

Does the placebo exist?

For some four decades, the medical world has held that placebos have clinical effects. Clinical research has built on this belief, especially drug research. A new drug must do measurably better than a placebo against which it is compared. Now, in a boldly assertive new study in the *New England Journal of Medicine*, two Danish researchers, Asbjorn Hróbjartsson and Peter C. Gøtzsche, say that the accepted wisdom of the medical world is wrong. Like the emperor who has no clothes, the placebo, according to Hróbjartsson and Gøtzsche, essentially has no clinical effects.

Hróbjartsson and Gøtzsche's study is called "Is the placebo powerless?" The subtitle describes the approach: "An analysis of clinical trials comparing placebo with no treatment." Clinical trials rarely compare placebo with no treatment—they focus on the effectiveness of the drug not the placebo (which, in any case, virtually everyone assumed had certain undetermined effects)—but Hróbjartsson and Gøtzsche, to the surprise of many, found 114 clinical trials that each offered the opportunity to compare placebo with no treatment. They conducted a meta-analysis of these studies, subcategorizing the outcomes into subjective or objective and binary (as in dead or alive) or continuous (as in gradations of fever), and concluded in their abstract that placebos "had possible small benefits in studies with continuous subjective outcomes and for the treatment of pain" but "no significant effects on objective or binary outcomes," and that, in all, there was "little evidence in general that placebos had powerful clinical effects." "Outside the setting of clinical trials," they declared, "there is no justification for the use of placebos."

Not surprisingly for findings that upset a seemingly fixed medical belief, the study has received broad attention both in the general press—The *New York Times*, for example, put a notice of the story on the front page—and among researchers, most sharply among researchers involved in efforts to identify the mind's contributions to physical health. For such researchers, the existence of placebos provides *prima facie* evidence that (somehow, to some degree, in an uncertain number of conditions) the mind can produce effects with clinical significance. If Hróbjartsson and Gøtzsche are right, then a cornerstone of the mind-body perspective is suddenly dust. Gina Kolata, who wrote The *New York Times* story on the placebo paper, drew attention to this point by noting that an editorial comment on the study by John C. Bailer III "said the findings called into question some mind-body beliefs."

The following five comments on Hróbjartsson and Gøtzsche's findings, generally from within the mind-body research community, are all critical to varying degrees (and, perhaps it should be said, with varying degrees of harshness). The arguments and the response of Hróbjartsson and Gøtzsche, with their careful distinctions of what is and is not under investigation in their study, raise important clarifications.

Here I offer two quick side notes. First, a striking aspect of the comments is the remarkable array of placebo studies they call upon (maybe *unearth* is the better term)—studies of oral placebos vs intravenous placebos, studies of the relapse rate in placebo effects, studies showing that the beliefs of doctors contributes to rates of cure in patients, studies of credible and uncredible placebos, and on and on. Why has some research group not sought funds to establish "The Placebo Project," to systemize all rigorous placebo studies and analyze them for what they tell or indicate or suggest about the powers of the mind to affect health and illness?

Second, I want to register my disapproval of Gina Kolata's effort to define "some mind-body beliefs" in her *New York Times* story. After noting John Bailer's assessment that Hróbjartsson and Gøtzsche's findings "called into question some mind-body beliefs," she then explains such beliefs as follows. "These are arguments that use the placebo effect to conclude that the mind can so profoundly affect the course of disease that people should be able to harness this power and think themselves well"—my emphasis. Who is Kolata talking to to come up with such a formulation? Apart from several popular writers, what mind-body researcher would ever claim that people could "think themselves well?" Is this claim made by researchers in psychoneuroimmunology? Is this claim made by researchers studying social support or stress or disclosure or hypnosis or psychosocial interventions for cancer patients or the placebo for that matter? By offering such a foolish view of mind-body beliefs, Kolata seems really to be saying that the mind-body enterprise is not worthy of scientific inquiry. This is a shameful position.

For those who have not had the opportunity to read the Hróbjartsson and Gotzsche study, the abstract outlines the main points.

Is the Placebo powerless? An analysis of clinical trials comparing placebo with no treatment

Asbjørn Hróbjartsson MD and Peter C. Gotzsche MD

Background: Placebo treatments have been reported to help patients with many diseases, but the quality of the evidence supporting this finding has not been rigorously evaluated.

Methods: We conducted a systematic review of clinical trials in which patients were randomly assigned to either placebo or no treatment. A placebo could be pharmacologic (e.g. a tablet), physical (e.g. a manipulation), or psychological (e.g. a conversation).

Results: We identified 130 trials that met our inclusion criteria. After the exclusion of 16 trials without relevant data on outcomes, there were 32 with binary outcomes (involving 3795 patients, with a median of 51 patients per trial) and 82 with continuous outcomes (involving 4730 patients, with a median of 27 patients per trial). As compared with no treatment, placebo had no significant effect on binary outcomes, regardless of whether these outcomes were

subjective or objective. For the trials with continuous outcomes, placebo had a beneficial effect, but the effect decreased with increasing sample size, indicating a possible bias related to the effects of small trials. The pooled standardized mean difference was significant for the trials with subjective outcomes but not for those with objective outcomes. In 27 trials involving the treatment of pain, placebo had a beneficial effect, as indicated by a reduction in the intensity of pain of 6.5 mm on a 100 mm visual-analogue scale.

Conclusions: We found little evidence in general that placebos had powerful clinical effects. Although placebos had no significant effects on objective or binary outcomes, they had possible small benefits in studies with continuous subjective outcomes and for the treatment of pain. Outside the setting of clinical trials, there is no justification for the use of placebos. (N Engl J Med 2001; 344:1594–602.)

Copyright © 2001 Massachusetts Medical Society.

Much ado about nothing

Robert Ader

Robert Ader PhD is the Director of the Center for Psychoneuroimmunology Research in the Department of Psychiatry at the University of Rochester Medical Center, Rochester, New York 14642.

*What we observe is not nature itself, but
Nature exposed to our method of questioning.*

Werner Heisenberg

A recent article in the prestigious *New England Journal of Medicine* (Hróbjartsson and Gøtzsche 2001) serves to highlight some of the difficulties in defining and analyzing placebo effects. The findings of this study are not impressive, as the accompanying editorial by John Bailar (2001) would have you believe. Indeed, by using amusing but inadequate analogy and relegating placebo effects to the realm of myth, the editorial does a disservice to any effort (however successful) to demystify placebo effects. Also, others, at least some of those quoted in the lay press, have been too quick to accept these data at face value, revealing long-held suspicions that there really isn't and couldn't be any such thing as a placebo effect. The reaction is reminiscent of the premature obituary for the contributing role of psychological factors when *H. pylori* was promulgated as the cause of duodenal ulcers. It also illustrates Viza's (1998) contention that "Modern medical logic would rather that treatments be inefficacious than incomprehensible."

Editorial comments aside, Hróbjartsson and Gøtzsche's meta-analysis of clinical trials comparing placebo interventions with no treatment is seriously flawed, and the results are unconvincing for a variety of reasons.

There are, to begin, procedural difficulties. In the same way that group assignment should be randomized and experimenters/observers should

have no knowledge of group assignment, the decision to include or exclude a study from a pooled statistical analysis should be made without knowledge of the outcome of the study. In the present instance, disagreements concerning eligibility were resolved by a discussion between investigators who had read the studies in full—not just the procedures. Thus, these data were subject to a potential source of bias comparable to that used by the authors to disqualify placebo trials involving objective outcomes.

Actually, it is not clear to what extent this analysis included studies in which observers were blinded. In the section on Methods, the authors state that, "We also excluded studies...if the person who assessed *objective* outcomes [italics added] was aware of group assignments" (p. 1595). Later, however, they report that "only two trials with binary objective outcomes...included observers who were clearly unaware of the group assignments" (p. 1597). Thus, in seven of nine trials with objective binary outcomes (Table 1, p. 1596), observers *were* aware of group assignments. According to the Methods, these seven studies did not meet an original criterion for inclusion in the analysis.

The authors claim that blinded evaluation of *subjective* outcomes is not possible. Certainly, patients in an untreated group would know they were not being treated, but that does not preclude a protocol in which observers are kept blind to the group assignment of the study population. From the enumeration of studies with either subjective (76) or objective (38) outcomes (Table 1, p. 1596), we can calculate that in at least 70% of the studies in this analysis (the 76 with subjective outcomes plus some with objective outcomes) observers were aware of group assignments. Since 112 of the 114 trials included a third, active drug treatment group, observer bias could have deflated the effect of placebo treatment.

Other procedural difficulties derive from the authors' conceptualization of placebo effects, which, in the present instance, is a very traditional one. Attention was directed to such simple variables as dropout rate; whether the original authors enunciated an interest in placebo effects per se; whether concealed randomization of

Discussion

Ader

patients (only five of 114 trials) mattered; or to the analysis of still other factors for which the necessary power was, admittedly, too low for meaningful statistical analysis. In contrast, more relevant studies of placebo effects, such as those measuring relapse rate (Suchman & Ader 1992) or adherence to placebo treatment (Horwitz & Horwitz 1993), seem to have been ignored. In addition, the decision to use only the first treatment in the case of crossover studies ignores the several studies of sequence effects (Batterman 1965; Moertel et al. 1976; Sunshine et al. 1964) and thereby underestimates placebo effects. There are, undoubtedly, several other "favorite" studies that were not included in this meta-analysis.

Granted, definitions of placebo and placebo effects are problematic. However, and especially in the case of a meta-analysis, we must understand what went into the analysis in order to interpret what came out. The authors' definition of a placebo as whatever the authors of individual studies said it was is one way to sidestep the problem of differentiating among the several phenomena described as placebo effects. It is, however, an inadequate way in which to justify the mix of placebo interventions included in this analysis. Definitions aside, no rationale is given for the assumption that different placebos are equivalent or even comparable and, therefore, reflect the same underlying mechanism, whatever that might be. From an analytical perspective, it makes no sense to this reader to equate a "nondirectional, neutral discussion between the patient and the treatment provider" (p. 1596), or even a verbal suggestion that symptoms will be gone in a week, with repeated administration of an inert substance over some longer period of time.

To further complicate the mix of trials in this analysis, the "no treatment" condition was described as "observation only or standard therapy; in the latter case, all patients in the trial received standard therapy, and the placebo was additional." In some unspecified proportion, then, the meta-analysis included comparisons of "placebo treatment versus no treatment" and "standard therapy plus placebo versus standard therapy alone." One would, of course, want to know if the magnitude of the difference between

placebo and no-treatment groups would be similar under these disparate conditions. Considering the different baselines from which one would be working, I daresay that would be unlikely.

Of the 40 clinical entities studied, only one, pain, was represented by more than seven different studies; 34 of the 40 entities were represented by three or fewer studies. Despite the presumed procedural variability, pain relief was achieved by placebo administration. That the effects were "small" probably results from the use of different placebos and procedures in the pain trials. It is not that other clinical entities did not show placebo effects; the number of samples was too small to permit evaluation of such effects, that is, the meta-analysis was premature.

