The Routledge Companion to Philosophy of Medicine

Miriam Solomon, Jeremy R. Simon, Harold Kincaid

Measuring Placebo Effects

Publication details
Jeremy Howick
Published online on: 25 Oct 2016

Accessed on: 08 Aug 2023
13
MEASURING PLACEBO EFFECTS

Jeremy Howick

1. Introduction

The term \textit{placebo} is Latin for “I shall please.” In medical research and practice, a placebo is a treatment that is capable of making people believe it is, or could be, the real treatment when in fact it is not that treatment. A sugar pill that is indistinguishable to a patient from real vitamin C, for example, could be a placebo. Within clinical trials, placebo controls are often used as a standard against which experimental treatments are compared. For example, when evaluating a new drug, investigators can give some patients a placebo and other patients the experimental drug. If the experimental drug outperforms the placebo by a sufficient margin, it is said to be effective. Patients (as well as, if feasible, doctors, outcome assessors, and data analysts) in these trials are often masked, which means they don’t know which treatment is the experimental drug and which is the placebo. Placebo use in clinical practice and clinical trials is both common and controversial (Howick, 2009, Howick et al., 2013a). A great deal of philosophical controversy surrounds what placebos are (Gøtzsche, 1994, Nunn, 2009), as well as whether placebos are ethical (Foddy, 2009, Howick, 2009). There is also a dispute—the one I will concern myself in this chapter—about the correct methodology that should be used to measure placebo response.

Henry Knowles Beecher’s early (1955) study claimed that one-third of the benefit of all treatments was due to placebo effects (Beecher, 1955). This is still the most widely cited paper in the field, and the beliefs that placebos are powerful persist because of it. Yet more recent studies conducted by Asbjorn Hróbjartsson and Peter Gøtzsche in Denmark conclude that placebos do not have important effects at all (Hróbjartsson and Gøtzsche, 2001, Hróbjartsson and Gøtzsche, 2004a, Hróbjartsson and Gøtzsche, 2010). In this chapter I will review the attempts to quantify placebo effects and list the methodological problems with these attempts. To anticipate, I will argue that while Beecher’s 1955 estimate is likely to be an overestimate, the more recent studies underestimate placebo effects. A large body of evidence shows that placebo effects are very powerful for treating some ailments, especially pain.

I will proceed by first describing Beecher’s 1955 study and the problems with it. The main objection to Beecher’s study is that he incorrectly infers from the fact that someone recovers after taking a placebo to the claim that the recovery was due to the placebo (section 2). I will then proceed to describe the more recent studies conducted by Hróbjartsson and Gøtzsche and list the methodological problems with them. The main issue with Hróbjartsson and Gøtzsche’s analysis is that their study includes heterogeneous placebos and conditions, making their average placebo effect estimate compatible with powerful placebo
MEASURING PLACEBO EFFECTS

effects for certain conditions such as pain (section 3). Before concluding (in section 5), I describe the evidence and mechanisms that explain why, despite the methodological difficulties in estimating placebo effects, it is reasonable to believe that placebos are effective for treating pain (section 4).

2. Beecher's Powerful Placebos

Although placebos have been used knowingly and unknowingly by physicians for at least several centuries (Kaptchuk, 1998), Henry Knowles Beecher was the first person we know of who rigorously attempted to quantify placebo effects. Beecher was a Harvard Medical School graduate and a doctor in North Africa, France, and Italy during World War II. Morphine was sometimes in short supply, and he noticed that some soldiers did not require any painkillers. Many of these patients had very serious wounds; for example, they had been shot. Beecher thus came to suspect that their mental states and attitudes, as much as their physical wounds, affected how much pain they experienced.

