 and C_2. Assume the two controls are described using identical wording in informed consent forms and that subjective priors are thus equalized across the two RCTs, implying

identical posterior belief functions.[13] Next, calculate for each of the RCTs the difference between the conditional expectation of the treatment-control difference in state H and the corresponding conditional expectation in state L. Applying equation (7), and noting that equality of prior beliefs implies equality of the posterior belief functions $(\widehat{\beta}_i)$, it is readily verified that:

$$\{[\mathbb{E}(Q_i^1|i \in \mathcal{T}, S=H) - \mathbb{E}(Q_i^1|i \in \mathcal{C}_1, S=H)] - [\mathbb{E}(Q_i^1|i \in \mathcal{T}, S=L) - \mathbb{E}(Q_i^1|i \in \mathcal{C}_1, S=L)]\} =$$
$$\{[\mathbb{E}(Q_i^1|i \in \mathcal{T}, S=H) - \mathbb{E}(Q_i^1|i \in \mathcal{C}_2, S=H)] - [\mathbb{E}(Q_i^1|i \in \mathcal{T}, S=L) - \mathbb{E}(Q_i^1|i \in \mathcal{C}_2, S=L)]\}.$$

The preceding equation states that if the control is varied across two RCTs, with subjective priors being held constant, the cross-state spread in the treatment-control difference will be equal across the two RCTs. For example, suppose both F^L and F^H first-order stochastically dominate F^{C_1} whereas F^{C_2} dominates F^L but is dominated by F^H. Then in the RCT using C_1 as a control one might observe a treatment-control difference of 10 in state H and a difference of 4 in state L. Under the more effective control C_2 one might see a treatment-control difference of 4 in state H and -2 in state L. Notice, the cross-state difference between treatment-control differences is identical under the two controls $(10 - 4 = 4 - (-2))$. However, with the more effective control C_2, the treatment-control difference is shifted downward so that the sign of the difference suffices to infer the state.

4.2 Relaxing Technical Assumptions

This subsection relaxes some of the technical assumptions utilized in the preceding subsection. To begin, recall Propositions 2, 3, and 4 placed no direct limit on the statistical distance between subjective and objective probability densities. However, each proposition assumes a degree of concordance between subjective and objective probabilities. Although concordance is arguably a reasonable working assumption in some settings, the following remark shows that concordance is not essential.

Remark 3 *The respective conclusions of Propositions 2, 3, and 4 hold if the respective assumptions*

[13] Informed consent laws are sufficiently broad so that controls need only be described in broad terms. See Hernandez et al (2014) and Golomb et al (2014).

imposed on subjective probability assessments instead satisfy the assumptions imposed in one of the two remaining propositions.

Proof. The respective restrictions on subjective probability assessments are sufficient to establish each $\beta_i' > 0$ implying each $\widehat{\beta}_i$ is strictly increasing implying each $\Psi_i[X_i(\cdot)] \equiv M_i(\cdot)$ is strictly increasing. With monotonicity established, the final step in the proof of each proposition remains the same.■

In summary, the preceding remark shows that in order to prove Propositions 2, 3, and 4, one could combine the respective assumptions for objective probabilities with either of the remaining two proposition's assumptions regarding subjective probability assessments. In this sense, the results of the propositions remain valid absent concordance. For example, the conclusion of Proposition 4, which assumed the control was first-order stochastically dominated in both states, would remain valid even if the control was instead perceived by all subjects as first-order stochastically dominating, with $F_i^C = F_i^H$ and f_i^H/f_i^L strictly increasing as assumed in Proposition 2.

The preceding remark illustrates how the restrictions imposed on subjective priors can be relaxed with the conclusions of Propositions 2, 3, and 4 still remaining valid. However, in some instances a regulator will want to avoid assuming that prior beliefs satisfy the type of MLRP conditions that were utilized in establishing monotonicity of subjective posteriors. The next remark shows that, even without imposing MLRP, the regulator can determine the true efficacy state if there exists a control that mimics either $S \in \{L, H\}$.

Remark 4 *If objective probabilities are such that $F^C = F^{S'}$ for $S' \in \{L, H\}$, the observation of a treatment-control difference different from zero reveals S' is not the true state.*

Proof. From equation (7) it follows that if $F^C = F^{S'}$ the treatment-control difference is zero if $S = S'$ and thus a non-zero difference reveals the other state.■

The more general implication of the preceding remark is that, even if one prefers to remain agnostic regarding the subjective priors held by test subjects, valuable information can be gathered

nevertheless by relying upon a battery of controls mimicking potential efficacy states. After all, a non-zero treatment-control difference in an RCT utilizing a control mimicking efficacy state S' reveals the true state is not S'. Moreover, the proposition shows that it is not necessary to find controls mimicking every efficacy state. Rather, controls mimicking a proper subset of efficacy states can be sufficient.

