

Illustration by Phillip Payette

By Andrew Lakoff

The Mousetrap

Managing the Placebo Effect in Antidepressant Trials

Drug development is a risky and uncertain practice even in the case of illnesses that lend themselves to measures of efficacy in readily quantifiable and biological terms. In the case of disorders such as depression or anxiety, clinical trials bear an added layer of complexity. In these instances, the identification of appropriate patients and endpoints depends crucially on techniques—and on “standardized” scales—whose reliability and validity are questionable even to those who use them most, namely, drug developers.

The difficulty of demonstrating efficacy in large-scale, randomized clinical trials has been a major obstacle in the development of novel anti-depressants, raising costs and delaying regulatory approval. Indeed, although the results of failed industry trials are not published, it is well known among developers that the majority of industry-sponsored antidepressant trials fail, and that some companies have had to run as many as ten trials before establishing the efficacy of a given drug. A number of explanations have been offered for this high rate of failure, including the heterogeneity of clinically tested populations, the unreliability of rating instruments,

fluctuations in the course of illness, and high rates of placebo response.

The high placebo response rate in depression trials poses an especially vexing problem. The placebo effect is unpredictable and seemingly unmanageable, and costs drug companies hundreds of millions of dollars in failed trials and delayed or shelved compounds. Because the response rate to placebo in depression ranges from twelve to fifty percent—not greatly different from response rates to antidepressants themselves (thirty to seventy percent)—it can even seem to impugn the efficacy of established and marketed drugs, used as active comparators in trials of novel compounds (1).¹ Moreover, it seems that the placebo response rate has actually been increasing in recent years, for unknown reasons (2).

Developers of antidepressants have tried a number of strategies to reduce placebo responses in order to demonstrate drug efficacy, but have generally been frustrated. However, a recent experiment indicates that in taking the opposite course, in actually trying to produce the placebo effect, researchers may be able to manipulate its uncertainties in the service of

bringing a drug to market. In addition, the developers' very pragmatic orientation toward the seemingly mysterious placebo response seems to resonate with many of the conceptual approaches that medical anthropologists have followed in considering the placebo response.

Uncertainties of the Placebo Effect

The placebo effect is intimately tied to uncertainty and to secrecy—its efficacy seems to be premised precisely on the patient's lack of knowledge. It represents an embodied relationship to the future, insofar as hope or expectation of cure can in fact produce somatic and psychic transformation. Although its mechanisms are unclear, it is undeniable that deception or "intentional ignorance" (3) is a key element in its potency. Many theories of placebo response attribute this ignorance to the patient, whereas the doctor is seen as the site of potentially manipulative knowledge and authority; such theories hold that the patient's trust in the doctor and the medical establishment drives the placebo effect (4).

Medical anthropologists take a more intersubjective view, analyzing the operations of the placebo effect in terms of the importance of context in the process of healing: the structured communication between doctor and patient activates a therapeutic potentiality that exists within individuals (5). Accordingly, in the therapeutic setting, the drug is both substance and symbol, and this duality is one of a number of elements that cannot be neatly untangled. Following Claude Levi-Strauss' analysis of the efficacy of symbols, anthropologists have re-christened the placebo effect the "meaning-response," situating it in terms of the performative and symbolic efficacy of biomedicine (6). In 1963, Levi-Strauss wrote that there were three complementary sources of magic's efficacy:

First, the sorcerer's belief in the effectiveness of his techniques; second, the patient's or victim's belief in the sorcerer's power; and, finally, the faith and expectations of the group, which constantly acts as a sort of gravitational field within which the relationship between sorcerer and bewitched is located and defined.

Thus, for Levi-Strauss, the experience of the sick person was the least important part of the system of symbolic "healing." The

fundamental problem, he thought, revolved around the relationship between the healer and the expectations of the group. For contemporary theorists of the symbolic efficacy of biomedicine, the shared sense of the authority of scientific knowledge and of its objects produces its own healing effect.