After the war was over, he investigated placebo effects by doing one of the first systematic reviews (a study that includes all available studies) with meta-analysis (statistical pooling of results of the studies within the systematic review). Beecher's review contained 15 placebo-controlled studies measuring the effects of painkillers for treating conditions ranging from postoperative pain to angina pectoris, headaches, and cough. The studies contained a combined total of 1,082 patients. He found that one-third of the patients who had received only placebos got better and concludes: “It is evident that placebos have a high degree of therapeutic effectiveness in treating subjective responses, decided improvement . . . being produced in 35.2 ± 2.2% of cases” (Beecher, 1955, p. 1696). Beecher's study is still the most widely cited in the field, and it was immortalized in his famous article “The Powerful Placebo.”

However, in the 1990s, researchers began to formally question Beecher’s results. In a provocatively titled article, “The powerful placebo effect: fact or fiction?” Kienle and Kiene (1997) accused Beecher of failing to “consider that many patients with a mild common cold improve spontaneously” (Kienle and Kiene, 1997, p. 1312). In this regard, Kienle and Kiene are correct: Beecher’s attribution of change in the placebo control group to the placebo control treatment is a fallacy. Many ailments, including the common cold and postoperative pain, usually go away quite quickly without any treatment at all. The so-called natural history of the disease, and spontaneous remission, are all potential causes of apparent recovery that have nothing to do with placebo effects. Claiming that one thing (giving a patient a placebo) causes another thing (reduction in pain) simply because the former comes first is what philosophers call the post hoc ergo propter hoc (after, therefore because of) fallacy.

There are at least two additional problems with Beecher’s method. First, polite patients taking the placebo could report improvement to please investigators, although no benefit was actually felt (Hróbjartsson et al., 2011). Noting that their doctors were trying to help them, these patients could exaggerate how much benefit they felt after being treated simply to please their friendly doctors. In addition, caregiver and doctor attention given to patients could lead to clinical improvement. Many systematic reviews have demonstrated that empathetic and encouraging caregivers improve outcomes in patients suffering from pain, anxiety, and depression (Crow et al., 1999, Di Blasi et al., 2001, Griffin et al., 2004, Derksen et al., 2013, Kelley et al., 2014, Kelm et al., 2014). Hence, the effects observed in the placebo groups within Beecher’s studies could be the effects of caregiver empathy and encouragement rather than the effects of administering, say, a sugar pill. One could, of course, classify practitioner empathy and encouragement as types of placebo effects. At the same time, empathy and encouragement appear to be distinct from the administration of a sugar pill. If the effects of caregiver empathy
and encouragement are different from the effects of administering a sugar pill, then the effects in Beecher’s studies might not be placebo effects.

Given all of the methodological problems with Beecher’s study, Kienle and Kiene thus conclude that “none of the original trials cited by Beecher gave grounds to assume the existence of placebo effects” (Kienle and Kiene, 1997, p. 1316).

3. Hróbjartsson and Gøtzsche’s Powerless Placebos

Taking up the challenge of estimating placebo effects more accurately, Peter Gøtzsche and Asbjorn Hróbjartsson conducted a systematic review and meta-analysis of trials that had three groups of patients:

1. patients not treated at all (often people in these groups were placed on waiting lists)
2. patients given a placebo
3. patients given a “real” treatment

They identified 114 trials. Typical pill placebos were lactose pills, typical physical placebos were procedures performed with the machine turned off (e.g., sham transcutaneous electrical nerve stimulation), and typical psychological placebos were theoretically neutral discussion between participant and dispenser. They then compared what happened to patients who received the placebo with patients in the untreated groups. This method avoided the post hoc ergo propter hoc fallacy because it included a comparison of patients whose ailment would have followed its natural history: the untreated group. They found that on average placebos only have small effects:

[There is] little evidence that placebos in general have powerful clinical effects. Placebos had no significant pooled effect on subjective or objective binary or continuous objective outcomes. We found significant effects of placebo on continuous subjective outcomes and for the treatment of pain but also bias related to larger effects in small trials. The use of placebo outside the aegis of a controlled, properly designed clinical trial cannot be recommended.