Consider next a regulator who, as immediately above, prefers to remain agnostic about the subjective probability assessments held by test subjects, but whose objective is to eliminate bias *on average*. The unconditional expectation of the state-contingent bias in equation (7) is

$$\mathbb{E}[m_i|i \in \mathcal{T}] - \mathbb{E}[m_i|i \in \mathcal{C}] = \int_{\mathcal{I}} \left[\begin{array}{l} \int_{\mathcal{P}} \Psi_i \left[\widehat{\beta}_i(p) \overline{Q}_i^{2H} + (1 - \widehat{\beta}_i(p)) \overline{Q}_i^{2L} \right] \\ [\lambda f^H(p) + (1-\lambda) f^L(p) - f^C(p)] dp \end{array} \right] di. \tag{16}$$

We then have the following remark.

Remark 5 *The unconditional expectation of bias is zero if the objective probability densities satisfy the ex post blinding condition*

$$f^C = \lambda f^H + (1-\lambda) f^L.$$

It is worth stressing that the condition for achieving zero bias in expectation only concerns properties of the *objective* probability densities. Intuitively, if the objective probabilities satisfy the stated condition, the treatment and control groups have an equal unconditional probability of each possible realization of p_i^1. This implies treatment and control groups have equal distributions of posteriors, equal distributions of mental effects, and a fortiori, equal expected mental effects. It is also worth stressing that the preceding remark only speaks to the unconditional expectation of bias. That is, while a control achieving satisfying the preceding ex post blinding condition achieves zero bias on average across states, there may well be large bias in both states (see Proposition 3).

Having discussed potential relaxation of the prior subsection's assumptions regarding the subjective probability assessments held by agents, we turn our attention next to relaxing the FOSD assumptions imposed on objective probabilities. In particular, we are interested in determining

whether bias can still emerge in state S' even if $F^{S'}$ does not first-order stochastically dominate F^C, and vice-versa. In fact, it is readily verified that FOSD relationships are not necessary to generate bias. For example, consider homogeneous Bayesian subjects having common knowledge that: $\mu^D = 0$; $\delta^L = \delta^H = 0$; $\mathcal{P} = [0,1]$, $f^H = 2p$; $f^L = 1$; $f^C = 4p$ for $p \leq 1/2$; and $f^C = 4(1-p)$ for $p > 1/2$, implying $\mu^L = \mu^C$. Suppose also

$$\begin{aligned} \Psi(X) &\equiv \max\{X - X^*, 0\} \\ X^* &\equiv \frac{\pi^L}{2} + \left[\frac{\lambda + \frac{\lambda t}{2}}{1 + \frac{\lambda t}{2}}\right]\left[\frac{2\pi^H}{3} - \frac{\pi^L}{2}\right]. \end{aligned} \tag{17}$$

In the preceding equation, X^* is expected future health quality conditional upon $p_i^1 = 3/4$. Since $\Psi(p) = 0$ for $p \leq 3/4$ and $f^L(p) > f^C(p)$ for $p > 3/4$ it follows that, if β is indeed strictly increasing, as shown below, then

$$\begin{aligned} \mathbb{E}[Q_i^1|i \quad &\in \quad \mathcal{T}, S = L] - \mathbb{E}[Q_i^1|i \in \mathcal{C}, S = L] \\ &= \int_{\mathcal{P}} \Psi\left[\beta(p)\overline{Q}^{2H} + (1 - \beta(p))\overline{Q}_i^{2L}\right][f^L(p) - f^C(p)]dp > 0. \end{aligned} \tag{18}$$

That is, despite the fact that F^L does not stochastically dominate F^C, here there is positive bias in state L. Intuitively, the likelihood of very good direct physiological effects is, by construction, higher for the treatment medication than the control. Under the type of convex brain-modulated physiological effect assumed in equation (17), the treatment medication has a stronger mental effect than the control in state L.

The preceding example illustrates a result of greater practical importance. Recall, despite the existence of bias under the stated assumptions in Propositions 3 and 4, the observation of a positive (negative) treatment-control difference would nevertheless be sufficient to conclude that the novel treatment generates a higher (lower) expected direct health outcome than the control. These results appear to lend support to the view of the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH, 2000) that the sign of the treatment-control difference offers a valid metric for rank-ordering medications. However, the following proposition shows that such a rule-of-thumb is not valid in general.