Signal and Noise: The Specificity Model

Contemporary biomedicine takes a very different view of both the role of the drug and the site of healing: the drug is understood to operate directly on a physical problem through its biochemical effects on the body of the patient. This notion of targeted drug effects correlates with biomedicine's tendency to order persons into classifications based on specific disease entities, such as "depression," that are presumed to exist outside the particular manifestation of illness in the individual (5, 8). Regulatory rationality combines these elements, assuming that such populations should then be treatable by the same kind of direct intervention—in this case, antidepressants. The goal of the clinical trial is then to test the efficacy of the drug on a population with a specific disease.

The problem is that it remains unclear whether patients classified as depressed all share the same disease. The key tool for assembling populations for clinical trials and measuring their response is the standardized rating scale—in the case of depression, the Hamilton Depression Rating Scale (HDRS). The scale attempts to transform the complexity and heterogeneity of individual illnesses into something collectively measurable, by turning stories into numbers—translating subjective experience into calculable entities. However, drug developers are not convinced that all those patients who are classified by standardized rating scales as "depressed" actually share the same illness. In fact, they are quite skeptical about the capacity of the standard rating scales to produce a consistent patient population for testing. From painful experience, they have learned that patients admitted under these criteria vary tremendously in their response to drugs and placebos—and also, that the scales are applied inconsistently by raters. Attempts to standardize the application of the scales—video training sessions, site audits, and other methods—seem not to have improved trial success rates.

For drug developers, it is actually the drug, rather than the depressed patient, that serves as a stable reference point. For pharmacologists, statisticians, and epidemiologists from companies with new compounds in late stages of development, whether the drug works is not really what is in question: at this point, given the hurdles it has already passed through, drug developers already have confidence in the efficacy of the substance. This can be seen in drug developers' use of the term "signal detection" to refer to the goal of the trial. Here the drug is already presumed to have specific efficacy—that is, a signal to transmit—and the problem is how to pick up the signal.

¹ Even drugs that are on the market often fail to distinguish themselves from placebo, so that trials performed relative to a marketed comparator can be uninformative. Trial drugs that perform as well as the comparator may reflect a lack of response for both the marketed and trial drugs. By the same token, the marketed comparator serves as a form of insurance for the trial drug. If neither the trial drug nor the comparator performs better than placebo, the presumption is that the failure lies in the trial, not the drug.

Reflections

From the perspective of drug developers, when trials fail, it is not necessarily that the drug does not work but that “noise” has crept into the process.

Noise Reduction

A biostatistician at a major pharmaceutical firm who is involved in developing a novel class of antidepressant compound expressed his frustration at the challenges of depression trials: “On the one hand,” he told me, “you want to standardize raters’ behavior as much as possible in order to glean consistent data—but then you might ‘dampen the signal’ by failing to note clinical signs not measured by the rating scales.” Yet if too much focus is placed on close clinical observation, rates of placebo response might increase because of the attachment that could then form between the rater and the subject. “There are so many different problems in this area,” he continued, “it’s like taking a balloon and trying to squeeze it in a certain place—the air just gets pushed elsewhere.”

Although advocates of alternative medicine have begun to see the placebo effect as a possible source of new forms of therapy, drug companies continue to view placebo effects as impediments to proving efficacy and bringing new drugs to market. For drug developers, understanding the operations of the placebo effect is not a question of anthropological curiosity but of corporate necessity. They are forced to confront the effect by the exigencies of regulatory guidelines. And yet their very pragmatic orientation to manipulating the placebo effect may also provide insight into its workings. As a part of these efforts, they have located a number of possible candidates for the locus of apparent placebo effect, including the patient, the evaluator, the measuring device, and the illness itself. A first important distinction they make is between *artifactual* and *real* placebo response.

Drug researchers cite at least two kinds of artifactual placebo response. One has to do with the motives of raters at the trial site. If the site is under pressure to rapidly enroll patients, raters may inflate their scores at the beginning of the trial; later, under more accurate measurement, the apparent improvement of subjects receiving placebo will artifactually magnify the placebo response. A second potential source of artifactual placebo response is statistical “regression to the mean,” whereby the patient has a rapidly fluctuating course of illness, and enrolls in the trial when it is at its worst. Then, if the illness improves over the course of the trial, one can again see what looks like, but is not really a placebo response.