(Hróbjartsson and Gøtzsche, 2001, p. 1599)

Even in cases where they found a significant placebo effect—notably for treating pain—Hróbjartsson and Gøtzsche question whether the observed effects were actually placebo effects. They raise a similar concern to Kienle and Kiene, noting that the outcomes in the placebo groups could have been due to a form of reporting bias, such as patients politely reporting they had recovered when in fact they did not actually experience any recovery. In their words:

Patients in an untreated group would know they were not being treated, and patients in a placebo group would think they were being treated. It is difficult to distinguish between reporting bias and a true effect of placebo on subjective outcomes, since a patient may tend to try to please the investigator and report improvement when none has occurred. The fact that placebos had no significant effects on objective continuous outcomes suggests that reporting bias may have been a factor in the trials with subjective outcomes.

(Hróbjartsson and Gøtzsche, 2001, p. 1597)

If we accept their results, Hróbjartsson and Gøtzsche seem to have completely overturned Beecher’s estimate. Yet while Hróbjartsson and Gøtzsche’s method overcomes the post hoc ergo
procrat hoc fallacy, their conclusions are questionable because of several problems with their review. The main problems are: (1) their review included heterogeneous studies that arguably should not have been lumped together so their average results fail to rule out powerful placebo effects for some conditions; (2) they failed to adequately put strictures on what counts as a placebo; and (3) they failed to recognize that untreated groups within clinical trials are not truly untreated. I will go over each of these problems in turn.

3.1. Problem with the Heterogeneity of the Studies within Hróbjartsson and Gøtzsche’s Review

Perhaps the most serious problem with Hróbjartsson and Gøtzsche’s review is heterogeneity. It is often acceptable to meta-analyze (pool) results, but the meta-analysis can be misleading if researchers include heterogeneous studies—apples and oranges—within the same analysis. Over 40 clinical conditions were included in the analysis, ranging from hypertension and compulsive nail biting to fecal soiling and marital discord. The placebo interventions were also diverse. Kirsch (2002), for example, notes that Hróbjartsson and Gøtzsche jumble together (along with placebo pills and injections) relaxation (described as a placebo in some studies and a treatment in others), leisure reading, answering questions about hobbies, newspapers, magazines, eating favorite foods and watching sports teams, talking about daily events, family activities, football, vacation activities, pets, hobbies, books, movies, and television shows as placebos. In this case it is problematic because it means that although they found a very small average placebo effect, some of the placebo treatments within the review had large effects. Just as “real” medicine is not necessarily effective on average for everything, neither are placebos. Instead, like “real” medicine, we might expect placebos to be effective for treating some ailments and not effective for treating others. To illustrate this problem, Howick et al. (2013b) reanalyzed the same three-armed trials from Hróbjartsson and Gøtzsche’s review, but instead of just measuring placebo effects (by comparing outcomes in placebo and no treatment groups), they also measured treatment effects (by comparing outcomes in treatment and placebo groups). They found that average treatment effects in that heterogeneous group of studies were also modest and not statistically significantly different from placebo effects. Unless we accept that medicine does not have significant effects, the Howick et al. study is arguably a reductio ad absurdum of Hróbjartsson and Gøtzsche’s methodology. If it is legitimate to pool heterogeneous placebos for diverse conditions, then it is also legitimate to pool similarly heterogeneous treatments for the same diverse conditions. Since pooling treatments in this way reveals minuscule treatment effects, and we know some treatments are highly effective, then it follows that pooling heterogeneous placebos for diverse conditions is not legitimate.

The highly significant heterogeneity of the interventions studied calls into question what conclusions can be drawn from the meta-analysis. An overall finding of insignificant placebo effects does not count against the reasonable view that placebos, like treatments, are common and powerful for certain disorders, but not for others. Similarly, certain placebos—for example, placebo injections—could be more effective than other placebos—for example, placebo pills. In section 4, I will review the evidence supporting the claim that placebo effects are quite large for treating pain.