The first questionable assumption underlying Hróbjartsson and Gøtzsche's meta-analysis, then, is that all "placebo" interventions are equivalent, that is, that there is a single placebo effect. This simplification is analogous to assuming that all so-called stressors are equivalent in their effects, a widely held assumption that is, nevertheless, contradicted by the data. To counter that the studies analyzed by Hróbjartsson and Gøtzsche revealed no differences among the placebo interventions speaks to the second, equally tenuous assumption, namely, that placebo effects are nonspecific. It seems questionable, however, to assume that the effects of any sort of placebo (pharmacologic, physical, or psychological) intervention would be equally efficacious for anemia, a mental handicap, and pain.

Even if one ignores the possibility that some placebo effects (for example, conditioned pharmacotherapeutic response) might be specific (Ader 1997), the likelihood that the psychobiological changes associated with even a single kind of placebo intervention would be relevant to and effective in alleviating the symptoms of any and all clinical entities seems quite remote. Despite the small number of trials for all the other clinical entities, the placebo effect seen in the data derived from the pain studies, themselves heterogeneous, suggests that it would be inadvisable to pool data from placebo trials with different medical outcomes. After all, if the

effects are nonspecific, there would be no reason to expect that studies on pain would be more likely to reveal placebo effects than any other single or combined group of studies of other medical conditions. Thus, evaluating the effects of (different) placebo interventions by combining studies on different disease processes may be analogous to evaluating the efficacy of an active drug by combining the results from studies of its effects on different disease processes.

In fact, there are probably several different kinds of “placebo” effects, and it can already be seen that at least some placebos can exert specific as well as nonspecific effects (Ader 1997). Similarly, the effects of a drug include extra-pharmacologic or “placebo” components that contribute to the organism’s response (recovery). A case in point is the difference between the effects of open and “hidden” drug administration (Amanzio et al. 2001). Also, like the experientially determined sequence effects that characterize placebo administrations, active drug effects are attenuated by the prior experience of ineffective (placebo) medication (Batterman 1965). Clearly, experience and learning influence pharmacotherapeutic responses.

In sum, the combined assumptions that different placebo interventions are equivalent and that placebo effects are applicable to all clinical situations, enabling Hróbjartsson and Gøtzsche to pool a heterogeneous selection of studies, make the entire analysis suspect and obviate any definitive conclusions. If there are a dozen studies, half of which show placebo effects and half of which do not, there is no statistical method that could determine whether or not there is really a placebo effect or the magnitude of any such effect. If a single well-designed study shows placebo effects, then the phenomenon is real. It remains only to define the conditions under which the effect does or does not occur. Therefore, unlike Bailar, we do not peek behind the curtain to uncover a hoax. Peeking behind the curtain and uncovering the factors responsible for placebo effects will not reduce the power of the placebo. On the contrary, an understanding of mediating

factors will more likely contribute to the constructive manipulation and application of placebo effects. It is also likely to provide a better understanding of the therapeutic actions of active drugs, the *in vivo* evaluation of which includes the effects of psychological, social, and cultural influences.

Publication of the Hróbjartsson and Gøtzsche paper was an error in editorial judgment compounded by the publication of Bailar’s accompanying editorial comment. The *New England Journal of Medicine* has been viewed as having a bias against matters psychological and a double standard with respect to the evaluation of such research. The publication of these two articles will reinforce that impression.

References

- Ader R 1997 The role of conditioning in pharmacotherapy. In: Harrington A, ed. *The Placebo Effect: An interdisciplinary exploration*. Cambridge: Harvard University Press
- Amanzio M, Pollo A, Maggi G, Benedetti F 2001 Response variability to analgesics: A role for nonspecific activation of endogenous opioids. *Pain* 90:205–215
- Bailar JC III 2001 The powerful placebo and the wizard of Oz. *New England Journal of Medicine* 344:1630–1632
- Batterman RC 1965 Methodology of analgesic evaluation: Experience with orphenadrine citrate compound. *Current Therapeutic Research* 7:639–647
- Horwitz RI, Horwitz SM 1993 Adherence to treatment and health outcomes. *Archives of Internal Medicine* 153:1863–1868
- Hróbjartsson A, Gøtzsche PC 2001 Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 344:1595–1602
- Moertel CG, Taylor WF, Roth A, Tyce AJ 1976 Who responds to sugar pills? *May Clinic Proceedings* 51:96–100
- Suchman AL, Ader R 1992 Classical conditioning and placebo effects in crossover studies. *Clinical Pharmacology and Therapeutics* 52:372–377
- Sunshine A, Laska E, Meisner M, Morgan S 1964 Analgesic studies of indomethacin as analyzed by computer techniques. *Clinical Pharmacology and Therapeutics* 5:699–707
- Viza D 1998 Commentary—From placebo to homeopathy: The fear of the irrational. *The Scientist* 12:8

A challenge to core beliefs

Howard Brody and David Weismantel

Howard Brody MD, PhD is professor of family practice and ethics and humanities in the life sciences at Michigan State University, East Lansing, MI. He is the author (with Daralyn Brody) of *The Placebo Response* (HarperCollins 2000). David Weismantel MD is assistant professor of family practice at Michigan State University.

Most people working today on the subject of the placebo response have based their intellectual approach—and, in many cases, their practical approach to the care of patients—on the idea that the placebo effect is real. The placebo effect, as we understand it, is much broader than what happens when a person swallows a sugar pill or is exposed to a similar dummy treatment. It comprises all those aspects of the healing context that have symbolic significance for the patient. Still, the historical account that we have always believed states that we first became convinced of the power of the placebo effect by seeing how patients reacted when given dummy medication, often improving remarkably. So a study that claims to show that people do not experience relief of symptoms when they take placebos in the context of research trials—Hróbjartsson and Gøtzsche's "Is the placebo powerless? An analysis of trials comparing placebo with no treatment" (2001)—upsets the very core of our belief system.

The timing of the publication of this study is ironic. We are making considerable progress today in identifying a scientific, mechanistic basis for the placebo effect. In particular, the relationship between the placebo effect, experimentally induced pain, and endorphin release has been convincingly traced by Amanzio, Benedetti, and their colleagues in Italy (Amanzio et al. 2001). The biochemical pathways most likely to be involved

in the placebo effect include pathways such as the cortisol–catecholamine pathway and the psychoneuroimmune responses, which have been demonstrated to alter measurable bodily processes and tissues. This means that if Hróbjartsson and Gøtzsche were to expand their study and eventually conclude that the placebo effect is real for subjective endpoints (as their study suggests) but not for objective endpoints (as their study concludes), their results would still be at odds with very persuasive hypotheses about mind–body interactions in health and illness. To be told today that placebos are powerless after all is something like being told that the human genome is a myth just after scientists had announced that they had finally been successful in mapping it in its entirety.

Naturally one's first impulse is to shoot the messenger. That won't work here (ethics aside). Both investigators are thoughtful scientists who know their methods well, and who have no apparent axe to grind. They are very careful not to overextend or overinterpret their results. They performed this meta-analysis as part of the international Cochrane Collaboration and submitted their methods for review before undertaking the study.

Nevertheless, their results conflict with the great bulk of the literature over the past 25 years. Admittedly, few of those studies were done with the care and rigor of the present analysis. Hróbjartsson and Gøtzsche are quite correct in noting that the majority of "studies" of the placebo response do not definitively control for such factors as the natural history of illness and regression to the mean. But the literature does include, for instance, several careful meta-analyses, such as those performed by Kleijnen, de Craen and their colleagues at Amsterdam (de Craen et al. 1996; Kleijnen et al. 1994), all of which showed the existence of placebo effects (though with many fewer studies in each case than were reviewed by Hróbjartsson and Gøtzsche). Imagine that 95% of the existing literature is worthless and must be eliminated; there would still be enough evidence in the

remaining 5% to support the existence of a placebo effect, even if it is more limited than we had thought.

What, then, might be the methodological problems with the present study that would account for its counter-intuitive and contrary results? First, as odd as it sounds, the literature search might have been inadequate. None of us would have guessed that there were as many as 114 papers employing a study design with both a placebo and a no-treatment arm. So, if there were indeed 114 papers, there might just as easily be 200 or 300. The search was done using the terms "placebo" and "no treatment." Since in the vast majority of cases the placebo and no-treatment arms were the control arms, not the focus of interest of the investigators, there is little guarantee that this search strategy would yield all relevant publications. The importance of this problem is that while the total of 114 sounds very impressive, the final yield was very few studies for any given condition, such as asthma, hypertension, etc. Many of the disease-specific sub-analyses of continuous outcomes yielded a trend toward a placebo effect but did not reach statistical significance. Adding relatively few more studies to any of those sub-analyses could have greatly changed the outcome.

A potentially greater problem is correctly noted by the authors themselves (and the accompanying editorial in the *New England Journal of Medicine*). Meta-analyses are generally supposed to evaluate studies with similar protocols, interventions, and outcome measurements. The mathematical combining of results from numerous separate studies hinges on this underlying similarity. The method of the present study, by contrast, is somewhat analogous to asking the question, "Are beta-blockers good medicine?" and combining the results of studies that look at the effects of beta-blockers for coronary artery disease, hypertension, migraine prophylaxis, and benign essential tremor. Perhaps Hróbjartsson and Gøtzsche have found a creative way of extending the meta-analysis methodology beyond its usually accepted limits. But much of what they have done is certainly contrary to the conventional wisdom.

Several possibilities exist for further investigation. There may be some value in a qualitative review of the present set of 114 papers, since Hróbjartsson and Gøtzsche have done us such a great service by identifying them. For example, upon reading the study, two of our colleagues independently proposed that perhaps the very act of being enrolled in a randomized clinical trial exerts so powerful a placebo effect that it outweighs the placebo effect actually due to taking a dummy medication. If this is so, then one might well see evidence of this factor by carefully reviewing the individual studies, or one might find that this explanation is quite implausible. If any such alternative explanation does seem plausible on initial review, the next step would be to design a structured, explicit review of the 114 studies to test this hypothesis further.

Another approach more in keeping with traditional meta-analysis methodology would be to seek and replicate this meta-analysis by focusing on the placebo effect in one disease condition only, such as hypertension. One would probably need to search the literature for all clinical trials involving "hypertension" and "placebo" and then hand-search the abstracts and methods to see which studies also included a no-treatment group. This is basically the method that Kleijnen and colleagues (1994) used successfully in their meta-analysis of the placebo effect in pain. (Incidentally, having both a placebo and a no-treatment arm is *not* the only design that would allow one to draw conclusions about the existence and size of a placebo effect. There may be other control conditions, such as having multiple placebo groups, or the hidden administration of an "active" medication, which would serve as well as a no-treatment comparison.)

We hope we have made clear our respect for the work of Hróbjartsson and Gøtzsche. They have done all of us a service by challenging us in this fashion. If we needed any more reminders that the average randomized, controlled, double-blind trial is a very poor vehicle by which to learn what we want to know about the placebo effect, they

Discussion

Greene et al.

have certainly provided us with an ample dose. Nonetheless, we think in the end there will prove to be more to the placebo story than this.