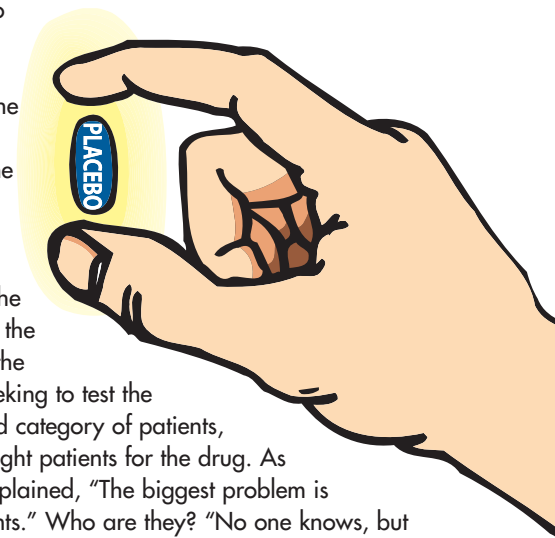
Real placebo response is attributed to the characteristics of either the trial site or the patients. For frustrated trial sponsors, one possible cause in the former instance has to do with “investigator behavior.” If the site-based investigators—contracted by the drug developers—perform what is termed “covert psychotherapy,” or in some other way give those who

have been assigned placebo the sense that they are being helped in any way, an unwanted placebo response may ensue. Here, developers presume a specific understanding of what the placebo response is: a reaction that arises from the patient’s hope, expectation, and/or an affective investment in the healer or in the treatment itself. One well-known authority on depression clinical trials argues that such “non-specific supportive contact” can even include the filling out of forms, especially if these are meant to reassure patients about their participation in the trial. He sternly advises those who wish to improve their trial results: “Patients who are overly sensitive to reassurance need to be identified and if possible excluded” (2).

The Right Patients for the Drug

This advice—to exclude overly sensitive subjects—relates to a more general set of trial design strategies based on the hypothesis that the source of excessive placebo response is the presence of a certain class of patients who are overly susceptible to placebo. If measuring devices such as rating scales are supposed to record the signal, the patients’ role is to transmit it—they are the drug’s medium. Given the heterogeneity of the depressed patient population, antidepressant developers here shift the locus of uncertainty in the trial from the drug to the patient: instead of seeking to test the drug on an established category of patients, they seek to find the right patients for the drug. As the biostatistician complained, “The biggest problem is getting the right patients.” Who are they? “No one knows, but there are a lot of different ideas.” He mentioned some of the possible clues to placebo-susceptibility: duration of illness, family history, and age of onset. But “they don’t hang together from one trial to the other,” he said. “Things disappear as you look at them more closely.”

The most salient subpopulations to be delineated before the trial begins are drug responders and placebo responders. Consistent with the notion of the patient as the “medium” of drug efficacy, subject populations are understood here as potential transmitters of efficacy. As S.A. Montgomery writes, “samples selected for trials should be able to deliver a predicted response to drug and not to placebo.” Unfortunately, he continues, standardized diagnostic criteria are “not up to the task of distinguishing between clear drug responsive patients and placebo responders.” However, he notes that seasoned



investigators seem to intuitively be able to make this distinction and exclude the latter from trials. Such tacit, non-standardizable knowledge, linked to experience, might then explain why some trial sites do better than others in demonstrating drug efficacy.

But how can one tell who is likely to respond to a placebo? The first attempts to delineate the characteristics of placebo-responders began after the recognition of the importance of the placebo effect in the years after World War II, when the double-blind, randomized controlled trial was accepted as the means to police fraudulent medications (9). In the rationale of the randomized controlled trial, the placebo effect was both an epistemological necessity and a practical obstacle to showing true drug efficacy. If one could ascertain before the trial who was likely to respond to placebo and therefore contaminate the results, one could ostensibly eliminate them from the trial beforehand and improve the chances for success.