3.2. Problem with Hróbjartsson and Gøtzsche’s Definition of Placebos

The estimate of the placebo effect is based on the average difference between outcome in the placebo and no treatment groups. This estimate depends on whether the placebos in the included trials were “real” placebos and whether the untreated groups were truly untreated. Yet
the placebo concept has proven notoriously difficult to define adequately. It is beyond the scope of this chapter to outline the definitional problems—see Howick, 2016 for a review—however, the common definitions of placebos as inactive or nonspecific are clearly mistaken: placebos can be active, at least with respect to some ailments, and their mechanism of action can be as specific as the action of a pharmacological drug (see section 4). To avoid the definitional problem, Hróbjartsson and Gøtzsche decided to adopt a practical approach and characterize placebos “practically as an [any!] intervention labeled as such in the report of a clinical trial.” But it hardly needs remarking that this approach is untenable. Suppose, for example, that someone reported using penicillin as a placebo in a trial of some new antibiotic as a treatment for pneumonia. The response will of course be “no one would, and if they did we would not take the trial seriously.” But this reaction seems exactly to show that we work with some concept that involves judgments about what can and cannot count as appropriate or legitimate placebos and placebo controls. Since it is unclear whether the placebos within Hróbjartsson and Gøtzsche’s study were actual placebos, their estimates of placebo effects must be questioned. Worse, Hróbjartsson and Gøtzsche go back on their alleged policy of accepting any treatment labeled as a placebo in the report of a clinical trial. For example, they exclude studies where “it was very likely that the alleged placebo had a clinical benefit not associated with the ritual alone (e.g., movement techniques for postoperative pain)” (Hróbjartsson and Gøtzsche, 2001, p. 1595). Here they seem to sneak in a definition of placebos as the effects of “rituals,” which is not acceptable: ritual feasting or fasting are not placebos. A similar problem arises because the untreated groups within Hróbjartsson and Gøtzsche’s study may actually have benefited from observation by and contact with health care practitioners.

3.3. Hawthorne Effects in the No Treatment Groups

The Hawthorne effect is named not after the scientist discovering the phenomenon but after a series of experiments in the Hawthorne works of the Western Electric Company in Chicago between 1924 and 1933. In one study, the lighting in the factory remained stable for the control group, whereas the lights in the part of the factory where the experimental group worked were made brighter. Productivity increased equally in both groups. In another version of the experiment, the researchers tried the opposite: lighting was kept stable in the part of the factory where the control group worked but was made less bright where the experimental group worked. In the second experiment, productivity also increased steadily and equally in both groups until the lights were so low in the experimental groups that the workers protested that they couldn’t see and production fell off (Roethlisberger and Dickson, 1939). Since the productivity increased in both groups whether or not the lights were turned up, turned down, or remained the same, it couldn’t be the lighting that caused the increase in productivity. The reason productivity increased was that the workers knew they were being watched. The effect of being watched is called the Hawthorne or observer effect. In the case of medical research, patients in clinical trials could experience some benefit (or change their behavior and responses) because they are being investigated.

Hawthorne effects could have influenced the untreated groups within Hróbjartsson and Gøtzsche’s analysis and confounded their results. “Untreated” participants within clinical trials are typically monitored and observed by researchers (Einarson et al., 2001). Knowing that their condition is being monitored, the patients might do various things that are good for their health, such as exercise more, eat healthier, or drink less. These things could all produce a health benefit for many conditions being treated. In addition, by virtue of the fact that they are being monitored, the “untreated” patients have contact with health care practitioners. Contact with practitioners has been shown to have an effect, especially if the practitioners are
empathetic (Di Blasi et al., 2001). This presents a problem with Hróbjartsson and Gøtzsche’s review that is similar to the objection they level at Beecher. The outcomes in the untreated groups might not be the result of natural history, but instead be the result of Hawthorne effects and the effects of contact with empathetic health care practitioners.