References

- Amanzio M, Pollo A, Maggi G, Benedetti F 2001 Response variability to analgesics: A role for nonspecific activation of endogenous opioids. *Pain* 90:205–215
- De Craen AJ, Roos PJ, de Vries AL, Kleijnen J 1996 Effect of color of drugs: Systematic review of perceived effect of drugs and of their effectiveness. *British Medical Journal* 313:1624–1626
- Hróbjartsson A, Gøtzsche PC 2001 Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 344:1595–1602
- Kleijnen J, de Craen AJ, van Everdingen J, Krol L 1994 Placebo effect in double-blind clinical trial: A review of interactions with medications. *The Lancet* 344: 1347–1349

The powerful placebo: Doubting the doubters

Penelope J. Greene, Peter M. Wayne, Catherine E. Kerr, Wendy A. Weiger, Eric Jacobson, Peter Goldman, and Ted J. Kaptchuk

Penelope J. Greene PhD [contact] is at the Center for Health Communication, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115 and the Center for the Study of Nutritional Medicine, Beth Israel Deaconess Medical Center, 1 Autumn Street (Kennedy 127), Boston, MA 02215; 617-632-0736; pgreene@hsph.harvard.edu. Peter M. Wayne PhD is at the New England School of Acupuncture, 40 Belmont St, Watertown MA 02472. Catherine E. Kerr PhD and Eric Jacobson PhD, are in the

Department of Social Medicine, Harvard Medical School, Boston, MA 02115. Wendy A. Weiger MD PhD, Peter Goldman MD, and Ted J. Kaptchuk OMD are at the Center for Alternative Medicine Research and Education, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215, and the Division for Research and Education in Complementary and Integrative Medical Theories, Harvard Medical School, Boston, MA 0215. The authors are among members of the Interdisciplinary Placebo Working Group, based at the Center for Alternative Medicine Research and Education at Harvard's Beth Israel Deaconess Medical Center.

Introduction

The recent study by Hróbjartsson and Gøtzsche (2001), a meta-analysis of 114 studies comparing placebo-treated groups with no-treatment groups, has had considerable impact in changing perceptions of the placebo effect from an inherent component of therapy to a “phenomenon” that does not exist at all. It is a bold paper that has been popularly taken as announcing the demise of the fundamental assumption that some kinds of psychosocial factors, such as suggestion, expectation, conditioning, hope, or anxiety reduction, have significant power to imitate pharmacologically active drugs or other therapeutic interventions. The authors instead suggest a placebo without “powerful effects,” whose apparent effects are primarily due to natural history of disease or regression to the mean.

The impact of this article has been amplified by extensive media coverage, much of which overlooked a significant part of the Hróbjartsson and Gøtzsche's conclusion: “We found significant effects of placebo on continuous subjective outcomes and for the treatment of pain.” Two media comments are representative: from *The Independent*, in London: “The oldest trick in the doctor's black bag—giving a patient a dummy pill to make them feel better—may have to be abandoned after scientists yesterday reported that the placebo effect is a myth” (Durham 2001); from *The Chicago Sun Times*: “The placebo effect turns

out to be a mirage...in a review of dozens of medical studies, two Danish researchers found no placebo effect" (Grossman 2001).

We argue that a concept as important as the powerful placebo should not be abandoned based on a single meta-analysis. Indeed, there is strong evidence of significant placebo effects from prospective and mechanistic placebo research. Furthermore, there are weaknesses in the methodology of the Hróbjartsson and Gøtzsche study that call into question their conclusions.

Hróbjartsson and Gøtzsche restrict their discussion to a narrow definition: They take "placebo effect" to mean the clinical outcome of the dummy control in a randomized controlled trial. However, a broader view of the placebo effect would include the entire package of "non-specific" effects of the patient-physician relationship and clinical context, including such behaviors as the communication of concern, monitoring and diagnostic procedures, labeling or explanation of the disease, and, more importantly, the impact of factors such as expectation, hope, and anxiety reduction. Hróbjartsson and Gøtzsche do not appear to tackle this broader placebo concept ("...placebos are generally control treatments with a similar appearance to the study treatments but without their specific activity") and recognize that their data do not address other factors associated with placebo effects ("We reviewed the effect of placebos but not the effect of the patient-provider relationship. We could not rule out a psychological therapeutic effect of this relationship, which may be largely independent of any placebo intervention."). Concerning the broader placebo concept, two recent reviews of randomized controlled trials comparing patients given different expectations or presented with different psychosocial medical behaviors have found that "expectancies are a mechanism for placebo effects [which have] received support across a range of clinical areas in a variety of studies" (Crowe et al. 1999) and that "physicians who adopt a warm, friendly, and reassuring manner are more effective than those who keep consultations

formal and do not offer reassurance" (DiBlasi et al. 2001).

Nonetheless, Hróbjartsson and Gøtzsche's challenge to the narrow meaning of placebo effect needs to be considered seriously. Since Beecher (1965) wrote his seminal article, the idea that inert controls of randomized controlled trials have powerful effects has justified both therapeutic and research uses of placebos (Kaptchuk 1998). Although a substantial body of subsequent research lent support to Beecher's conclusion, his original arguments were weak. Among other issues (see Kienle & Kiene 1996), he did not mention natural history or regression to the mean as alternative explanations for any improvements in placebo-treated patients, a peculiar lapse since Beecher included many studies explicitly considering natural history as the cause of improvement in the placebo group (Kienle & Kiene 1997). To Beecher's credit, he enhanced his argument for the importance of inert controls by focusing on psychological studies of placebos. For example, he cited Wolf's demonstration (1950) that sugar pills given with explicit expectation "instructions" could be more effective than drugs, or even completely reverse the known effect of drugs. Despite the weaknesses of Beecher's original paper, the idea of a powerful placebo in therapy and research became entrenched.

This paper addresses shortcomings of Hróbjartsson and Gøtzsche's challenge. First, we review evidence supporting robust placebo effects that are not included in their meta-analysis. Second, we examine methodological problems that affect their conclusions.

Evidence for a powerful placebo effect

Important evidence for a placebo effect comes from three areas of research: clinical trials comparing more than one placebo, trials comparing outcomes from the same placebo in different contexts, and studies investigating the mechanisms by which placebos work.

Discussion

Greene et al.

A. Comparisons of different placebos

Hróbjartsson and Gøtzsche's meta-analysis compares placebo arms in randomized controlled trials with "no-treatment" arms.* Another approach to examining placebo effects in controlled trials is having one type of placebo control for another type of placebo. If placebo effects are nothing beyond natural history or regression toward the mean, the effects of two types of placebos should be similar.

Several clinical trials, reviewed by Kaptchuk et al. (2000), have prospectively compared two different types of placebo (for example, saline injection and sugar pill) to see whether there is a differential placebo effect. In an early study of hypertension that prospectively compared the effects of injected vs oral placebo, injected placebo produced significantly lower blood pressure than oral placebo (Grenfell et al. 1961). Four other trials similarly suggested that device placebos in varicose veins and osteoarthritis were superior to oral placebos (Kaptchuk et al. 2000). Although these trials have methodological shortcomings, they demonstrate that different *routes* of placebo administration produce different magnitudes of placebo effect.

The results of these prospective trials are supported by a recent meta-analysis examining 22 trials for migraine comparing the effects of injected placebo vs oral placebo. Placebo injection produced significantly greater relief than placebo pills (de Craen et al. 2000). Similarly, another meta-analysis found that the magnitude of placebo effect differed with the number of placebo pills administered. In 51 trials for duodenal ulcers (de Craen et al. 1999), the healing rate of patients receiving placebo medication four times a day was significantly greater than for patients receiving placebos twice a day.

These studies show that the magnitude of placebo effects varies with different dosages and methods of administration. If placebos have no effect beyond natural history or regression toward the mean, changing the type of placebo should make no difference.

B. Same placebo under different conditions

Other randomized controlled trials have prospectively looked at how differential awareness

or knowledge among practitioners or patients affected responses to the same placebo. Gracely and colleagues (1985) randomized dental patients into two groups: patients in the first group received placebo, narcotic analgesic, or narcotic antagonist; the second group received either placebo or narcotic antagonist, with no possibility of receiving narcotic analgesic. Treating dentists knew the group assignment of each patient, but remained blind to the actual medication individual patients received. Pain experienced by placebo recipients was significantly worse in the second group (with no possibility of receiving a narcotic analgesic) than in the first group (which had the possibility of receiving analgesic), suggesting that physician knowledge without explicit communication could affect patient outcomes. Another trial of physician expectation on hypertension drugs similarly found that variations in practitioner belief led to different responses to the same placebo (Shapiro et al. 1954).

Dahan et al. (1986) randomized patients with insomnia into two groups. One group and their nurse thought they were in a study comparing a new hypnotic benzodiazepine drug to placebo. The other group was not informed that they were in a study and were treated as if they were routine patients. Both groups received only placebo. The placebo hypnotic activity was significantly higher in the *uninformed* group. Another experiment examined whether knowledge of the possibility of receiving a placebo changed the magnitude of the effect on cancer pain. In this instance, the placebo effect was significantly higher with *informed* consent compared with no informed consent (Bergmann et al. 1994), an outcome in a different direction from the preceding study, but nevertheless showing different responses to the same placebo.

Furthermore, between 1969 and 1987, numerous asthma studies have demonstrated that different instructions to subjects inhaling the same placebo saline can have dramatic, and reversible, effects in either positive or negative directions (examples include Butler & Steptoe 1986; Godfrey & Silverman 1973; Luparello et al. 1968; McFadden et al. 1969; Neild & Cameron

*The majority of the original trials focused on comparisons with treatment arms, without explicit hypotheses concerning placebo/no-treatment comparisons.

1985; Pastorello et al. 1987; Spector et al. 1976). For example, inhalation of saline mist can result in bronchoconstriction or bronchodilation depending on the accompanying suggestion, and the outcome is reversible with a subsequent suggestion of the opposite effect.

Numerous other experiments have shown that labeling the placebo with different expectations causes different placebo effects (Kaptchuk 2001). To repeat, if the placebo effect was only natural history or regression to the mean, these differential effects should not have occurred.

C. Mechanisms of placebo effects

Recent studies aimed at elucidating the neurophysiological mechanisms underlying placebo responses further support the existence of placebo phenomena. This research is exemplified by a number of elegant studies by Benedetti and colleagues testing the hypothesis that endogenous opioids play a central role in placebo analgesia.

In one study of experimentally induced ischemic arm pain (Amanzio et al. 2001), subjects received intravenous injections of the nonopioid analgesic ketorolac, with and without naloxone (an opioid blocker), in either an open or hidden fashion. Subjects receiving open injections of ketorolac who knew they were receiving a pain killer reported more pain relief than did those receiving hidden injections of the same drug. However, when the opioid blocker naloxone was added to the open injections of ketorolac, their pain levels were reduced to the same level as those receiving hidden injections. This suggests the placebo analgesic effect of visible injection is mediated by endogenous opioids. A parallel, clinical experiment reported in the same paper compared the effectiveness of hidden vs open injections of four analgesics (including ketorolac) on patients experiencing postsurgical pain. Again, in all cases, open injections were significantly more effective than hidden ones. Amanzio et al.'s experiments clearly support the existence of placebo analgesia as well as the hypotheses proposed by earlier researchers that placebo analgesia is mediated, at least in part, by endogenous opioids (Grevert et al. 1983; Levine et al. 1978; Levine & Gordon 1984; critically reviewed by ter Riet et al. 1998).

Another study conducted by this research group demonstrated that opioid-mediated placebo analgesia could be spatially directed (Benedetti et al. 1999). In this study, pain was produced simultaneously in four limbs by intradermal capsaicin injections. Placebo or analgesic cream was applied to only one or two of four limbs in a double-blind fashion. Subjects consistently reported reduced pain on sites treated with the placebo cream; however, these placebo effects were then completely antagonized by naloxone. These results suggest that expectation of anatomically localized analgesia may result in selective release of opioids with ensuing modulation of only those signals of pain that originate from the specified area (also see Montgomery & Kirsch 1996).