Proposition 5 *The observation of a positive treatment-control difference does not imply $\mu^S > \mu^C$ and the observation of a negative treatment-control difference does not imply $\mu^S < \mu^C$.*

Proof. Given the discussion preceding equation (18), the first claim follows if β is strictly increasing. Let

$$\begin{aligned}\gamma(p) &\equiv \frac{tR^H(p)+(1-t)}{tR^L(p)+(1-t)} \ \forall\ p \in (\underline{p},\overline{p}) \\ &\Rightarrow \ln[\gamma(p)] = \ln[tR^H(p)+(1-t)] - \ln[tR^L(p)+(1-t)] \\ &\Rightarrow \frac{\gamma'(p)}{\gamma(p)} = \frac{[tR^H(p)+(1-t)]'}{[tR^H(p)+(1-t)]} - \frac{[tR^L(p)+(1-t)]'}{[tR^L(p)+(1-t)]}.\end{aligned}$$

From equation (14) it follows that β is strictly increasing if γ is strictly increasing. Note

$$\begin{aligned}p &\in \left[0,\frac{1}{2}\right] \Rightarrow \gamma(p) = \frac{1-\frac{t}{2}}{\frac{t}{4p}+(1-t)} \\ p &\in \left[0,\frac{1}{2}\right] \Rightarrow \gamma(p) = \frac{2tp+4(1-t)(1-p)}{t+4\,(1-t)\,(1-p)}.\end{aligned}$$

Differentiating, it follows γ is increasing on both intervals.

For the second claim, consider the same density functions but assume Ψ has a convex kink and a concave kink benefiting the control, which has thinner tails. In particular, assume

$$\Psi[X(p)] \equiv \min\left\{\sqrt{\frac{1}{2}}, \sqrt{\max\left\{p-\frac{1}{4},0\right\}}\right\}. \tag{19}$$

It follows

$$\int_0^1 \Psi[X(p)][f^C(p)-f^L(p)]dp = \frac{29}{80}\sqrt{2} - \frac{1}{15} - \frac{7}{24}\sqrt{2} > 0.\blacksquare$$

Recall, the preceding subsection established that, although RCTs might well be biased, under the assumptions in Propositions 3 and 4, the observation of a positive (negative) treatment-control difference is sufficient to conclude $\mu^S > \mu^C$ ($\mu^S < \mu^C$). The two key ingredients in proving those propositions were FOSD assumptions on objective probabilities and technical restrictions on priors causing posteriors to be monotone. The examples in the preceding proposition maintained monotone posteriors but failed to impose FOSD between F^C and F^L. In the absence of a FOSD relationship,

the relative magnitude of mental effects becomes sensitive to the functional form of Ψ. For example, in the proof of the proposition, $\mu^S = \mu^C$, but the treatment-control difference was positive under a convex Ψ (equation (17)) that advantaged the treatment given its fatter right tail. Conversely, the treatment-control difference was negative given a convex-concave Ψ (equation (19)) that advantaged the control given its thin tails.

Phrased differently, even the "gold standard" treatment-control difference implicitly relies upon technical assumptions about probability distributions or functional forms. In particular, stronger distributional assumptions, e.g. FOSD, allow a greater degree of agnosticism regarding functional forms of mental effects.

4.3 Tests for RCT Bias

It would be useful to formulate statistical tests that would allow researchers to assess whether RCTs correctly measure incremental health improvements relative to the chosen control, $\mu^S - \mu^C$, or whether, as we have argued, RCTs generate biased measures of incremental health improvements with the bias itself being systematically related to the objective properties of the control medication.

For our first bias test, we propose adopting the differences-in-differences (DiD below) test statistic, commonly employed in applied microeconometric work, but tailoring it to medical RCTs in order to isolate the role played by the control medication.[14] To motivate this test, consider two different RCTs where the same novel drug is given to the treatment groups, but with the two control groups receiving distinct medications C_1 and C_2 generating mean health outcomes μ_1^C and μ_2^C, with C_2 first-order stochastically dominating C_1. Suppose C_1 and C_2 are described using identical wording in informed consent forms and thus subjects in the two RCTs form the same subjective priors. This implies identical posterior belief functions. Finally, to conserve notation, assume test subjects are homogeneous.[15]

Consider then computing the expectation of the difference between the treatment-control differ-

[14] See Angrist and Pischke (2009) for a general discussion of DiD.