There is good reason to believe that the untreated groups did benefit from Hawthorne effects and the effects of caregiver attention. A recent systematic review found that the untreated groups within Hróbjartsson and Gøtzsche’s review experienced a 24% improvement compared with baseline (Krogsboll et al., 2009). There are two plausible explanations for the 24% improvement from baseline. First, it could be natural history of disease. Second, the improvement is partly due to Hawthorne effects and effects of contact with health care practitioners. If it is the second—and this is my preferred explanation since 24% is a large effect that is not likely to be explainable solely in terms of natural history—then Hróbjartsson and Gøtzsche have underestimated the placebo effect. Because the untreated groups were not, in fact, untreated, Hróbjartsson and Gøtzsche’s review underestimated placebo effects. This is due to simple arithmetic, noting that the placebo effect is estimated by subtracting outcomes in the placebo effect with outcomes in the untreated groups.

3.4. Other Biases within Hróbjartsson and Gøtzsche’s Review

Hróbjartsson and Gøtzsche updated the 2001 meta-analysis in 2004 (Hróbjartsson and Gøtzsche, 2004b, Hróbjartsson and Gøtzsche, 2010). The 2010 review included 202 studies—almost twice as many as the initial review. The more studies they included, the greater the placebo effects became, both in terms of statistical significance and (in the case of pain) clinical relevance. Moreover, Hróbjartsson and Gøtzsche recognized the problem that the untreated participants may have been treated and hence led to an underestimation of placebo effects. In their words: “Patients in a no-treatment group also interact with treatment providers, and the patients are therefore only truly untreated with respect to receiving a placebo intervention” (Hróbjartsson and Gøtzsche, 2004a, p. 97). They also acknowledge the problem of heterogeneity: “we cannot exclude the possibility that in the process of pooling heterogeneous trials the existence of such a group [of trials that showed a significant placebo effect] was obscured” (Hróbjartsson and Gøtzsche, 2004a, p. 97). Despite the growing evidence for placebo effects, and recognition of at least a few of the problems with their analysis, their conclusions remained just as skeptical, concluding that they “did not find that placebo interventions have important clinical effects in general” (Hróbjartsson and Gøtzsche, 2010). Yet given that their updated review found greater placebo effects, one may question whether Hróbjartsson and Gøtzsche were justified in maintaining the same conclusion.

Finally, Hróbjartsson and Gøtzsche’s study may not be applicable to routine clinical practice because placebos are likely to have a smaller effect in clinical trials than in clinical practice. In a clinical trial, patients believe that they might receive the placebo or the experimental treatment. There is therefore some doubt as to whether they will receive the “real” treatment. This could lead to lower expectations about recovery and worse outcomes compared to the case where they believed they were sure to receive the real treatment and thus had higher expectations. If a doctor prescribed a placebo in clinical practice, however—and several studies have shown that most doctors have prescribed a placebo in clinical practice (Fassler, 2009, Howick et al., 2013a)—the patient believes they are receiving a “real” treatment, and hence have higher expectations. The higher expectations, in turn, could lead to greater placebo effects in clinical practice. Since Hróbjartsson and Gøtzsche’s review attempts to measure placebo effects within clinical trials, it may not apply to clinical practice where placebo effects could be greater.
3.5. Summary of the Evidence for Placebo Effects

In sum, Beecher's estimate of placebo effects is problematic for failing to take natural history into account and is likely an overestimate. However, Hróbjartsson and Gøtzsche's studies are likely to underestimate placebo effects, most notably because they fail to consider variation in placebo effectiveness and because they fail to acknowledge Hawthorne effects. The area where there is least controversy about the effectiveness of placebos is pain. In the next section I will review the evidence for both the existence of and the mechanism for placebo analgesia.

