Causal diagrams, the placebo effect, and the expectation effect

Eyal Shahar1
Doron J Shahar2
1Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, 2Department of Mathematics, College of Science, University of Arizona, Tucson, AZ, USA

Abstract: Using causal diagrams, a formal research methodology, we analyzed several definitions of placebo and the placebo effect. We conclude that placebo is an ambiguous, redundant term and that the so-called placebo effect conceals far more interesting effects that are attributed to the patient's expectation. Biomedical research will benefit from abandoning the term placebo effect and focusing instead on a deeper understanding of the expectation variable, including its causes, effects, and effect modifiers. This avenue of research should be pursued by observational cohorts that are nested within clinical trials.

Keywords: causal diagrams, effect modification, placebo, expectation

Introduction

Despite decades of interest in the placebo effect,1 2 the meaning of the term remains controversial.3 Many writers allude to the positive effect of a treatment on a favorable outcome – when that treatment has no biological effects; others extend the term to include psychological effects of any treatment. The counterpart of the placebo effect – the effect of a treatment on a negative outcome through a negative expectation – is sometimes called the nocebo effect.4 Both resonate with the idea of a self-fulfilling prophecy, where a belief in some outcome is one of the causes of that outcome. This idea, however, is the only unifying theme in writings about the placebo (or nocebo) effect: expectations can affect the outcome. Beyond that, disagreements prevail.

Even terminology is unclear. Numerous definitions of the word “placebo” have been proposed,5 many of which are incomprehensible. Moreover, several different effects have been called the “placebo effect”, yet some of them do not even refer to placebo treatments.1 3 To compound misery with madness (as the author Christopher Paolini would put it), at least three study designs claim to estimate the placebo effect, but the term does not have the same meaning in all three. Furthermore, each design embeds assumptions that may produce bias.

We used causal diagrams to gain better insight into the placebo effect. Since causal diagrams express causal assertions clearly, they are an excellent tool for scientific reasoning, which proved helpful in understanding study design and sources of bias.5 8

Causal diagrams, effect modification, and conditioning

A causal diagram depicts a theoretical causal structure for a certain set of variables.9 The variables are displayed along the time axis (left to right), and an arrow points from
each cause to its presumed effect. Figure 1 shows an example that includes two treatment variables, the treatment offered ($T_{OFFERED}$) and the treatment taken ($T_{TAKEN}$), and an outcome ($D$). $R$, $Q$, and $C$ are generic variables.

A natural path between two variables is any alternating sequence of variables and arrows that connects the two ends and passes only once through each variable. The arrows themselves may point in any direction. A causal path is any path in which all the arrowheads point in the same direction, namely, any path through which one variable affects the other. For example: $R \rightarrow T_{OFFERED} \rightarrow T_{TAKEN} \rightarrow D$; $T_{OFFERED} \rightarrow T_{TAKEN} \rightarrow D$; and $T_{TAKEN} \rightarrow D$ (Figure 1). A confounding path contains a shared cause of the two ends of the path, such as $T_{TAKEN} \leftarrow C \rightarrow D$ and $T_{OFFERED} \leftarrow Q \rightarrow D$ (Figure 1). The shared cause is called a confounder. Finally, a colliding path contains two successive arrows that point to the same variable (a collider), such as $T_{OFFERED} \rightarrow T_{TAKEN} \leftarrow C \rightarrow D$ (Figure 1).

Mathematical theorems establish a link between a causal diagram and the expected association between any two displayed variables. In particular, the marginal (crude) association between two variables is created by causal paths and confounding paths that run from one to the other. For example, according to Figure 1 the marginal association between $T_{TAKEN}$ and $D$ is the “sum” of three paths: $T_{TAKEN} \rightarrow D$ (causal path); $T_{TAKEN} \leftarrow C \rightarrow D$ (confounding path); and $T_{TAKEN} \leftarrow T_{OFFERED} \leftarrow Q \rightarrow D$ (confounding path).

Central to cause-and-effect relation is the idea of effect modification.\(^{10}\) In the simplest case, two variables at the same time-point modify each other’s effect on some value of their common outcome. For instance, the effect of the treatment on a favorable outcome might vary according to the patient’s expectation, and the effect of the patient’s expectation might vary according to the treatment. Note that effect modification is a reciprocal phenomenon. That is, if one variable modifies the effect of another, the latter variable also modifies the effect of the former.\(^{10}\) Effect modification is displayed in a causal diagram by placing a lower case letter, denoting a value of the modifier, above the arrow of the modified effect, as shown in Figure 1. This intuitive notation indicates that the effect of one variable (eg, $Q \rightarrow T_{OFFERED}$) depends on the value of the other ($R = r$).