[15] This is without loss of generality.

ences across the two RCTs. Applying equation (7), we have:

$$\begin{aligned} & \left\{\mathbb{E}[Q_i^1|i \in \mathcal{T}, S] - \mathbb{E}[Q_i^1|i \in \mathcal{C}_1, S]\right\} - \left\{\mathbb{E}[Q_i^1|i \in \mathcal{T}, S] - \mathbb{E}[Q_i^1|i \in \mathcal{C}_2, S]\right\} \qquad (20) \\ = \; & \mu_2^C - \mu_1^C \\ & + \Delta\overline{Q}^2 \int_{\mathcal{P}} \Psi' \left[\widehat{\beta}(p)\overline{Q}^{2H} + (1 - \widehat{\beta}(p))\overline{Q}^{2L}\right] \widehat{\beta}'(p)[F_1^C(p) - F_2^C(p)]dp. \end{aligned}$$

Notice, if the two RCTs were delivering unbiased estimates of the novel treatment's incremental health benefit, the integral in the preceding equation would be zero and the DiD would simply reflect the difference between mean health outcomes under the different control medications. In contrast, our theory predicts that, provided posterior beliefs are monotone, the integral term here will be positive and the DiD will exceed $\mu_2 - \mu_1$. Intuitively, we predict greater systematic pessimism of the control group receiving C_1 rather than C_2, which increases the treatment-control difference in the C_1 controlled trial and increases the DiD.

The assumption of common prior beliefs across the RCTs is not necessary. To see this, consider the same thought-experiment, but now admit the possibility that prior beliefs differ across the two RCTs, implying that the posterior belief functions will differ across the two RCTs. Here we have

$$\begin{aligned} DiD \; = \; & \mu_2 - \mu_1 \qquad (21) \\ & + \Delta\overline{Q}^2 \int_{\mathcal{P}} \Psi' \left[\widehat{\beta}_1(p)\overline{Q}^{2H} + (1 - \widehat{\beta}_1(p))\overline{Q}^{2L}\right] \widehat{\beta}_1'(p)[F_1^C(p) - F^S(p)]dp \\ & - \Delta\overline{Q}^2 \int_{\mathcal{P}} \Psi' \left[\widehat{\beta}_2(p)\overline{Q}^{2H} + (1 - \widehat{\beta}_2(p))\overline{Q}^{2L}\right] \widehat{\beta}_2'(p)[F_2^C(p) - F^S(p)]dp. \end{aligned}$$

Suppose F_1^C is first-order stochastically dominated by F^S, whereas F^S is first-order stochastically dominated by (or equal to) F_2^C. Here too the DiD exceeds the difference between mean health outcomes under the respective controls.

Phrased differently, according to conventional theories of placebo effects, pharmaceutical manufacturers should be indifferent between the choice of control, with the treatment-control difference being reduced one-for-one with changes in μ^C, which implies that the experimenter must simply add the treatment-control difference back to μ^C in order to deduce the mean health outcome under

the novel medication. In sharp contrast, our analysis implies that the treatment-control difference changes *more* than one-for-one with changes in μ^C. This implies that a pharmaceutical manufacturer hoping to generate upward bias in estimated health outcomes would prefer to use a low efficacy control since this generates systematically more pessimistic control group beliefs and hence greater upward bias in the assessed treatment-control difference.

A limitation of the proposed DiD test, as a device for detecting RCT bias, is that it presumes the researcher knows the true difference between mean health outcomes induced by the two control medications. If this is not the case, the following triangle equality test can be applied instead.

The proposed triangle equality test requires the researcher to conduct three different RCTs. The first two RCTs are as described above for the DiD test: the novel drug will be given to the respective RCT treatment groups, with the respective RCT control groups receiving different control medications C_1 and C_2 generating (unknown) mean health outcomes μ_1^C and μ_2^C. Again, C_1 and C_2 are to be described using identical wording in informed consent forms so that subjects in these two RCTs form the same subjective priors implying identical posterior belief functions. In the third RCT, subjects are randomly assigned to trial arms given either C_1 or C_2, and are informed of this.

Next consider that

$$\mu^S - \mu_1^C = (\mu^S - \mu_2^C) + (\mu_2^C - \mu_1^C). \tag{22}$$

From the preceding equation it follows that, if RCTs are in fact delivering unbiased estimates of pure physical effect differences, the following triangle equality must hold:

$$\begin{aligned} \mathbb{E}[Q_i^1|i \quad &\in \quad \mathcal{T}, S] - \mathbb{E}[Q_i^1|i \in \mathcal{C}_1, S] \\ &= \quad \{\mathbb{E}[Q_i^1|i \in \mathcal{T}, S] - \mathbb{E}[Q_i^1|i \in \mathcal{C}_2, S]\} + \{\mathbb{E}[Q_i^1|i \in \mathcal{C}_2, S] - \mathbb{E}[Q_i^1|i \in \mathcal{C}_1, S]\} \end{aligned} \tag{23}$$

That is, if RCTs are unbiased, the difference between the novel treatment and control 1 should be equal to the difference between the novel treatment and control 2 plus the difference between control 2 and control 1.