Research by Benedetti's group also illustrates progress in understanding neurophysiological mechanisms underlying placebo phenomena unrelated to pain. For example, Benedetti et al. (1999) reported that respiratory depression commonly following administration of narcotics could be elicited with a placebo following conditioning with the opioid agonist buprenorphine. This placebo respiratory depressant effect was then completely blocked by naloxone, indicating that, like placebo analgesia, this response is mediated by endogenous opioids. However, unlike studies of placebo analgesia in which patients are required to cognitively assess pain, subjects in this study were completely unaware of their breathing patterns.

Together, these studies demonstrate that under controlled experimental conditions, specific responses to placebo can be elicited by either visual suggestion or conditioning, and then blocked by a specific antagonist.

A growing body of literature from the field of mind-body research also supports the existence of placebo phenomena and sheds light on the mechanistic processes and pathways underlying them. Indeed, some researchers have suggested that placebo treatments may be seen as types of mind-body interventions, and pathways mediating effects of mind-body interventions may also mediate placebo effects (Baime 1999; Stephano et al. 2001). For examples, research from the fields

Discussion

Greene et al.

of behavioral medicine (Schrodt & Tasman 1999), biofeedback (Green & Shellenberger 1999), hypnosis (Wickramasekera 1999), and meditation (Baime 1999) supports the hypothesis that “psychosocial” elements such as expectation, belief, and anxiety reduction affect healing and maintenance of health. For some physiological processes such as immune and cardiovascular functions, a number of the neurological mechanisms linking mind and body have been elucidated (Stephano et al. 2001; Watkins 1997).

In summary, mechanistic studies as well as the broader area of mind–body medical research not only support the existence of placebo effects, but also have begun to elucidate tangible physiological pathways that mediate these effects. These studies anchor placebo effects in biological reality.

Methodological issues

Meta-analysis provides a powerful tool for combining information across a range of studies; however, it is not without limitation (Hedges & Olkin 1985; Rosenthal & DiMatteo 2001). Below we

* It has not escaped our attention that there is a fifth point, concerning an apparent trend among binary subgroup outcomes, that appears initially to contradict Hróbjartsson and Gøtzsche’s conclusions. Although none of the subgroup pooled relative risks are statistically significant, there appears to be a definite trend of beneficial effect of placebo. If placebo were truly equivalent to no treatment, we would expect that in some subgroups placebo would appear to be superior to no treatment, while in other subgroups no treatment would appear to be superior to placebo, rather than such a one-sided pattern. However, on further investigation, the 32 studies listed in the data set submitted to the *New England Journal of Medicine* do not contradict their conclusions, after all. (But see footnote in the right hand column of this page) On the other hand, their conclusion that objective continuous outcomes did not exhibit placebo effects raises a related “trend” issue. In Table 1, the 95% confidence interval is (–0.27 to 0.03). Given the overall pattern for continuous outcomes, this would seem an appropriate instance for Rothman’s concern regarding “significance questing” (1986) when there is a “close call.” The binary outcomes (overall, subjective, and objective) are similarly “close calls.”

†The reader is not given a reason for the preference for binary data. We also do not know how many of Hróbjartsson and Gøtzsche’s own selections were binary or continuous, or whether there were continuous outcomes possible when they selected binary.

discuss four methodological problems that cast doubt on Hróbjartsson and Gøtzsche’s analyses and conclusion: their use of binary and continuous outcomes; selection of studies; publication bias; and the weakness of meta-analysis itself in this context.*

A. Binary vs continuous outcomes

The authors selected “the main objective or subjective outcome of each trial.” When a main outcome was not available in an original study (38 cases), they “used the outcome that [they] felt was most relevant to patients.” They explain that “binary outcomes (e.g. the proportions of smokers and nonsmokers) were preferred to continuous ones (e.g. the mean number of cigarettes smoked).”† This preference may have affected their results for three reasons.

First, the studies with binary outcomes—the type Hróbjartsson and Gøtzsche preferred—did not demonstrate statistically significant placebo effects whereas a large subset of studies with continuous outcomes did.** Further, some “binary” outcomes are not necessarily true dichotomies as is, say, “death/survival.” Of the 32 binary outcomes, many reflect underlying continuous outcomes such as nausea, depression, or pain, compared with an outcome such as fertility, an aspect of a more reasonable dichotomy. Had such variables been included as continuous (that is, amount of improvement in pain, depression, or psychological adjustment to seizure disorders), we cannot know if they would have exhibited results similar to the 82 continuous outcome studies: that is, a significant placebo effect of the pooled continuous studies, and a significant placebo effect for the pooled subset of subjective continuous studies. (Of course, if the original studies used in the meta-analysis provided only dichotomous outcomes underlying

** The power of binary/binomial/proportion-based tests is weaker than for difference of means-based tests. Thus, given the smaller number of binary studies ($n=32$) than continuous ($n=82$), it may be a statistical artifact of the smaller number of studies in cases like this that has led to the lack of overall significant outcome in the binary group. (See also the comment on Rothman’s “significance questing” in footnote* in the left hand column of this page.)

possibly continuous outcomes, then Hróbjartsson and Gøtzsche had no choice but to categorize them that way.)

Second, not using continuous data when available leads to loss of information, such as the magnitude of outcomes within and between groups. Consider their own example distinguishing between binary and continuous outcomes: cigarette smoking. Imagine, hypothetically, smokers randomized to two groups—placebo (perhaps a placebo nicotine patch, or a “counseling” session unrelated to conventional smoking-cessation programs) and no-treatment—so that each group has a reasonably similar distribution of “numbers of cigarettes smoked.” At the conclusion of the trial, the placebo group might have a significantly lower mean number of cigarettes smoked than at baseline (perhaps as dramatically as a 50+% average reduction). However, suppose not a single person in this “improved” placebo group stopped smoking entirely. Assume further that the number of cigarettes smoked by the no-treatment group remained the same. Drawing conclusions using the binary outcome, comparing proportions of smokers and nonsmokers, both groups began with 100% smokers and both ended with 100% smokers. “Therefore” the placebo was no different than no-treatment at all. *Q.E.D.* However, using the continuous outcome (the mean), the placebo group was, indeed, significantly “better” than no treatment at all. *Q.E.D.* (again!) In fact, varying the choice of such statistical approaches can in some cases actually yield opposite effects. Indeed, the choice of

measures and statistical procedures used can affect the outcome dramatically (an analytic example of published data is given by Greene 1977). Thus, it is at least possible that using binary rather than continuous outcomes, when there was a choice, could have affected the results,* as suggested above.

Third, the *choice* of outcome selected can dramatically affect conclusions. The authors’ “choice” of the outcome they “felt was most relevant to patients” when there was not a single “main” outcome in the original study is certainly subjective, with possible biases, however subconscious or inadvertent. Indeed, the recognition of precisely these types of subconscious biases led to the layers of blinding in research. Related to this is their preference for binary outcomes rather than continuous. If the original researchers of a study identified two (or more) outcomes and at least one (but not all) was binary, it would be possible to “view” a binary outcome as “most relevant,” or even as “the” primary goal of the research. Their choice of the word “preferred” implies precisely such a subjective decision, conscious or not.†

B. Selection criteria for studies

The manner in which Hróbjartsson and Gøtzsche excluded studies with high dropout rates may mask important differences in dropout rates between arms. They excluded studies if the overall dropout rate exceeded 50%. However, an overall dropout rate of 50% may nevertheless mask very differing dropout rates in subgroups: verum, placebo, and no-treatment arms. Quite apart from the argument that to be relatively consistent with

* This problem is by no means restricted to this study, or even to meta-analytic contexts. It could have similarly affected the outcome of any of the original articles with binary outcomes. For example, in the preceding example, change the phrase “placebo group” to “treatment group” and change the phrase “no treatment group” to “placebo group.” That is the traditional placebo-controlled clinical trial, and the statistical issues of how data are collected, coded, and analyzed can be very much a factor in the subsequent results. Indeed, this issue is directly related to the way Beecher’s 1955 study has been interpreted—or misinterpreted. That is, Beecher reported that approximately 35% of the subjects “treated” with a placebo improved. That has, not infrequently, been equated, inaccurately, with a 35% reduction in the magnitude symptoms.

† Hróbjartsson and Gøtzsche do mention that the “effects of placebo were also unrelated to...whether we had identified the main outcome on the basis of clinical relevance (data not shown).” However, because the very choice of the outcome is so crucial (depending on a variety of factors such as how the variables were coded, what cut-off points were used, what choices were available, and what discretion they had), it is of more than passing curiosity how their choices were made and how this lack of effect was investigated.

Discussion

Greene et al.

the inclusion/exclusion criteria, no single arm should have a dropout rate exceeding 50%, it would be exceedingly interesting to know the dropout rates for each arm separately. Comparing the dropout rates for the verum, placebo, and no-treatment groups might uncover a pattern suggesting that the subjects were indeed responding differently to placebo compared with no treatment.*

Further, examining both the pattern of dropouts for the three arms and the reasons given (when available) would be worthy of more study. For example, if the no-treatment subjects in the worst condition were disproportionately likely to drop out (perhaps to seek some kind of treatment elsewhere), then those remaining in the no-treatment group would, of course, be those who were doing relatively well. This may be less of a factor for those in a placebo group, as they have reason to think that they might be receiving a real treatment, and may therefore be willing to wait for it to “work.” Therefore, in some cases, the final comparison between the placebo and no-treatment groups might be comparing most of the placebo group with the “least sick” of the no-treatment group, thus attenuating, or even eliminating or reversing, any measured placebo effect.†

Additionally, the 114 studies are analyzed as if they are all independent cases. While Hróbjartsson and Gøtzsche avoided more than one publication on the same research (29 studies were identified as having been published more than once and duplicates were excluded), avoiding duplicate publications from the same studies, or even from the same authors and/or research facilities, does not assure independence. Specifically, studies of

the same condition/disease/treatment/placebo are reasonably more similar than those with different conditions/diseases/treatments/placebos. Selecting more than one study from each such group could violate assumptions of independence: differing numbers of studies for such groups could disproportionately weight the “average” outcomes. (The number per condition ranges from 1 to 29; for example, the “pain” studies included 27 continuous and 2 binary outcomes.) Corrections for possible relationships and disproportionalities of this sort have been created (for example, Tomberlin et al. 1981), and the conclusions with and without such corrections can differ dramatically.

Publication bias

Regarding the question of whether publication bias is a problem in all meta-analyses: in an ordinary meta-analysis (looking at outcomes of a verum treatment), publication bias would tend to skew the meta-analysis results in a positive direction because trials with positive results for the verum treatment, and thus relatively less of an effect in the placebo group, are the ones that tend to get published. However, in a placebo meta-analysis, publication bias would tend to skew the results in a negative direction because trials with more positive results for the placebo would tend not to get published.