In the 1950s, pioneer placebo researcher Louis Lasagna used personality questionnaires and Rorschach tests to characterize the typical placebo-reactor in post-operative pain trials. When asked, "What sort of people do you like best?" placebo reactors were apparently more likely to respond, "Oh, I like everyone." They were more often active churchgoers than non-reactors, and had less formal education (10). With regard to the Rorschach results, the researchers noted:

In contrast to the non-reactors the reactors were... more anxious, more self-centered and preoccupied with internal bodily processes, and more emotionally labile.... [T]he reactors are in general individuals whose instinctual needs are greater and whose control over the social expression of these needs is less strongly defined and developed than the non-reactors.

In the 1970s, researchers found that placebo reactors scored higher on the "Social Acquiescence Scale," based on agreement with proverbs such as, "obedience is the mother of success," "seeing is believing," and "one false friend can do more harm than 100 enemies" (11).

This line of research linking placebo response to personality characteristics eventually faded, having failed to provide convincing evidence, and researchers have more recently focused on somatic rather than psychological factors. They hypothesize that milder severity of illness, a more rapidly fluctuating course, or certain kinds of somatic complaints correlate with placebo response (12). Accordingly, one explanation for increasing rates of placebo response rate might then be that less severely ill patients are now being used more often, given the shortage of patients for clinical trials (2). But efforts to operationalize these criteria bring other problems for drug developers, such as limiting the potential indication of the approved drug or extending the length of the trial in the search for better-qualified patients.

The Mousetrap: Staging the Trial

Given the difficulty of trying to identify placebo responders by delineating their characteristics, antidepressant researchers have turned to a more pragmatic approach that might be named the "mousetrap technique," after the play within a play that Hamlet staged in order to goad his father's murderer into revealing his guilt. Here, experimenters in effect stage the trial before it actually begins, giving all patients placebo for a week, and then eliminating those who respond from the trial. This staged trial is called a "single-blind placebo run-in period," inasmuch as the doctors know that all of the patients are receiving a placebo, while patients remain ignorant that the real trial has yet to begin (13). With this approach, it does not matter why patients respond to placebo, nor does the knowledge or technique of the investigator matter. One simply needs to know which patients have responded in order to eliminate them. However, these efforts have also proven disappointing—placebo response rates during these run-in periods tend to be low, and so it has not been possible to eliminate most of the potential placebo responders. In the actual trial that follows the run-in period, other subjects continue to respond to the placebo, drowning out the signal of drug efficacy and undermining the trial.

As a result, some trial consultants have called for the abandonment of the run-in period (14). But researchers at one major pharmaceutical company recently noticed some interesting things from looking at the post-trial data from trials with placebo run-ins (15). They saw that the treatment response curve tended to be flat until the point of randomization to drug or placebo—at week two or three. Sharp improvement then set in, even in patients who remained on placebo after the run-in period. The investigators hypothesized that there was something about the shift from the run-in period to the actual trial that increased the placebo response. From an anthropological perspective, this suggestion is intriguing in terms of thinking about how the placebo effect might work: What changes at this point in the trial, after all, is not the subjects' level of knowledge, but rather the evaluators'. As the trial moves from "single-blind" to "double-blind," it is the evaluators who shift from a condition of knowledge to one of ignorance, from certainty to uncertainty.

The researchers then designed an experimental study to pursue whether it was the double-blind nature of the post-randomization phase that was causing this rise in response. They built in a "double-blind placebo run-in period," in which neither the patients nor the site-based evaluators knew of the placebo run-in period. Strikingly, the run-in placebo response rate rose to twenty-eight percent of the subjects, as opposed to five percent in the single-blind run-in period. By eliminating these placebo responders from the efficacy analysis at the end of the trial, the evidence of drug efficacy improved significantly. Unlike the partial simulation of the single-blind run-in, the experimenters "staged" a full simulation of the trial before the

Reflections

Science in the cultural context

actual trial, in order to make visible the heretofore hidden population of potential placebo responders.