In contrast, according to our proposed theory of placebo effects, only the third RCT, which compares the two controls, can be presumed to yield an unbiased estimate of health effect differences, since subject experiences in that RCT are uninformative about the efficacy state of the novel drug. Applying integration by parts in equation (7), our theory implies the triangle equality will be satisfied only if

$$\begin{aligned} & \mu^S - \mu_1^C + \Delta\overline{Q}^2 \left[\begin{array}{c} \int\limits_{\mathcal{P}} \Psi' \left[\widehat{\beta}(p)\overline{Q}^{2H} + (1-\widehat{\beta}(p))\overline{Q}^{2L} \right] \widehat{\beta}'(p) \\ [F_1^C(p) - F^S(p)]dp \end{array} \right] \\ = \; & \mu^S - \mu_2^C + \Delta\overline{Q}^2 \left[\begin{array}{c} \int\limits_{\mathcal{P}} \Psi' \left[\widehat{\beta}(p)\overline{Q}^{2H} + (1-\widehat{\beta}(p))\overline{Q}^{2L} \right] \widehat{\beta}'(p) \\ [F_2^C(p) - F^S(p)]dp \end{array} \right] \\ & + \mu_2^C - \mu_1^C. \end{aligned} \tag{24}$$

In turn, satisfaction of the preceding condition demands that, in violation of our theory, mental effects are invariant to the objective properties of the control medications, so that

$$\begin{aligned} & \int\limits_{\mathcal{P}} \Psi' \left[\widehat{\beta}(p)\overline{Q}^{2H} + (1-\widehat{\beta}(p))\overline{Q}^{2L} \right] \widehat{\beta}'(p) F_1^C(p) dp \\ = \; & \int\limits_{\mathcal{P}} \Psi' \left[\widehat{\beta}(p)\overline{Q}^{2H} + (1-\widehat{\beta}(p))\overline{Q}^{2L} \right] \widehat{\beta}'(p) F_2^C(p) dp. \end{aligned} \tag{25}$$

5 Treatment Group Measure

This section examines the role played by the treatment group measure. We begin first with an analysis of how bias varies with t. We then move on to deriving testable implications regarding how health outcomes within treatment and control arms respectively will vary with t.

To begin we have the following result.

Proposition 6 *As the measure of the treatment group goes to zero, bias goes to zero, regardless of the true efficacy state S.*

Proof. As t tends to 0, $\widehat{\beta}_i$ tends to λ_i, and $\widehat{\beta}'_i$ tends to 0. This implies the bias term in equation (7) tends to 0.■

Intuitively, bias arises from systematic differences in posterior beliefs across treatment and control groups. In turn, such systematic differences arise naturally if treatment and control groups draw their direct physiological states from different distributions. However, if t is close to zero, subjects rationally place little weight on their own draw of p_i^1 in forming posterior beliefs.

Importantly, the proposition shows that RCTs featuring very small treatment probabilities have the attractive property of virtually eliminating bias in *both* efficacy states. By way of contrast, we recall from the preceding sections that clever changes in the control medication cannot achieve this. For example, mimicking controls ($f^C = f^S$) only eliminate bias in one of the two efficacy states, and controls mimicking the unconditional treatment density, $f^C = \lambda f^H + (1-\lambda) f^L$, only reduce the unconditional expectation of bias to zero.

However, as a practical matter there are a number of reasons one may want to avoid test designs featuring small treatment groups. First, there is the concern over standard errors. Second, there may be ethical concerns in withholding treatment for large numbers. Finally, such designs may fail to attract voluntary participation, even if monetary incentives were provided to encourage participation.

With this in mind, a natural question to ask is whether bias is monotonically increasing in t. To address this question, we now express the bias in state $S \in \{L, H\}$ as a function of the trial design parameter t:

$$B^S(t) \equiv \int_{\mathcal{I}} \left[\underbrace{\int_{\mathcal{P}} M_i(p,t) f^S(p) dp}_{\text{Treatment Arm Mental}} - \underbrace{\int_{\mathcal{P}} M_i(p,t) f^C(p) dp}_{\text{Control Arm Mental}} \right] di. \tag{26}$$

Differentiating the preceding equation we obtain:

$$\frac{dB^S(t)}{dt} = \int_{\mathcal{I}} \left[\int_{\mathcal{P}} \left(\begin{array}{c} \Psi_i' \left[X_i(p,t) \right] \left(\Delta \overline{Q}_i^2 \right) \left(\frac{\partial \widehat{\beta}_i}{\partial \beta_i} \right) \\ \left(\frac{\lambda_i (1-\lambda_i) f_i^C(p)}{[t(\lambda_i f_i^H(p) + (1-\lambda_i) f_i^L(p)) + (1-t) f_i^C(p)]^2} \right) [f_i^H(p) - f_i^L(p)][f^S(p) - f^C(p] \end{array} \right) dp \right] di. \tag{27}$$

Consistent with the empirical evidence provided by Malani (2006), the preceding equation suggests that in general bias will vary with the treatment probability parameter t. In fact, from the preceding expression, we have the following proposition presenting sufficient conditions for the absolute value of bias to be increasing in t.