Finally, any study of cause and effect requires us to alter the distribution of some other variables through selection, analysis, or both. Such an alteration, formally called conditioning, may take several forms: restricting the sample to some values (or one value) of a categorical variable; excluding missing values; adding a covariate to a regression model; and more. In its extreme form — restricting a variable to one of its values — conditioning dissociates a variable from all its causes and all its effects (because a value is not associated with anything). In a causal diagram conditioning is depicted by drawing a box around a variable; the resultant dissociation is denoted by drawing two lines over every surrounding arrow.

All causal diagrams are imperfect, delivering an incomplete representation of causal reality. The diagrams we present here are no exception. Still, they help to formalize the discussion and may offer a new insight.

**A causal structure for treatment and outcome**

Many variables affect the treatment that a patient is offered in medical practice or within a trial. We denote these variables collectively by the single letter $R$: the effects of

---

**Notes:** $T_{OFFERED}$ denotes the treatment offered; $T_{TAKEN}$ denotes the treatment taken; $D$ denotes the outcome; $R$, $Q$, and $C$ are generic variables; $r$ denotes a value of the modifier $R$; $q$ denotes a value of the modifier $Q$.

**Figure 1** A causal diagram.

**Figure 2** A causal structure, incorporating causes of the offered treatment ($R$), blinding status ($B$), and the patient’s expectation ($E$) of the outcome ($D$). Notes: $T_{OFFERED}$ denotes the treatment offered; $T_{TAKEN}$ denotes the treatment taken; a lower case letter denotes a value of the modifier (eg, $t_{OFFERED}$ is a value of $T_{OFFERED}$).
these variables are denoted collectively by an arrow with a three-pronged tail emanating from \( R \) (Figure 2). Some of the variables that are represented by \( R \), such as the setting of care, also affect blinding status \( (B) \). For example, patients are more likely to know what they are offered in routine practice than in a blinded trial.

The offered treatment \( (T_{\text{OFFERED}}) \) affects the outcome through three possible causal paths (Figure 2): first, that which is offered obviously affects that which will be taken \( (T_{\text{OFFERED}} \rightarrow T_{\text{TAKEN}}) \). Second, whatever is offered affects the patient expectation about the outcome (denoted \( E \)), which in turn affects the outcome \( (T_{\text{OFFERED}} \rightarrow E \rightarrow D) \). Third, the offered treatment affects the outcome through causal paths that involve the physician (summarized in Figure 2 by a direct, curved arrow). For instance, subsequent medical care depends on what the patient has already been offered.

Figure 2 also shows three intuitive theories of effect modification: first, the effect of the offered treatment \( (T_{\text{OFFERED}}) \) on the patient’s expectation of the outcome \( (E) \) varies according to blinding status \( (B) \), and vice versa. For instance, if patients do not know what they are offered \( (B = \text{blinded}) \), their expectations of benefit should be similar for every treatment (a null effect of \( T_{\text{OFFERED}} \) on \( E \)). In contrast, that effect is not expected to be null if the patients know what they are offered. Similarly, the effect of the offered treatment on the outcome via subsequent medical care \( (T_{\text{OFFERED}} \rightarrow D) \) also depends on blinding status \( (B) \). For example, not knowing what their patient is offered, physicians will provide similar subsequent care, but their medical care might differ if they know that the patient is offered placebo rather than a drug. Finally, the treatment taken \( (T_{\text{TAKEN}}) \) and the patient expectation \( (E) \) also modify each other with respect to the outcome. For instance, expectation might have a stronger effect on a favorable outcome when a drug is taken than when a placebo is taken.

### On definitions

Some of the confusion about placebo treatments can be traced to ambiguous or conflicting definitions of the term. But above all, it is rooted in misunderstanding of the term “definition” itself and its purpose. A short explanation follows.

Consider a long, commonly used phrase, such as “the formation of a blood clot inside a blood vessel that obstructs the flow of blood”. Since it would be cumbersome to continually repeat the phrase, the shorthand “thrombosis” may substitute for the entire phrase. To define something is to do just that: to replace a phrase whose meaning is already clear (called definiens) with another, shorter phrase or word (called definiendum). The whole sentence “[definiendum] denotes [definiens]” is called a definition. Placebo, for example, is a definiendum (whose definition is yet to be discussed).

Before a phrase becomes a definiendum, it is a meaningless collection of letters (or it already has a meaning and is being redefined). After becoming a definiendum, the phrase is no more than an alternative name for its definiens. Definienda, just like acronyms, do not serve any deep purpose, but they help to quicken communication. Without them, communication would be as exhilarating as reading the dictionary. A definiens serves even less of a role: it is merely a meaningful phrase, just given an alternative name.