4. The Evidence and Mechanisms for Placebo Analgesia

Even Hróbjartsson and Gøtzsche's skeptical review acknowledges—albeit with what I argued were partly unjustified skeptical caveats—that placebo analgesia is effective. A large body of evidence has demonstrated the effects of placebo analgesia. For example, a recent systematic review of 198 trials (involving 16,364 patients) of placebo treatments for osteoarthritis pain found that placebos were effective (Zhang et al., 2008). In one of the more innovative studies that points towards a mechanism for placebo analgesia, Benedetti et al. (2004) used four common painkillers on a total of 278 patients who had undergone thoracic surgery for different pathological conditions (Benedetti et al., 2004, p. 680). The patients were then, of course unbeknownst to them, randomized into “overt” and “covert” groups with sex, age, weight, and pain baseline-balanced. The overt group was treated by doctors who “gave the open drug at the bedside, telling the patient that the injection was a powerful analgesic and that the pain was going to subside in a few minutes” (Benedetti et al., 2004, p. 681). Then, one dose of analgesic was administered every 15 minutes until the patients reported a 50% reduction of pain. The covert group, on the other hand, had the analgesic delivered by a preprogrammed infusion without any doctor or nurse in the room. Both the covertly and overtly treated patients already had an intravenous line and machine attached to them because of their operations. The results were that over 30% more analgesic was required by the patients who were treated covertly. The $P$-values reached statistical significance, ranging from 0.02 to 0.007 depending on the drug.

How come people who expected a positive outcome—those who were treated overtly—felt less pain? The answer lies in the brain's reward mechanism. The expectation of a positive reward (such as decreased pain) excites the body in a way that it produces enough endogenous endorphins to cause a reduction in pain (Benedetti, 2009). Combining the first four letters of the word endogenous with the last letters of the word morphine gives us the term endorphin. From the human cell receptor's point of view, morphine and endorphins are the same thing. To demonstrate that the body's endorphins are responsible for the analgesic effects of placebos, several researchers have done interesting randomized trials where they gave some patients a drug called naloxone, which antagonizes (blocks) the effects of opiates such as morphine and endorphins, and didn’t give it to other patients. Now if placebo analgesia works by activating the body's ability to produce endorphins, then blocking the endorphin activity would reduce the placebo effect. In fact, this is just what happened. In the patients who received naloxone, the placebo effect was diminished (ter Riet et al., 1998, Sauro and Greenberg, 2005).

In short, dozens of empirical studies demonstrate an analgesic effect of placebos, and at least one mechanism explaining how placebo analgesia works—endogenous opiates—has also been established in rigorous studies. The existence of a placebo analgesia effect does not fully exonerate Beecher for his methodological flaws, but it does suggest that, at least...
MEASURING PLACEBO EFFECTS

for killing pain, Hróbjartsson and Gøtzsche are mistaken that placebos have insignificant effects.

5. Conclusion

What can we conclude about the power of placebos from these studies? The first thing is to acknowledge the fact that placebos won’t work for everything. If someone has a leg amputated, taking a sugar pill or believing it will grow back is unlikely to have any effect. In addition, many ailments—especially the most common ones people visit their doctors for, such as mild to moderate pain, depression, anxiety, flu, colds, and so on—will go away on their own and without any treatment (placebo or not). Beecher’s early and widely cited systematic review of placebo effects made the methodological error of failing to take natural history into account, and it led to an overestimation of placebo effects. However, Hróbjartsson and Gøtzsche’s conclusion that placebos do not have much of an effect at all is equally problematic. Just as “real” treatments are exceptionally powerful for treating some conditions (for example, antibiotics for meningitis), so placebo treatments can be effective for treating some conditions, especially mild to moderate pain and depression, anxiety, and smoking cessation, and they can generally improve quality of life. By lumping all types of placebos for all conditions together, Hróbjartsson and Gøtzsche obscure important placebo effects for conditions like pain. For example, the effects of placebo analgesia have been demonstrated empirically, and their mechanism of action—endogenous opiates—has been established in numerous studies. Further focused research is required to clarify the placebo concept and provide accurate estimates of placebo effects for specific conditions.

References


**Further Reading**

Beecher, H. K. 1955. The powerful placebo. *J Am Med Assoc*, 159, 1602–6. This is the original study that provides a placebo effect estimate; it is also one of the earliest examples of a systematic review and meta-analysis.