Proposition 7 *Suppose objective probabilities are such that f^H/f^C and f^C/f^L are strictly increasing. Suppose further subjective probabilities for all $i \in \mathcal{I}$ are such that f_i^H/f_i^C and f_i^C/f_i^L are strictly increasing. Then if the objective and subjective probability densities all cross at a single point, the absolute value of bias in both states is strictly increasing in t.*

Proof. We apply equation (27) focusing on the final product term in the integrand. Let p^* denote the single crossing point. The result claimed for state H follows from:

$$p \neq p^* \Rightarrow [f_i^H(p) - f_i^L(p)][f^H(p) - f^C(p)] > 0.$$

The result claimed for state L follows from:

$$p \neq p^* \Rightarrow [f_i^H(p) - f_i^L(p)][f^L(p) - f^C(p)] < 0.\blacksquare$$

The following proposition presents sufficient conditions for bias to be increasing in t.

Proposition 8 *Suppose objective probabilities are such that both f^H/f^C and f^L/f^C are strictly increasing. Suppose further subjective probabilities satisfy the conditions stated in Proposition 4. Then if the objective and subjective probability densities all cross at a single point, bias in both states is strictly increasing in t.*

Proof. We apply equation (27) focusing on the final product term in the integrand. Let p^* denote the single crossing point. The result claimed follows from:

$$p \neq p^* \Rightarrow [f_i^H(p) - f_i^L(p)][f^S(p) - f^C(p)] > 0.\blacksquare$$

The preceding two propositions notwithstanding, bias is not necessarily increasing in the trial design parameter t. For example, suppose there is positive bias in state H. But suppose the probability densities do not have a single crossing point. In this case, the term $(f_i^H - f_i^L)(f^H - f^C)$ in the integrand in equation (27) is potentially negative on some intervals. By letting the slope of Ψ go to zero for p outside all such intervals one obtains $dB^H/dt < 0$.

Finally, as detailed in the introduction, Malani (2006) presents empirical evidence regarding the arm-by-arm responsiveness of subjects to changes in the treatment probability. We consider now the testable implications of our model in this context. Returning to equation (26), consider now the derivative of the treatment arm mental effect with respect to the treatment probability parameter t:

$$\int_{\mathcal{I}} \left[\int_{\mathcal{P}} \begin{pmatrix} \Psi_i' \left[X_i(p,t) \right] \left(\Delta \overline{Q}_i^2 \right) \left(\frac{\partial \widehat{\beta}_i}{\partial \beta_i} \right) \\ \left(\frac{\lambda_i (1-\lambda_i) f_i^C(p)}{[t(\lambda_i f_i^H(p) + (1-\lambda_i) f_i^L(p)) + (1-t) f_i^C(p)]^2} \right) [f_i^H(p) - f_i^L(p)] f^S(p) \end{pmatrix} dp \right] di. \tag{28}$$

It is apparent from the preceding equation that the sign of the slope is ambiguous, since the sign depends upon the underlying probability densities. In particular, under, say, the MLRP, the term $f_i^H(p) - f_i^L(p)$ in the integrand will be negative for low values of p. It follows that the preceding integral will be negative if the novel treatment has sufficiently low efficacy, with f^S attaching sufficiently high probability to low realizations of p. Conversely, the term $f_i^H(p) - f_i^L(p)$ in the integrand will be positive for high values of p. Thus, the integral will be positive if the novel treatment has high efficacy, with f^S attaching high probability to high realizations of p. Thus, our theory predicts measured health quality in the treatment arm will generally increase (decrease) in t if the treatment medication has sufficiently high (low) efficacy.

Returning to equation (26), consider next the derivative of the control arm mental effect with respect to the treatment probability parameter t:

$$\int_{\mathcal{I}} \left[\int_{\mathcal{P}} \begin{pmatrix} \Psi_i' \left[X_i(p,t) \right] \left(\Delta \overline{Q}_i^2 \right) \left(\frac{\partial \widehat{\beta}_i}{\partial \beta_i} \right) \\ \left(\frac{\lambda_i (1-\lambda_i) f_i^C(p)}{[t(\lambda_i f_i^H(p) + (1-\lambda_i) f_i^L(p)) + (1-t) f_i^C(p)]^2} \right) [f_i^H(p) - f_i^L(p)] f^C(p) \end{pmatrix} dp \right] di. \tag{29}$$

It is apparent from the preceding equation that the sign of this slope is also ambiguous, since the sign depends upon the underlying probability densities. In particular, under, say, the MLRP,

the term $f_i^H(p) - f_i^L(p)$ in the integrand will be negative for low values of p. Thus, the integral will be negative if the control has low efficacy, with f^C attaching high probability to low realizations of p. Conversely, the term $f_i^H(p) - f_i^L(p)$ in the integrand will be positive for high values of p. Thus, the integral will be positive if the control has sufficiently high efficacy, with f^C attaching high probability to high realizations of p. Thus, our theory predicts measured health quality in the control arm will generally increase (decrease) in t if the control medication has sufficiently high (low) efficacy.