Weakness of meta-analysis used for placebo

Meta-analysis has frequently become used for comparisons of multiple studies of similar medical outcomes, such as medications for duodenal ulcers or migraine, with similar or differing dosages or administration methods across a range of trials. However, meta-analyses have been criticized when used to summarize results drawn from heterogeneous trials using widely varying methods and outcomes (for example, Hedges & Olkin 1985; Rosenthal & DiMatteo 2001). This criticism may be especially apropos for a meta-analysis of “placebo” and “no-treatment” combining treatments for 40 different conditions.

* This would be similar to a study comparing the dropout rates for treatment and placebo groups (Rochon et al. 1999).

† This effect can also occur in regular clinical trials. Again, substitute “treatment” for “placebo” and “placebo” for no-treatment.” If subjects in the placebo arm who are the sickest or doing the “worst” drop out at a higher rate than others, then those who remain in the placebo arm are, by definition, doing relatively well. Therefore, in some cases, the final comparison between the two groups might be comparing most of the treatment group with the “least sick” of the placebo group, thus attenuating, or even eliminating or reversing, any measured verum effect.

Placebos are meant to mimic a verum therapy, and are likely to be far more variable than verum treatments: for each verum treatment, there are many—and diverse—possible placebos. In Hróbjartsson and Gøtzsche’s meta-analysis, there are in fact two sources of variability: variable underlying medical conditions, and variable placebos for each verum treatment per condition. Increased variability makes it more difficult to detect differences that do exist, assuming constant sample size.*

The possibility that this variability across the different randomized controlled trial placebo conditions may be affecting the study’s results is supported by Moerman’s (2000) meta-analysis of 117 trials in 44 countries of two very similar drugs designed to treat ulcers, Cimetidine and Ranitidine. He found that a mean of 35% of patients had their ulcers healed, but the rates varied from 0 to 100% in these studies. A particularly striking finding was that “the placebo healing rate in 6 German studies averaged 59%, twice as high as in the rest of the world ($P=0.00018$) and three times that of two of its neighboring countries, Denmark and the Netherlands ($P=0.011$).” Moerman’s conclusion reinforces the argument that placebo effects may be highly dependent on their local context—the larger range of “non-specific” effects such as patient-physician relationship and patients’ hopes, expectations, and anxieties, which the authors have excluded from their meta-analysis. The heterogeneity in such a wide range of trials may make such a separation untenable.

* Indeed, the authors do find significant heterogeneity among the binary outcomes ($P=0.003$) and among the continuous outcomes ($P<0.001$).

† The actual conditions listed by Hróbjartsson and Gøtzsche were: “hypertension, asthma, anemia, hyperglycemia, hypercholesterolemia, seasickness, Raynaud’s disease, alcohol abuse, smoking, obesity, poor oral hygiene, herpes simplex infection, bacterial infection, common cold, pain, nausea, ileus, infertility, cervical dilatation, labor, menopause, prostatism, depression, schizophrenia, insomnia, anxiety, phobia, compulsive nail biting, mental handicap, marital discord, stress related to dental treatment, orgasmic difficulties, fecal soiling, enuresis, epilepsy, Parkinson’s disease, Alzheimer’s disease, attention-deficit-hyperactivity disorder, carpal tunnel syndrome, and undiagnosed ailments.”

Imagine a meta-analysis of clinical trials intended to determine the “general efficacy of verum treatments” combining treatments for 40 different conditions such as diabetes, anxiety, smoking cessation, cancer, and both bacterial and viral infections, with possibly differing treatments per condition.† How much more meaningful is a meta-analysis of clinical trials intended to determine the “general effect of placebo treatments” under a similarly heterogeneous collection of conditions, with possibly differing placebos per condition and per verum treatment?

Conclusion

Hróbjartsson and Gøtzsche’s provocative study will undoubtedly shift future conceptual and evidentiary discussions concerning placebo effects. We hope that it generates continuing debate and renewed interest in examining the placebo concept. However, results from other areas of research, and methodological limitations in their meta-analysis, raise serious doubts about their findings.

Our future understanding of placebo phenomena would be greatly enhanced through more rigorous standards in clinical trial design. If the following four practices were employed in trials testing verum interventions, data acquired from the placebo arms would become much more meaningful:

- inclusion of a third, “no-treatment” arm where ethically appropriate;
- questioning of all patients to determine if they received treatments other than the trial intervention (“no treatment” patients may be more likely to seek other therapies, in which case they would no longer truly be “no treatment”);
- consideration of how closely a placebo mimics the verum treatment;
- checking for the unblinding of patients (that is, asking the patients in verum and placebo arms whether they believe they received the verum or placebo treatment).

Trials including these four components allow for more definitive conclusions regarding the efficacy of verum treatments as well as more

Discussion

Greene et al.

rigorous analyses of placebo effect. If this approach became common practice, a much stronger meta-analysis of placebo effects would be possible in a few years' time.

Acknowledgment

We would like to thank Akemi Joy Tanaka and Jacqueline B. Savetsky for assistance.

References

- Amanzio M, Pollo A, Maggi G, Benedetti F 2001 Response variability to analgesics: A role for nonspecific activation of endogenous opioids. *Pain* 90:205–215
- Baime MJ 1999 Meditation and mindfulness. In: Jonas WB, Levin JS, eds. *Essentials of Complementary and Alternative Medicine*. Philadelphia: Lippincott Williams & Wilkins
- Beecher HK 1955 The powerful placebo. *Journal of the American Medical Association* 159(17):1602–1606
- Benedetti F, Amanzio M, Baldi S, Casadio C, Maggi G 1999 Inducing placebo respiratory depressant responses in humans via opioid receptors. *European Journal of Neuroscience* 11:623–631
- Benedetti F, Arduino C, Amanzio M 1999 Somatotopic activation of opioid systems by target-directed expectations of analgesia. *Journal of Neuroscience* 19: 3639–3648
- Bergmann JF, Chassany O, Gandiol J et al. 1994 A randomized clinical trial of the effect of informed consent on the analgesic activity of placebo and naproxen in cancer pain. *Clinical Trials Metaanalysis* 29(1):41–47
- Butler C, Steptoe A 1986 Placebo responses: An experimental study of psychophysiological processes in asthmatic volunteers. *British Journal of Clinical Psychology* 25:73–183
- Crowe R, Gage H, Hampson S, Hart J, Kimber A, Thomas H 1999 The role of expectancies in the placebo effect and their role in the delivery of health care: A systematic review. *Health Technology Assessment* 3: 1–98
- Dahan R, Caulin C, Figea L, Kanis JA, Caulin F, Segrestaa JM 1986 Does informed consent influence therapeutic outcome? A clinical trial of the hypnotic activity of placebo in patients admitted to hospital. *British Medical Journal* 293(6543):363–364
- De Craen AJM, Tijssen JGP, de Gans J, Kleijnen J 2000 Placebo effect in the acute treatment of migraine: subcutaneous placebos are better than oral placebos. *Journal of Neurology* 247:183–188
- De Craen AJM, Moerman DE, Heisterkamp SH, Tytgat GNJ, Tijssen JGP, Kleijnen J 1999 Placebo effect in the treatment of duodenal ulcer. *British Journal of Clinical Pharmacology* 48:853–860
- DiBlasi Z, Harkness E, Ernst E, Georgiou A, Kleijnen J 2001 Influence of context effects on health outcomes: A systematic review. *Lancet* 357:757–762
- Durham M 2001 Placebo effect is all in the doctor's mind. *The Independent (London)*. May 25:3
- Godfrey S, Silverman M 1973 Demonstration by placebo response in asthma by means of exercise testing. *Journal of Psychosomatic Research* 17:293–297
- Gracely RH, Dubner R, Deeter WR, Wolskee PJ 1985 Clinicians' expectations influence placebo analgesia. *Lancet* 1:43
- Green JA, Shellenberger 1999 Biofeedback therapy. In: Jonas WB, Levin JS, eds. *Essentials of Complementary and Alternative Medicine*. Philadelphia: Lippincott Williams & Wilkins
- Greene PJ 1977 Culture and genetic variability. *Nature* 267:375
- Grenfell RF, Briggs AH, Holland WC 1961 Double blind study of the treatment of hypertension. *Journal of the American Medical Association* 176:124–128
- Grevert P, Albert LH, Goldstein A 1983 Partial antagonism of placebo analgesia by naloxone. *Pain* 16:129–143
- Grossman KN 2001 Placebos don't do much good after all. *The Chicago Sun Times*, May 24:30
- Hedges LV, Olkin I 1985 Estimation of a single effect size: parametric and nonparametric methods. *Statistical methods for meta-analysis*. Academic Press
- Hróbjartsson A, Gøtzsche PC 2001 Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 344(21): 1594–1602
- Kaptchuk TJ 1998 Powerful placebo: The dark side of randomized controlled trials. *Lancet* 344:995–998
- Kaptchuk TJ 2001 The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology* 53:786–792
- Kaptchuk TJ, Goldman P, Stone DA, Stason WB 2000 Do medical devices have enhanced placebo effects? *Journal of Clinical Epidemiology* 53: 786–792
- Kienle GS, Kiene H 1996 Placebo effect and placebo concept: A critical methodological and conceptual analysis of reports on the magnitude of the placebo effect. *Alternative Therapies in Health and Medicine* 2(6):39–54
- Kienle GS, Kiene H 1997 The powerful placebo effect: Fact or fiction? *Journal of Clinical Epidemiology* 50(12):1311–1318
- Levine JD, Gordon NC 1984 Influence of the method of drug administration on analgesic response. *Nature* 312:755–756
- Levine JD, Gordon NC, Fields HL 1978 The mechanism of placebo analgesia. *Lancet* 2:654–657
- Luparello T, Lyons HA, Bleecker ER, McFadden ER 1968 Influences of suggestion on airway reactivity in asthmatic subjects. *Psychosomatic Medicine* 30: 819–825
- McFadden ER, Luparello T, Lyons HA, Bleecker E 1969 The mechanism of action of suggestion in the induction of acute asthma attacks. *Psychosomatic Medicine* 31:134–143

- Moerman DE 2000 Cultural variations in the placebo effect: Ulcers, anxiety, and blood pressure. *Medical Anthropology Quarterly* 14(1):51–72
- Montgomery GH, Kirsch I 1996 Mechanisms of placebo pain reduction: An empirical investigation. *Psychological Science: A Journal of the American Psychological Society* 7:174–176
- Neild JE, Cameron IR 1985 Bronchoconstriction in response to suggestion: Its prevention by an inhaled anticholinergic agent. *British Medical Journal* 290:674
- Pastorello EA, Codecasa LR, Pravettoni V et al. 1987 The role of suggestion in asthma. Effects of inactive solution on bronchial reactivity under bronchoconstrictor or bronchodilator suggestion. *Annals of Allergy* 59:336–338
- Rochon PA, Binns MA, Litner JA et al. 1999 Are randomized control trial outcomes influenced by the inclusion of a placebo group? A systematic review of nonsteroidal anti-inflammatory drug trials for arthritis treatment. *Journal of Clinical Epidemiology* 52(2):113–122
- Rothman K 1986 Significance questing. *Annals of Internal Medicine* 154(3):445–447
- Rosenthal R, DiMatteo MR 2001 Meta-Analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology* 52:59–82
- Schrodt GR, Tasman A 1999 Behavioral medicine. In: Jonas WB, Levin JS, eds. *Essentials of Complementary and Alternative Medicine*. Philadelphia: Lippincott Williams & Wilkins
- Shapiro AP, Myer T, Reiser MF, Ferris EB 1954 Comparison of blood pressure response to Verloid and to the doctor. *Psychosomatic Medicine* 16:478–488
- Spector S, Luparello TJ, Kopetzky MT, Souhrada J, Kinsman RA 1976 Response to asthmatics to methacholine and suggestion. *American Review of Respiratory Disease* 113:43–50
- Stephano GB, Fricchione GL, Slingsby BT, Benson H 2001 The placebo effect and relaxation response: Neural processes and their coupling to constitutive nitric oxide. *Brain Research Reviews* 35:1–19
- ter Riet G, de Craen AJM, de Boer A, Kessels AGH 1998 Is placebo analgesia mediated by endogenous opioids? A systematic review. *Pain* 76:273–275
- Tomberlin TJ, Greene PJ, Hartung J 1981 A correction for diffusion. *The Proceedings of the American Statistical Association*
- Watkins A 1997 *Mind–Body Medicine: A Clinician’s Guide to Psychoneuroimmunology*. New York: Churchill Livingstone
- Wickramasekera I 1999 Hypnotherapy. In: Jonas WB, Levin JS, eds. *Essentials of Complementary and Alternative Medicine*. Philadelphia: Lippincott Williams & Wilkins
- Wolf S 1950 Effect of suggestion and conditioning on the action of chemical agents in human subjects—the pharmacology of placebos. *Journal of Clinical Investigation* 29:100–109