### What is placebo?

Both \( T_{\text{OFFERED}} \) and \( T_{\text{TAKEN}} \) may take many values. To simplify the discussion, and without losing generality, we will assume throughout that both variables take one of three values: drug, placebo, or nothing. But what is placebo in the first place? What is, for example, the meaning of \( T_{\text{OFFERED}} = \text{placebo} \)? The answer is not that simple.

The word “placebo” originated from Latin and meant “to please”. Later, placebo (definiendum) was used in medicine to denote “any medication prescribed more to please the patient than to treat the disease” (definiens). Finally, the word made its way into biomedical science, but the transition was not smooth.

The placebo-like things that medical researchers were studying were not placebos in the old sense of the word. Therefore, the word was redefined to match those research topics in which they were interested. Not all researchers, however, were interested in the same placebo-like things, so different definitions were introduced. Even worse, many definitions were too vague to be considered definitions at all, as the following examples show.

1. A placebo is a treatment that does not objectively affect the outcome.
2. A placebo is an inert/inactive/ineffective treatment.
3. A placebo is a treatment that has a non-specific effect on the outcome.
What exactly is a non-specific effect vis-à-vis a specific effect?

4. A placebo is a treatment that is inert in the usual sense, but not in its effects.

As before, the word inert is unclear, and “inert in the usual sense” is even worse. Is there “inert in an unusual sense”?

5. A placebo is a treatment that on its own has no beneficial effect on the outcome.

At first glance, this definition seems to agree with the standard usage of the word placebo. But the phrase “on its own” is suspect. Does it mean that placebo alone has no effect, whereas placebo and some other drug do have an effect? What does “on its own” mean – exactly?

The vagueness of these definitions and others also becomes evident when we try to translate them to the language of causal diagrams. Since causal diagrams express causal relations unambiguously, and since all definitions of placebo mention an effect, any clear definition of placebo must correspond to a causal structure. The examples above fail that criterion. We cannot depict them in a causal diagram.

The following definition is a little better:

A placebo is a treatment that has no direct physical effect, but may have a psychological effect.

This is perhaps the most widely used definition, but it is still not rigorous enough. Referring to Figure 2, the definition implies that the causal path $T_{\text{taken}} \rightarrow D$ is elaborated to include some “physical” intermediary, say $P$: $T_{\text{taken}} \rightarrow P \rightarrow D$. Then, a treatment $t_0$ is called placebo if $T_{\text{taken}} = t_0$ is a null effect for the contrast between $T_{\text{taken}} = t_0$ and $T_{\text{taken}} = \text{nothing}$. But $P$ is left unspecified, and unknown variables are not acceptable in a causal diagram. Moreover, the body–mind distinction between the so-called physical effects and psychological effects is not sharp, because a psychological effect can operate through biological variables.

As explained later, causal inquiry does not require us to adopt any definition of placebo. For completeness, however, we offer the following definition, which incorporates the idea of effect modification: a treatment is called placebo if the effect of that treatment (versus no treatment) is null when the patient has no particular expectation about the outcome. Figure 3 depicts the definition. Let $E$ be the patient’s expectation about some outcome, $D = d$ (eg, $\text{VITAL STATUS} = \text{alive}$), and let $t_0$ be some value of $T_{\text{taken}}$. The value $t_0$ is called placebo if the effect of $T_{\text{taken}} = t_0$ versus $T_{\text{taken}} = \text{nothing}$ on $D = d$ is null when $E = \text{indifference}$. Note that the effect in question may be null – but is not required to be null – for other values of $E$. That is, placebo treatment may affect the outcome when the patient has some expectation about the outcome. Of course, in such a case the placebo effect is not null and effect modification is present.

The proposed definition of placebo holds for nocebo as well. In fact, the difference between the standard usage of the two terms has to do with the nature of $D$ – “good outcome” or “bad outcome” – not with structural differences in a causal diagram. As such, we do not differentiate between them.

What is the placebo effect?

Unlike placebo, the various definitions of the placebo effect are simple to understand, typically reflecting different cause-and-effect relations (Figure 4). One group of definitions focuses on the effect of a treatment variable ($T_{\text{taken}}$ or $T_{\text{offered}}$), and another group on the effect of the expectation variable ($E$). Which effect deserves to win the title, if any, is not an interesting question, as explained in the section “On definitions”. Although several definitions rely on the word placebo – whose definition varies – this difficulty will be set aside. Regardless, whenever placebo is mentioned, blinding is implied: that is, placebo recipients think that they are getting, or may be getting, a treatment that may help them.