6 Concluding Remarks

This paper illustrates a fragility associated with double-blind RCTs, often viewed as the gold standard in medicine for estimating pure non-placebo physiological effects. As we show, when subjects receive interim signals, and when positive expectancy about future health quality leads to better present-day health quality, the expectation of mental effects cannot be presumed equal across treatment and control groups in RCTs, since beliefs regarding efficacy will vary systematically with the objective probability distributions governing direct physiological states. It follows that the difference between mean health outcomes across treatment and control groups is a biased estimator of the mean of the direct (non-placebo) physiological effect.

We do not argue that all RCTs are vulnerable to the problems highlighted within the model, but rather we argue that RCTs will tend to become biased if subjects observe signals and update beliefs prior to measurement of health outcomes. Conversely, if subjects do not receive interim signals, or if uncertainty regarding the data generating process weakens the updating process, then the biases we posit are less of a concern. This latter argument suggests a potential benefit to greater opacity in informed consent forms, since greater opacity is likely to diminish the propensity for belief updating.

Constructively, we used the model as a framework for analyzing the details of RCT implementation beginning first with an analysis of the role played by the control medication. It was shown that choice of control can be used to alter bias. For example, high-efficacy controls dominate low-efficacy

controls if the goal is to be conservative in the sense of eliminating or reducing the probability of upward bias. Controls of intermediate efficacy serve the purpose of reducing the probability of upward bias, and can also reveal the latent efficacy state as the sign of the treatment-control difference varies across states. Controls mimicking the distribution of health outcomes of the treatment medication in a given state eliminate bias in that state, while controls mimicking the average unconditional density of treatment medication health outcomes cause the expectation of bias to go to zero. Finally, we proposed a novel difference in differences test to detect RCT bias and to test whether our posited control medication effect is indeed operative.

We turned next to analysis of the choice of treatment probability. Here a set of technical conditions were offered such that bias increases in the treatment probability. However, it was also shown that bias magnitude can be non-monotone in treatment probabilities. Finally, in terms of normative implications it was shown that the adoption of balanced panels may not be optimal. This is because bias arising from hope-based placebo effects approaches zero as the treatment probability approaches zero. Thus, in large samples, one may prefer unbalanced panels featuring very small treatment groups.

Continuing with our analysis of the role of treatment probability, we assessed the model's ability to explain some stylized facts that are at odds with the canonical theory of placebo effects. Consistent with the stylized facts, our model predicts that within-arm health outcomes will vary with treatment probability, although the predicted effects can be non-monotone. Further, the responsiveness of treatment and control arms to changes in treatment probability are expected to differ, consistent with the notion that RCTs are biased.

Before closing, it is worth discussing why it would be, as a general matter, inappropriate to credit a studied drug with the mental effects measured during an RCT. First and foremost, as quoted in the introduction, regulators have stated that their goal is to strip out mental effects. Second, as our analysis shows, the expectancy of medical subjects is related to their assessment of the probability of approval and production of a drug, captured by the model parameters (π^L, π^H).

In reality, these parameters are likely to vary over time and cross-sectionally with the financial constraints of companies, regulatory stringency, and governmental funding capacity. They do not represent the type of physiological constants of interest to physicians. Third, as shown, expectancy in a current RCT reflects in part the value of the treatment in counter-factual states. Thus, again, expectancy in a current experimental round is not generally representative of long-term expectancy since information sets change over time, implying changes in beliefs and changes in mental effects. Fourth, it was shown that positive mental effects during a current RCT reflect the anticipated value of next-generation drugs, not just the value of the drug being studied. The failure to account for this effect would lead to downward bias in the estimated marginal benefit of the next-generation drug.

Finally, it is likely that the RCT methodology will gain popularity in the context of social experiments. With social experiments, the distinction between direct effects and indirect mental effects is of less obvious importance than in medicine where regulatory approval of new drugs and procedures is often predicated upon demonstrating positive direct effects. Nevertheless, if the goal of social experiments is to guide policy interventions, understanding and empirically measuring the distinct causal roles played by direct and indirect effects would appear to be important.

References

[1] Angrist, Joshua D. and Jorn-Steffen Pischke, 2009, *Mostly Harmless Econometrics: An Empiricist's Companion,* Princeton University Press.