Apples, oranges, and placebos: Heterogeneity in a meta-analysis of placebo effects

Irving Kirsch and Alan Scoboria

Irving Kirsch PhD is a professor of psychology at the University of Connecticut, where he does research on placebo effects, hypnosis, and the effects of suggestion on memory. Alan Scoboria is a graduate student in clinical psychology at the University of Connecticut where he also does research on placebo effects, hypnosis, and the effects of suggestion on memory.

In a meta-analysis of heads-on comparisons of placebo versus no treatment, Hróbjartsson and Gøtzsche (2001) reported a significant placebo effect of 0.28 standard deviations (SD) in studies reporting continuous outcome scores and a nonsignificant effect in studies reporting dichotomous outcomes. Contrary to the conclusions of the authors, this indicates that the placebo effect is significant but may be attenuated when actual outcomes are condensed into binary categories (for example, cured versus not-cured). The most reliable finding in the Hróbjartsson and Gøtzsche meta-analysis, however, was the substantial and significant heterogeneity in the outcomes produced by placebo, regardless of how those outcomes were assessed. In other words, some of the placebos were significantly more effective than others. This, in itself, validates the existence of a placebo effect. One placebo cannot be more effective than another unless placebos are capable of producing an effect.

The heterogeneity of outcomes reported by Hróbjartsson and Gøtzsche (2001) is not surprising. The magnitude of the placebo effect depends on many factors. One of these is the condition being treated. Placebos have been shown

Discussion

Kirsch and Scoboria

to have substantial effects on depression (Kirsch & Sapirstein 1999), asthma (Sodergren & Hyland 1999) and phobic anxiety (Kirsch et al. 1983), for example, but they may have little or no effect on many other conditions. Thus, an estimate of the mean placebo effect size for clinical disorders as diverse as colds, depression, Raynaud's disease, infertility, schizophrenia, orgasmic potential, epilepsy, enuresis and nicotine addiction is not very meaningful.

Another factor affecting the magnitude of the placebo effect is the type of placebo being used. It has been shown, for example, that placebo injections are more effective than placebo pills (Traut & Passarelli 1957), and that placebo morphine is more effective than placebo aspirin (Evans 1974). It especially does not make sense to lump together "pharmacologic (e.g. a tablet), physical (e.g. a manipulation) and psychological (e.g. a conversation)" placebos (Hróbjartsson & Gøtzsche 2001). So-called "psychological placebos" (that is, psychological procedures used as controls in psychotherapy outcome studies) are an extremely heterogeneous set of procedures. They include, for example, listening to stories, reading books, attending language classes, viewing films, participating in "bull" sessions, playing with puzzles, sitting quietly with a silent therapist, and discussing current events (see Prioleau et al. 1983). Indeed, simply being placed on a waiting list has been labeled a placebo in some studies (for example, Sloane et al. 1975). The effectiveness or ineffectiveness of procedures of this sort tell us nothing about the effectiveness of placebos as used in clinical trials of medication.

Psychological placebos vary greatly in credibility. Some of them are very believable to patients, but others are not, and it has been demonstrated empirically that credible placebos are significantly more effective than less credible placebos, at least in the treatment of phobic anxiety (Kirsch & Henry 1977; McReynolds et al. 1973). Thus, the inclusion of unbelievable control procedures is likely to have attenuated Hróbjartsson and Gøtzsche's estimate of the placebo effect and contributed to the significant heterogeneity that they found.

It makes no sense to evaluate the magnitude of the placebo effect in general, as Hróbjartsson and Gøtzsche have attempted to do. The definition of placebo "as an intervention labeled as such in the report of a clinical trial" (Hróbjartsson & Gøtzsche 2001) is inadequate, and its use as a criterion for including studies in a meta-analysis produces misleading results. Instead of assessing the placebo effect in general, one needs to assess the magnitude of particular types of placebos for particular conditions.

Despite the inclusion of studies with atypical and incredible placebos for the treatment of disorders that may be especially insensitive to placebo effects, it is surprising that Hróbjartsson and Gøtzsche found as small a placebo effect as they did. What is particularly surprising about this result is its inconsistency with previous meta-analyses of the effects of placebos. Smith et al. (1980) reported an effect size of 0.56 SD for psychological placebo compared to no-treatment. Andrews and Harvey (1981) reanalyzed these data but excluded trials with healthy volunteers: they reported an effect size of 0.55 SD. Shapiro and Shapiro (1982) restricted their meta-analysis to published reports in which two or more different treatments were compared in the same study—the same criterion used by Hróbjartsson and Gøtzsche—and reported an effect size of 0.71 SD for placebo treatment. Kirsch and Sapirstein (1999) compared placebo pills to no-treatment for depression and reported an effect size of 0.79. These are substantially greater than the effect sizes reported by Hróbjartsson and Gøtzsche.

Not only did Hróbjartsson and Gøtzsche fail to address the discrepancy between their study and previous meta-analyses of the placebo effect, they did not even cite the prior research. They also seem to have missed some of the studies analyzed in those meta-analyses, including the seminal study in which the effects of a psychological placebo were assessed and found to be greater than those of no-treatment (Paul 1966). They also failed to include subsequent studies showing an even greater placebo effect when more credible psychological placebos were used (for example, Kirsch & Henry 1977; Kirsch et al. 1983).

Credible placebos can produce powerful effects on some conditions, incredible placebos have much smaller effects, and placebos of any kind have no effect on some conditions. Hence the significant heterogeneity in effect sizes that Hróbjartsson and Gøtzsche found in their meta-analysis. The overall effect sizes they reported are arbitrary. They depend on the nature of the placebos that were used and on the kinds of conditions that were treated in the studies that were included in their meta-analysis. In fact, most of the studies they analyzed did not include conventional placebos (that is, placebo pills) at all. Thus, they tell us nothing about the power of placebos as used in clinical trials.

References

- Andrews G & Harvey R 1981 Does psychotherapy benefit neurotic patients? A reanalysis of the Smith, Glass and Miller data. *Archives of General Psychiatry* 36:1203–1208
- Evans FJ 1974 The placebo response in pain reduction. In: Bonica JJ, ed. *Advances in Neurology*, Vol 4. Pain New York: Raven
- Hróbjartsson A, Gøtzsche PC 2001 Is the placebo harmless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 344:1594–1602
- Kirsch I, Henry D 1977 Extinction vs credibility in the desensitization of speech anxiety. *Journal of Consulting and Clinical Psychology* 45:1052–1059
- Kirsch I, Sapirstein G 1999 Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication. In: Kirsch I, ed. *How expectancies shape experience*. Washington, DC: American Psychological Association
- Kirsch I, Tenen H, Wickless C, Saccone AJ, Cody S 1983 The role of expectancy in fear reduction. *Behavior Therapy* 14:520–533
- McReynolds WT, Barnes AR, Brooks S, Reagen NJ 1973 The role of attention-placebo influences in the efficacy of systematic desensitization. *Journal of Consulting and Clinical Psychology* 41:86–92
- Paul GL 1966 *Insight vs. desensitization in psychotherapy*. Stanford, CA: Stanford University Press
- Prioleau L, Murdock M, Brody N 1983 An analysis of psychotherapy versus placebo studies. *Behavioral and Brain Sciences* 6:275–310
- Shapiro DA, Shapiro D 1982 Meta-analysis of comparative therapy outcome studies: a replication and refinement. *Psychological Bulletin* 92:581–604
- Sloane RB, Staples FR, Cristol AH, Yorkston NJ, Whipple K 1975 *Psychotherapy versus behavior therapy*. Cambridge, MA: Harvard University Press
- Smith ML, Glass GV, Miller TI 1980 *The benefits of psychotherapy*. Baltimore, MD: Johns Hopkins University Press
- Sodergren SC, Hyland ME 1999 Expectancy and asthma. In: Kirsch I, ed. *How expectancies shape experience*. Washington, DC: American Psychological Association
- Traut EF, Passarelli EW 1957 Placebos in the treatment of rheumatoid arthritis and other rheumatic conditions. *Annals of the Rheumatic Diseases* 16:18–22

The placebo efficacy study: Problems with the definition of the placebo and the mechanisms of placebo efficacy

Ian Wickramasekera

Ian Wickramasekera PhD is Professor of Psychiatry at the Saybrook Graduate School, San Francisco, CA, USA, and Clinical Professor of Psychiatry at Stanford Medical School, Palo Alto, CA, USA.

Hróbjartsson and Gøtzsche (2001) claim to evaluate the empirical efficacy of “placebo,” but they failed to distinguish between placebo studies that experimentally manipulate placebo mechanisms to amplify placebo efficacy and those that reduce it, and they fail to use a uniform criteria to define and identify placebo studies.

Their study appears to have an important strength because it is based on research that compared a “no-treatment” control group to a placebo control group, which is rarely done in clinical trials. But, as the authors admitted, they did not *verify* that patients in an alleged “no-treatment” control group in fact received no treatment. If the patients in an untreated control group sought treatment outside of the trials more often than did patients in the placebo groups, the effects of placebo would be less apparent, as they

Discussion

Wickramasekera

appear to be in this study. Untreated human patients are not passive chess pieces on a clinical trial chess board unless their passivity can be verified with respect to *not* seeking treatment outside the clinical trial. Patients with aversive clinical symptoms who know they are not receiving any treatment in the clinical trial are likely to seek help outside the clinical trial. Likewise, patients who believe they are receiving only a “placebo treatment” are not likely to take or use it enthusiastically. This study made no effort to determine if the patients in the placebo group even verbally reported that they believed they were receiving a credible therapy, much less did it attempt to verify behavioral compliance with a placebo therapy. Such verification, by at least verbal report of patient belief, has become standard in behavioral science studies of clinical efficacy comparing active therapy interventions with placebo therapies.