According to Figure 4A, the placebo effect is defined as the effect of the treatment taken (placebo versus nothing) on some value of the outcome, ignoring the patient’s expectation. This definition will fail, however, whenever expectation is an effect modifier (Figure 2), because the effect, as stated, does not exist; its estimation is subject to effect modification bias.

Figure 4B offers a remedy by taking expectation into account: the placebo effect is defined as before, but may vary – depending on the patient expectation of some outcome. A third definition (Figure 4C) substitutes the offered treatment for the treatment taken. Again, expectation at the time of offering might be an effect modifier (not shown).

Figure 4D–F all focus on the effect of expectation: first, ignoring treatment (Figure 4D); second, when the treatment is called placebo (Figure 4E); third, for any possible treatment
Causal diagrams, placebo, and expectation

(Figure 4F). Again, the first definition will fail whenever the expectation effect is modified by the treatment taken (Figure 2). The second definition eliminates that difficulty by requiring the treatment to be placebo, which in turn, relies on a clear definition of placebo.

Similar to the second definition, the third definition takes effect modification into account (Figure 4F). But there is a much bigger difference between the two. The third definition does not refer to placebo at all! There is no need to argue about the definition of placebo, or even about the definition of the placebo effect, because effect estimation may be pursued for every type of treatment, even if a special value called placebo does not exist. The last definition may be called “the expectation effect” rather than “the placebo effect”. Some authors may call it “the response expectancy effect”, which seems unnecessarily long.

Ultimately, there are two interesting causal variables – treatment and expectation – which may reciprocally modify each other’s effect on the outcome (Figure 2). The business of biomedical research is to estimate those effects, and any effect modification between them, and between them and other variables.

More on the expectation effect

The ideas of placebo treatment and effect are inherently confusing and sharply differ from classic ideas about values and effects of variables. As proposed here, these difficulties can be avoided if we focus instead on the interesting effects of the expectation itself.

The expectation effect is simply the causal path $E \rightarrow D$ (Figure 2): the tendency of an outcome to realize when it is expected (to some degree), as compared with that tendency when there is no expectation (or relative to some other degree of expectation). Of course, the arrow $E \rightarrow D$ might conceal many causal paths with intermediary variables. That collection of paths is sometimes called “multiple placebo effects”, which are “explained” by the intermediaries. Both ideas, however, can survive with no reference to placebo.

Modifiers of the expectation effect are of great interest and an integral part of causal inquiry: does that effect vary by the treatment taken (Figure 2), by age, by sex, or by any other cause of the outcome? In contrast, the many causes of the expectation itself should be viewed as separate research topics. The offered treatment ($T_{OFFERED}$) and blinding status are two causes of expectation (Figure 2), but there are others. A mental state of expectation is likely affected by variables such as prior experience, trust, and education – to name a few.

Recalling that effect modification is a reciprocal phenomenon, researchers may also be interested in modification of the treatment effect by the level of expectation (Figure 2) and by other causal variables. The two avenues of research are complementary and worth pursuing. Neither requires any reference to placebo.
A misleading emphasis on the placebo effect, rather than the expectation effect, has led to several research designs whose validity may be questioned. We scrutinize, next, three designs in light of the analysis that was offered here.

**A three-arm trial**

Figure 5 shows the causal structure of a blinded trial. The special circumstances of a given trial, such as the study protocol, the randomization procedure, and blinding, may be viewed as conditioning on \( R \) and \( B \). With perfect blinding, the paths \( T_{OFFERED} \rightarrow E \rightarrow D \) and \( T_{OFFERED} \rightarrow D \) will generate null effects, because a concealed treatment should have no effect on the patient’s expectation of the outcome, or on the behavior of the physician. Therefore, perfect blinding ensures that the offered treatment affects the outcome only through taking that treatment (Figure 5). Under this ideal causal structure, the effect of a drug may be tested against any competitor, whatever its name might be: placebo or otherwise.

The placebo effect is often studied by adding a third, no-treatment arm (Figure 6A). In fact, many authors define the placebo effect as the contrast between the placebo group and the untreated group (Figure 4C). As we saw, however, that effect may be modified by expectation, and in such cases the design embeds effect-modification bias. Instead of estimating multiple different effects of placebo (whatever the word means), we pursue some biased average.

The contrast between the placebo group and the untreated group does not estimate the expectation effect either, because not all placebo recipients necessarily share the same expectation, given an unknown treatment. The placebo group in a blinded trial, just like the treatment group, comprises an unknown mixture of patients who expect benefit from the offered treatment and those who don’t, because patients are told that it may be a drug of uncertain benefit, or placebo (Figure 6A). Similarly, indifference is not uniformly shared among patients in the third arm (no treatment), because some patients expect to benefit from the natural healing mechanisms of the body. Moreover, the offered treatment is not the only cause of expectation!