[2] F. J. Anscombe; R. J. Aumann, 1963, A Definition of Subjective Probability, *Annals of Mathematical Statistics* 34.

[3] Bothwell, Laura E., and Scott H. Podolsky, 2016, The Emergence of the Randomized Controlled Trial, *New England Journal of Medicine* 375 (6), 501-503.

[4] Caplin, Andrew and John Leahy, 2001, Psychological Expected Utility Theory and Anticipatory Feelings, *Quarterly Journal of Economics*, 55-79.

[5] Chan, Tat Y. and Barton H. Hamilton, 2006, Learning, Private Information, and the Economic Evaluation of Randomized Experiments, *Journal of Political Economy* 114 (6), 997-1040.

[6] Chassang, Sylvain, Gerard Padro i Miguel, and Erik Snowberg, 2012, Selective trials: A principal-agent approach to randomized controlled experiments, *American Economic Review* (102), 1279-1309.

[7] Chassang, Sylvain, Erik Snowberg, Ben Seymour, and Cayley Bowles, 2015, Accounting for Behavior in Treatment Effects: New Applications for Blind Trials, *PLOS One*, 10(6), e0127227. doi: 10:1371/journal.pone.0127227..

[8] Deaton, Angus, 2010, Instruments, Randomization, and Learning about Development, *Journal of Economic Literature* 48 (2), 424-455.

[9] Di Blasi, Zelda, Elaine Harkness, Edzard Ernst, Amanda Georgiou, and Jos Kleijnen, 2001, Influence of Context Effects on Health Outcomes: A Systematic Review, *The Lancet* (357), 757-762.

[10] Epstein, Larry G., 2006, An Axiomatic Model of Non-Bayesian Updating, *Review of Economic Studies* 73, 413-436.

[11] Epstein, Larry G., Jawwad Noor and Alvaro Sandroni, 2008, Non-Bayesian Updating: A Theoretical Framework, *Theoretical Economics* 3, 193-229.

[12] Fisher, Ronald A., 1935, *The Design of Experiments*. London: Oliver and Boyd.

[13] Golomb, Beatrice, Sabrina Koperski, Murray Enkin, and Jeremy Howick, 2010, What's in Placebos?, *Annals of Medicine*, October.

[14] Haygarth, John, 1801, *Of the Imagination as a Cause and as a Cure of Disorders of the Body: Exemplified by Fictitious Tractors and Epidemical Convulsions*. Bath: Crutwell.

[15] Hernandez, Astrid, Josep Banos, Cristina Llop and Magi Farre, 2014, The Definition of Placebo in the Informed Consent Forms of Clinical Trials, *PLOS ONE*.

[16] International Conference of Harmonization, 2000, Choice of Control Group and Related Issues in Clinical Trials E10, Department of Health and Human Services: Center for Biological Evaluation and Research.

[17] Kaptchuk, Ted J., 1998, Powerful Placebo: The Dark Side of the Randomised Controlled Trial, *The Lancet* 351, 1722-1725.

[18] MacLeod, A.K., J.M. Williams, and D.A. Bekerian, 1991, Worry is Reasonable: The Role of Explanations in Pessimism about Future Personal Events, *Journal of Abnormal Psychology*, 478-486.

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

[20] Malani, Anup, 2008, Patient Enrollment in Medical Trials: Selection Bias in a Randomized Experiment, *Journal of Econometrics*, 341-351.

[21] Philipson, Tomas, and Jeffrey Desimone, 1997, Experiments and Subject Sampling, *Biometrika*, 619-630.

[22] Rothwell P.M., 2005, External Validity of Randomised Controlled Trials: To Whom do the Results of this Trial Apply?, *The Lancet* 365, 82-95.

[23] Rothwell P.M., 2006, Factors That Can Affect the External Validity of Randomised Controlled Trials, *PLOS Clinical Trials*, 1-5.

[24] Shapiro, A.K. and E. Shapiro, 1984, Patient-Provider Relationships and the Placebo Effect, in Matarazzo, Weiss, Herd, Miller and Weiss eds.: *Behavioral Health: A Handbook of Health Enhancement and Disease Prevention*, New York, N.Y. Wiley Interscience, 371-383.

[25] Stewart-Williams, Steve, and John Podd, 2004, The Placebo Effect: Dissolving the Expectancy versus Conditioning Debate, *Psychological Bulletin* 130 (2), 324-340.

[26] Thomas, K.B., 1987, General Practice Consultations: Is There Any Point in Being Positive?, *British Medical Journal* (294), 1200-1202.

[27] Turner, Judith A., Richard A. Deyo, John D. Loeser, Michael von Korff, and Wilbert E. Fordyce, 1994, The Importance of Placebo Effects in Pain Treatment and Research, *Journal of the American Medical Association* 271 (20), 1609-1614.