As noted, this study failed to pay any attention to the hypothesized theoretical mechanisms that have been empirically shown to amplify the placebo effect (Roberts et al. 1993; Turner et al. 1994; Wickramasekera 1977, 1980, 1985). This is a more serious flaw that goes to the heart of the study’s claims to evaluate the efficacy of the placebo effect. In its definition and its selection of placebo therapy groups, it ignores all the growing literature on hypothesized theoretical mechanisms (patient and therapist subjective belief in the efficacy of the therapy, conditioning and memory mechanisms, etc.) that have been empirically shown to amplify or reduce the efficacy of the placebo effect (Ader 1985, 1989, 1997; Kirsch & Sapirstein 1998; Moerman 1983; Roberts et al. 1993; Siegel 1999; Siegel et al. 1982; Turner et al. 1994; Wickramasekera 1977, 1980, 1985). Studies show a broad range of placebo effects. For example, with regard to clinical pain treated with drugs or surgery, on both objective and subjective measures of efficacy the mean placebo effects range from 20% to 100% (Turner et al. 1994). It seems likely that the explanation for such a range is at least partly due to the involvement of different possible mechanisms that enhance or diminish a placebo’s effect. The failure to consider a placebo’s relationship to

such mechanisms is a major deficiency in scholarship. It is blind empiricism.

In the last 30 years, there have been at least two empirically replicated theoretical mechanisms accounting for placebo effects—first, the degree of patient and therapist belief in the efficacy of the therapy proposed (Roberts et al. 1993); second, the hypothesis that placebo responses can be conditioned (Ader 1985; Ader 1997; Hernstein 1962; Wickramasekera 1977, 1980, 1985) or learned and retained in memory from previous specific therapies. The presence of these two empirically documented mechanisms has been shown to increase the efficacy of the placebo component, even of active drug therapies; their absence reduces the placebo effect. These mechanisms should have been considered in planning any serious study of placebo efficacy.

In regard to belief, Hróbjartsson and Gøtzsche made no effort to determine if the patients and therapists believed that those in the placebo group were receiving an equally credible therapy. This needs to be done routinely if for no other reason than to insure patient participation and compliance with the “placebo therapy” or any other active drug therapy. Turner et al. (1994) have shown that compliance is likely to be an important factor of placebo studies, and it clearly is in testing any active drug. Roberts et al. (1993) review of five treatments (drug and surgical) that have since been abandoned (due to the results of blinded controlled randomized clinical trials) involving 6931 patients, found that when both the patient and therapist believed in the efficacy of the therapy, the mean placebo rate was 70%.

Further, it has been shown that the empirical efficacy of a placebo compared to an active therapy in drug depression studies depends on whether the placebo is “active” (produces side-effects like the active drug therapy) or inert (Greenberg & Fisher 1997). It appears that in drug therapy depression studies the efficacy of the placebo is related to the magnitude of placebo-produced side-effects (Kirsch & Sapirstein 1998). Patients who do not get “drug side-effects” come to believe quickly that they are on a placebo and not receiving the active therapy.

It has also been shown by Kirsch and Saperstein (1998), in a meta-analysis of antidepressants and other drugs, that the more powerful the drug the more powerful the placebo. This meta-analysis of 19 double-blind studies ($n=2318$) found that the magnitude of the placebo effect was highly related ($r=0.90$) to the magnitude of the drug effect. Evans (1974) had previously reached a similar conclusion from an analysis of six double-blind studies of other medications.

Hróbjartsson and Gøtzsche claim that patients in a placebo group would think they had received treatment, but they do not provide any direct or indirect evidence to support this minimal but crucial claim of patient belief. Actual clinical practice *always* involves treatments that therapists have reason (based on theory or clinical observation) to believe are effective and patients believe are likely to help them. It is known that simply telling patients that they are in a clinical trial in which they may get a placebo therapy under double-blind conditions can cut in half the efficacy of a placebo therapy (Cohen et al. 1977). Hróbjartsson and Gøtzsche even report a large study (Schulz et al. 1995) of 33 meta-analyses showing that blinding in clinical trials reduces the efficacy of placebo therapy.

The authors also totally ignore the large body of empirical evidence from animal and human studies that placebo effects can be conditioned through the mechanisms of learning and memory and that these effects can be amplified or reduced experimentally (Ader 1985, 1997; Hernstein 1962; Siegel 1999; Siegel & Kretzler 1997; Siegel et al. 1982; Wickramasekera 1977, 1980, 1985). This conditioned mechanism can be inferred only indirectly from conventional double-blind crossover designed studies (Ader 1989). A conditioning demonstration of placebo effects needs direct experimental manipulation of the reinforcement property of the active ingredient in a drug or therapy and the associated and inevitable learned or memory component in all blinded or unblinded clinical trials (Wickramasekera 1977, 1980, 1985).

The definition and composition of the placebo groups in this study ignored the above empirical evidence bearing on the belief and conditioning

mechanisms of efficacy of the placebo. The authors followed a policy of blind empiricism in selecting their placebo groups for this meta-analysis and very likely accepted an unreliable definition of some placebos by simply accepting any group that a study "labeled a placebo group." Their definition of a placebo was simply based on any "intervention labeled as such in the report of a clinical trial" (Hróbjartsson & Gøtzsche 2001). Hence, we have no assurance that the placebo groups in this study were characterized by any uniform criteria.

The above basic considerations of ignored mechanisms of placebo efficacy and lack of uniform criteria of definition of the placebo cast doubt on both the process by which the placebo groups in this study were constituted and on any conclusions about placebo efficacy drawn by the study.

References

- Ader R 1985 Conditioned immunopharmacological effects in animals: implications for a conditioning model of pharmacotherapy. In: White L, Tursky B, Schwartz GE, eds. *Placebo: theory, research and mechanisms*. New York: Guilford Press
- Ader R 1989 Conditioning effects in pharmacotherapy and the incompleteness of the double-blind, crossover design. *Integrated Psychiatry* 6:165–170
- Ader R 1997 The role of conditioning in Pharmacotherapy. In: Harrington, ed. *Placebo*. Cambridge: Harvard University Press
- Cohen HD, Graham C, Fotopoulos SS, Cook MR 1977 A double-blind methodology for biofeedback research. *Psychophysiology* 14(6):603–608
- Evans FJ 1974 The placebo response in pain education. In: Bonica JJ, ed. *Advances in neurology: Vol 4. Pain*. New York: Raven
- Greenberg RP, Fisher S 1989 Examining antidepressant effectiveness: findings, ambiguities, and some vexing puzzles. In: Fisher S, Greenberg RP eds. *The limits of biological treatments for psychological distress*. Hillsdale, NJ: Erlbaum
- Hernstein RJ 1962 Placebo effect in the rat. *Science* 138: 677–678
- Hróbjartsson A, Gøtzsche PC 2001 Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 344(21): 1594–1632
- Kirsch L, Sapirstein G 1998 Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication. *Prevention & Treatment* 1: 1–17

Discussion

Hróbjartsson and Gøtzsche respond

- Roberts AH, Kewman DG, Mercier L, Hovell M 1993 The power of nonspecific effects in healing: implications for psychosocial and biological treatments. *Clinical Psychology Review* 13: 375–391
- Siegel S 1999 Multiple chemical sensitivity as a conditional response. *Toxicology and Industrial Health* 15:323–330
- Siegel S, Hinson RE, Krauk MD, McCully J 1981 Heroin “overdose” death: the contribution of drug-associated environmental cues. *Science* 216:436–437
- Siegel S, Kreutzer R 1997 Pavlovian conditioning and multiple chemical sensitivity. *Environmental Health Perspectives* 105:521–526
- Turner JA, Deyo RA, Loeser JD, Von Korff M, Fordyce WE 1994 The importance of placebo effects in pain treatment and research. *Journal of Nervous and Mental Disease* 271(20):1609–1614
- Wickramasekera I 1977 The placebo effect and medical instruments in biofeedback. *Journal of Clinical Engineering* 2(3):227–230
- Wickramasekera I 1980 A conditioned response model of the placebo: predictions from the model. *Biofeedback and Self-Regulation* 5(1):5–18
- Wickramasekera I 1985 A conditioned response model of the placebo effect: predictions from the model. In: White L, Tursky B, Schwartz GE, eds. *Placebo: theory, research, and mechanisms*. New York: Guilford

Hróbjartsson and Gøtzsche respond

Core belief in powerful effects of placebo interventions is in conflict with no evidence of important effects in a large systematic review

The five interesting, and at times polemical, comments on our review, “Is the placebo powerless? An analysis of clinical trials comparing

placebo with no treatment” (Hróbjartsson & Gøtzsche 2001), did not demonstrate flaws in our review, so we see no reason for changing our conclusion that there is little evidence that placebos in general have powerful effects.

Broad and narrow concept of “placebo effect”

A general problem in placebo research is the vagueness of the central concept of “placebo effect.” It is not a coincidence that such a concept does not appear in our paper. The phrase has been used to describe phenomena as different as patients’ improvement after a placebo intervention, the effect of a placebo intervention, psychologically mediated effects in general, the effect of the patient-provider interaction, the effect of suggestion, the effect of expectancies, and the effect of patients’ experience of meaning, etc. We think the notion is associated with too different (though overlapping) phenomena to serve as a conceptual tool for clear analyses.

All five comments discuss “the placebo effect.” What we investigated was the effect of placebo interventions (as defined by the researchers conducting the randomized trials we reviewed), or what Greene et al. call “the narrow definition” of placebo effect. We could not show that patients receiving placebos were markedly better off than patients not receiving placebos. Our result is neutral to many of the above meanings of the term “placebo effect.” It is therefore a misinterpretation of our work to claim that, for example, our results show lack of psychologically mediated effects on health or that the patient-doctor relation is not important.

Brody and Weismantel focus on whether “the placebo effect is real,” and Greene et al. find it implausible that placebos should have no effects “beyond natural history or regression to the mean.” It is not correct, however, to conclude from our findings that there is no effect of placebo. Our focus was clinical, that is, whether placebos induce effects of a magnitude that is important to patients. We were unable to find such effects. We feel the question whether there is, or is not, an effect of

placebo is theoretically intriguing but of secondary clinical interest.

It is impossible to define the notion of “placebo intervention” in a way that all researchers and clinicians would agree on. In the absence of such a definition, we chose to include all interventions that had been called placebos in randomized clinical trials (excluding only the few interventions that clearly had an effect beyond the treatment ritual). Our review therefore reflects the practical concept of placebo treatments as used in clinical research.

The various interventions called placebos have in common the appearance of standard treatments, albeit without their essential ingredients, but differ from each other in various other ways. Kirsch and Scoboria, and Ader question whether different types of placebos produce different effects. As we have reported in our paper, we tested whether the effects of pharmacological placebos (for example, tablets), physical placebos (for example, machines turned off), and psychological placebos (for example, neutral conversations) differed, but found no statistically significant difference.

Broad or narrow reviews

Kirsch and Scoboria, Brody and Weismantel, Ader, and Greene et al. question the broad approach of our systematic review. We chose a broad approach because we wanted to investigate the common assumption that all types of placebos can potentially cause effects on all types of clinical conditions. We were aware that subtypes of placebos could be of importance and that effects could differ between clinical conditions. We therefore planned several sub-analyses, and our review can perhaps best be described as consisting of a structured group of overlapping meta-analyses. As we expected the studies to be heterogeneous, we analysed the data with random effects models (that incorporate the heterogeneity). A broad approach for meta-analysis is often appropriate when there is no good a priori reason to exclude some conditions from consideration as is the case for placebo and, for instance, homeopathy (Gøtzsche 2000).