From a methodological standpoint, the attempt to estimate the expectation effect from a three-arm study is prone to information bias: none of the groups is necessarily homogenous with respect to the expectation variable. Stated differently, the treatment group to which a patient is assigned might be a poor measurement of expectation.

**Open versus concealed treatment**

Another design for studying the expectation effect is a two-arm trial comparing an open treatment with a concealed treatment (Figure 6B). Ethical and practical issues aside, all patients receive the treatment, but one group is told so (open treatment) and the other is not (concealed treatment). The contrast between the two groups is attributed to different expectations, given drug treatment. It is assumed that patients who know they are treated expect benefit, whereas their unaware counterparts have no such expectation.

On the one hand the design is better than a three-arm trial because no reference is made to placebo; the effect may be estimated for any drug, whether labeled placebo or not. On the other hand, the underlying assumption may be false again, because neither group of patients necessarily shares the same level of expectation. Not every patient who is being treated, and knows that, believes that the treatment will help. Likewise, patients who think that they are not treated might still expect a positive outcome, say, because they trust natural defense mechanisms more than they trust a drug.
The balanced placebo design

Figure 6C shows the so-called balanced placebo design, essentially a $2 \times 2$ factorial design. Patients are randomized into one of four study groups based on the offered treatment (drug or placebo) and what they are told about their treatment (“it is a drug” or “it is placebo”). Evidently, in two groups patients are deceived about what they are offered, but we set ethical issues aside.

Like any other $2 \times 2$ factorial design, four causal parameters may be estimated, including modified effects. Looking down the columns (Figure 6C), we can estimate two effects of offering a drug (versus offering placebo): 1) when the patient is told “it is a drug”; 2) when the patient is told “it is placebo”. Looking across the rows, we can estimate two effects of whatever the patient is told: 1) when a drug is offered; 2) when placebo is offered.

Again, neither patients who are told “it is a drug” nor patients who are told “it is placebo” necessarily share the presumed expectation about the outcome. Telling the patient something is one of the causes of expectation, but not the only one.

All three designs – three-arm, concealed treatment, and the balanced design – are based on a tradeoff between randomization and information bias. Since patients cannot be randomized to their level of expectation (likely a continuous property), the variable is measured by one of its causes (what the patient is told), which is amenable to randomization. But that approach might not be in the best interest of science. If we want to study the effects of expectation, we should try to measure the variable as accurately as we can, regardless of any attempt to manipulate its value. The effects of mental state variables, like those of many other variables (genes, weight, smoking), should be studied by an observational design.

Discussion

Placebo and the placebo effect are confusing ideas, but neither is essential for the advancement of knowledge. The relevant causal variables are treatment and expectation, which may modify each other’s effect on the outcome. Biomedical science should direct its attention to studying the causes and effects of the patient’s expectation, rather than endlessly debating the meaning and significance of placebo treatments. To that end, medical researchers should benefit from collaboration with psychologists.

The collaboration should begin by discussing expectation itself, a psychological variable whose meaning is not entirely clear. Psychological variables are difficult to study because we learn to infer them from behaviors, rather than sense them directly, and our language has not developed to express them accurately. For instance, “happy” and “sad” seem to be two extreme values of one variable (just as “tall” and “short” are two extreme values of the variable height). To our knowledge, however, no name exists for the variable whose values range from “happy” to “sad” (“mood” is far too broad).

The variable “expectation” does have a name, but we wonder about the nature of that variable, as illustrated by the following questions: is there a different expectation variable for each outcome (expectation of falling, expectation of surviving, etc), or maybe one variable contains the expectation of all outcomes simultaneously? Is there expectation of an outcome of which our mind is not aware? For example, before people knew that cancer exists, did they still have an expectation of getting cancer? These questions and others should be discussed and answered before designing any study about “expectation”. They are critical in deciding how to measure the variable(s).

Whatever expectation means, measuring it is no small challenge, because we are dealing with a mental state variable, which might even include a subconscious component. Manipulating the variable one way or another is not the best method, and self-reporting might not be accurate. Nonetheless, this kind of challenge is inherent in most, if not all, psychological research. The preferred approach should be observational cohorts that are nested within randomized blinded trials – with no deception of patients. Let’s not deceive ourselves that deception can be used to randomize patients to their level of expectation.

Disclosure

The authors do not have any conflict of interest in this work.

References