The success of the double-blind run-in experiment is worth reflecting on in terms of what it may reveal about the placebo effect. It seems to show that placebo response is significantly higher when the doctors do not know that they are giving a placebo—that the doctors' uncertainty may be as important as the patients'. Recall that Levi-Strauss argued for the centrality of the sorcerer's belief in his own fabrication to the efficacy of his magic. It is often assumed in the literature on the placebo that it is the patient's trust combined with the doctor's purposeful deception that is the source of placebo response. Yet in the experiment outlined above, the doctor is deceived as well: this study shifts the locus of placebo effect from the patient—as site of a “meaning-response”—to the doctor–patient relationship as a site of uncertainty and potentiality.

Although anthropologists have long understood the role of meaning and intersubjectivity in healing, here such knowledge was cleverly put to use in a very pragmatic context: The experimenters harnessed the actions of the placebo effect in order to eliminate placebo responders. Rather than trying to reduce the placebo effect by unraveling its mechanisms, the experimenters actually increased it by multiplying the sites of uncertainty. In order to take advantage of the placebo effect in this way, the researchers did not need to know why what they were doing was increasing it, only how to do so. The trial was effective—which was the practical task at hand—not because they had solved the mystery of the placebo effect, but because, in creating the double-blind run-in, they had found a way to eliminate placebo responders without needing to know anything about their characteristics. And in doing so, they had unwittingly provided evidence for anthropologists' argument that the hopes and expectations of the *doctor* are as crucial as those of the patient in the performance of healing.

REFERENCES

1. Walsh, B.T., Siedman, S.N., Sysko, R., and Gould, M. Placebo response studies of major depression: Variable, substantial, and growing. *JAMA* **287**, 1840–1847 (2002).
2. Montgomery, S.A. The failure of placebo-controlled studies. *Eur. Neuropsychol.* **9**, 271–276 (1999).
3. Kaptchuk, Ted J. “Intentional ignorance: A history of blind assessment and placebo controls in medicine. *Bull. Hist. Med.* **72**, 389–433 (1998).
4. Nuland, S. The uncertain art: Mind, body, and the doctor. *Am. Scholar* (Summer, 2001).
5. Kleinman, Arthur. *Rethinking Psychiatry. From Cultural Category to Personal Experience*. Free Press, New York (1988).
6. Moerman, D. Physiology and symbols: The anthropological implications of the placebo effect, in *The Anthropology of Medicine: From Culture to Method*. (Romanucci-Ross, L. Moerman, D.E., and Tancredi, L.R. eds) Bergin and Garvey, Westport, CT (1997).
7. Levi-Strauss, Claude. The sorcerer and his magic, in *Structural Anthropology*, Basic Books, New York (1963).
8. Rosenberg, C. *Explaining Epidemics, and Other Studies in the History of Medicine*. Cambridge University, Cambridge (1992).
9. Harrington, A. “Seeing” the placebo effect: Historical legacies and present opportunities, in *The Science of the Placebo: Toward an Interdisciplinary Research Agenda* (in press).
10. Lasagna, L., Mosteller, F., von Felsinger, J.M., and Beecher, H.K. A study of the placebo response. *Am. J. Med.* (June 1954).
11. McNair, D.M., Gardos, G. Haskell, D.S., and Fisher, S. Placebo response, placebo effect, and two attributes. *Psychopharmacology* **63**, 245–250 (1979).
12. Schatzberg, A.F. and Kraemer, H.C. Use of placebo control groups in evaluating efficacy of treatment of unipolar major depression. *Biol. Psychiatry* **47**, 736–744 (2000).
13. Quitkin, F.M., McGrath, P.J., Stewart, J.W. et al. Placebo run-in period in studies of depressive disorders. Clinical, heuristic and research implications. *Brit. J. Psychiatry* **173**, 242–248 (1998).
14. Trivedi, M.H. and Rush, H. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology* **11**, 33–43 (1994).
15. Faries, D.E., Heiligenstein, J.H., Tollefson, G.D., and Potter, W.Z. The double blind variable placebo lead-in period: Results from two antidepressant clinical trials. *J. Clin. Psychopharmacol.* **21**, 561–568 (2001).

Andrew Lakoff, PhD, is an NIMH postdoctoral fellow in the Department of Social Medicine at Harvard Medical School. He is a medical anthropologist currently pursuing research on the process of drug development.