Ader’s claim that our aim was to identify a “single placebo effect” is a misunderstanding.

The pooled result is a rough overall estimate based on a heterogeneous data set combining different clinical scenarios. It is obviously unjustified to interpret such a result as valid for all conditions, settings, types of placebos, etc. In the discussion of the original paper, we stated clearly that the pooling of heterogeneous trials could have obscured a subpopulation of trials in which there was an important effect of placebo. However, the almost neutral result indicates that if placebo has positive effects in some situations, it must have negative effects in others. There are not many examples of interventions in health care where an overall neutral effect hides a beneficial effect in some patients and a harmful one in others. It should also be noted that, despite several sub-analyses, we were not able to identify subgroups of trials with clinically important effects of placebos with any reasonable level of certainty. The most promising condition was pain, but we identified bias in these studies, and even if the bias was ignored, the estimated effect was small. We believe it is important to look at the general pattern of our results instead of focusing on subgroup results.

Kirsch and Scoboria claim that the substantial heterogeneity in itself is a proof of effect of placebo. This is wrong. Heterogeneity can occur without any true effect, for example, because of publication bias or other biases (Egger M et al. 1997), or because of differences in the natural course of the diseases.

Methodological issues

Trial identification and inclusion

Brody and Weismantel question whether we have identified all trials. Our literature search was very detailed, and we found so many trials that it is very unlikely that the existence of some additional ones could have had an impact on our conclusion. Greene et al. repeat the point we made in our discussion that the publication bias of three-armed placebo trials could deflate the estimation of effect of placebo. We tested whether trials in which the

Discussion

Hróbjartsson and Gøtzsche respond

estimation of placebo was an explicit objective differed from the others, without finding any difference. We will, of course, incorporate any new trials in the updated Cochrane version of our analysis.

Ader comments that several studies of sequence effects were not included in our review. However, we studied the effect of placebo interventions not sequence effects. Ader furthermore states that “undoubtedly, several other ‘favorite’ studies were not included in this meta-analysis.” This accusation is not substantiated by evidence. In fact, we defined, and published, our inclusion criteria before conducting the systematic review (Hróbjartsson & Gøtzsche 1999). No trial that satisfied the inclusion criteria was excluded. Kirsch and Scoboria criticize that we did not quote more previous research. We quoted the research we found was most relevant, but space restriction in the *New England Journal of Medicine* impeded a larger sample of references. Kirsch and Scoboria furthermore state that we “seem to have missed a ‘seminal study’” comparing placebo and no treatment, and refer to a book by G. L. Paul published in 1966. We identified two overlapping non-clinical studies by Paul (1967, 1968), including the follow-up report of the 1966 study. Both studies were excluded as they involved college students, and not patients, and because they were not clearly randomized studies, and thus the trials did not comply with predefined inclusion criteria.

Ader, and Greene et al. raise the question whether lack of blinding of the reviewers could induce bias. Before we conducted this review we published articles in which we indicated that we expected to find an effect of placebo (Gøtzsche 1994, Hróbjartsson 1996), so if our expectancy would induce bias, it is not likely it would be in a negative direction. Furthermore, a study of the impact of blinding on reviewing trials could not demonstrate any bias (Berlin 1997). As long as there is no evidence of bias associated with unblinded reviewers, we recommend the standard practice of systematic reviewing without blinded reviewers, since blinding is very resource-demanding.

Credibility of placebo interventions

Wickramasekera points out that we did not verify the credibility of the placebo treatments. The information provided in the trial reports was not adequate for meaningful analyses, however. Furthermore, our aim was not to investigate the effect of beliefs or expectations as such. Theoretically, it is possible that certain experimental manipulations of patients’ beliefs and expectations could have an important impact on their health. However, the standard placebo intervention presented in an ordinary clinical trial has not been shown to have such effects. Kirsch and Scoboria indicate that “conventional placebos (placebo pills)” were most credible, and that too few trials with such placebos had been analyzed. The concern seems groundless as we analyzed 45 trials that included 5,462 patients receiving pharmacological placebos. Few meta-analyses can provide such impressive numbers which, in addition, lend sufficient power to the analyses.

Outcomes

Kirsch and Scoboria, and Greene et al. correctly point out that binary outcomes have less power than continuous ones, but the loss of power in analyses including more than 3,000 patients, as we identified in the trials with binary outcomes, is relatively small and cannot explain our mostly negative finding.

Greene et al. question our preference for binary outcomes. We argue that such outcomes are usually more clinically meaningful than continuous ones. For example, the main clinical aim of treating tobacco addiction is to quit smoking, not so much to reduce the numbers of cigarettes smoked. Only a handful of trials reported both binary and continuous outcomes.

We agree with Greene et al. that the selection of which outcome we thought “was most relevant to patients” implies a decision with some subjectivity. However, we decided on the selection process as a trade-off between the risks of two types of bias. The first is the potential subjectivity in the selection of outcomes for the review. The

second is the danger of selective reporting of positive results in the primary trials. As we described in our paper, we tested whether there was a difference in effect between the trials with a clear indication of a primary outcome and the trials in which we selected the outcome based on our conception of clinical relevance. We found no statistically significant difference. Greene et al. suggest a study of the proportion of drop-outs in the placebo and no-treatment groups. We have made no formal analyses of this, but our impression is that there was no clear difference in drop-out rates between the no-treatment groups and the placebo groups.

Ader claims that “in seven out of nine trials with objective binary outcomes, observers were aware of group assignments.” This is incorrect. In the seven trials with objective binary outcomes, it was unclear whether observers were blinded or not. The sub-analyses comparing the effect of placebo in the two groups of trials found no statistically significant difference.

Greene et al. focus on the statistical methods used and indicate that other types of methods could have changed close call p-values from nonsignificant to significant. We are less concerned with p-values, and more with whether the confidence intervals overlap with effects that are clinically relevant.

The pooling of data

Ader finds it problematic that we pooled trials comparing placebo with no placebo, and trials comparing experimental treatment plus placebo with only experimental treatment. As we reported in our paper, we found no statistically significant difference between the two types of trials (placebo as add-on treatment or not). The result may seem surprising if one assumes an important effect of placebo, but is to be expected if the effect is absent or small.

References to other types of research on placebo

A considerable part of the comments refers — paradoxically — not to our study but to other

types of research. Brody and Weismantel acknowledge that “the majority of ‘studies’ of the placebo response do not entirely control for such factors as the natural history of the illness and regression to the mean.” They imagine that 95% of the existing literature is worthless. We agree, although the percentage could be higher. After studying (and contributing to) the placebo literature for almost a lifetime, Shapiro & Shapiro (1997) gave it the following arid general characterization: “There is no systematic approach in published studies of the placebo effect. All ... are anecdotal reports, clinical impressions, theoretical formulations and post-hoc extrapolations of chance findings”

It is an important task to identify high quality empirical studies of placebo. Our review is based on all randomized clinical trials we could identify, comparing patients who were treated with placebo with patients who were not treated with placebo. We are generally surprised by the type, and sometimes quality, of research on placebo referred to by the commentators. Restriction in time made a comprehensive analysis of all the many references unfeasible, but the main points are discussed below.

First, numerous references are given to studies “showing placebo effects” which do not control for natural remission. For example, Greene et al. refer to Moerman’s (2000) review of ulcer studies, in which claims of effects of placebos are presented without controlling for the natural remission of ulcers. Furthermore, Wickramasekera refers to a review of “non-specific effects” (Roberts et al. 1993) and to a review of pain trials (Turner et al. 1994). Both reviews lack no-treatment groups. This practice has been prevalent in placebo research for decades, but is nonetheless a fundamental and serious methodological flaw. No causal inference can be made about the effects of placebo interventions without adequate control for natural remission, regression to the mean, and unknown factors. We recommend that researchers claiming effects of placebo stop referring to studies without adequate control groups.

We are also surprised that several commentators refer to studies that look at effects, not of placebo, but of other aspects of the patient-provider relationship. Thus, they are not of primary

Discussion

Hróbjartsson and Gøtzsche respond

relevance to our investigation. For example, Greene et al. refer to two important studies of the impact of informed consent (Dahan et al. 1986, Bergman et al. 1994), and to a string of studies of the impact of different instructions on asthma. The aim of our study was not the effect of informed consent, or instructions. Brody and Weismantel refer to review of interactions between placebo and active treatment (Kleijnen et al. 1994), and to a review of the impact of the color of tablets by de Craen et al. (1996), as “showing a placebo effect.” We studied the effect of placebo interventions overall, not the effects of defined components of placebo interventions, as, for example, the color of tablets, nor did we investigate the effects of interactions. Furthermore, de Craen and colleagues’s conclusion was that “little research has been carried out,” and the available evidence “suggests” color “may” have an effect. This does not constitute clear-cut evidence for effects of the color of tablets. Ader refers to studies measuring relapse rates or adherence to treatment; again neither were our research objectives.

Third, Greene et al. refer to studies that investigate whether different forms of placebo differ in effect (Kaptchuk et al., 2000, Grenfell et al. 1961). It is problematic when Greene et al. characterize these studies as having “methodological shortcomings” and, at the same time, state that they “demonstrate that different routes of placebo administration produce different magnitudes of placebo effect.” Trials with methodological shortcomings have a high risk of bias and will often overestimate the effect of an intervention. Results from such trials are therefore not reliable. Greene et al. also refer to two systematic reviews, one comparing placebo as injection with oral placebos for migraine (de Craen et al. 2000 a), and one comparing two oral placebos with four oral placebos against ulcers (de Craen et al. 2000 b). However, the reviews evaluate the results in trials comparing active treatment with one type of placebo, with the results of other trials comparing active treatment with another type of placebo. Lack of direct randomization between the two types of placebo interventions makes such reviews vulnerable to confounding. This was recognized by the authors of the ulcer review who

prudently stated that “we cannot rule out that in this nonrandomized comparison the observed difference was caused by some unrecognized confounding factor or factors.” Thus, though the studies are clearly of high standard as well as being interesting, they do not provide reliable evidence that different types of placebos produce different effects.

Fourth, Kirsch and Scoboria, Wickramasekera, and Greene et al. refer to several previous meta-analyses comparing placebo and no-treatment, but they do not clarify that the reviews referred to are fundamentally different from ours. Whereas we directly compare patients who have been randomized to placebo and no treatment, both Kirsch & Scoboria (1998), Smith et al. (1980), and Shapiro & Shapiro (1982) (as well as Dush 1986 and Grissom 1996 who are not referred to) evaluate differences in effects in trials that compare an active treatment with placebo with other trials that compare active treatment with no treatment. Again, there is no direct randomization between placebo and no treatment, so the design is not reliable. We are not surprised that the effect reported in these trials is somewhat larger than we found.

Brody and Weismantel, and Greene et al. provide a fifth type of reference to studies that investigate the mechanisms of placebo reactions. Internal opiate secretion has been suggested as a mechanism for the possible analgesic effect of placebo. Such research is very interesting, however a systematic review could not conclude that the hypothesized mechanism was proven beyond reasonable doubt (ter Riet et al. 1998). Even assuming that a biologically plausible mechanism can be firmly established, such a result is not necessarily at odds with our review, because we do not rule out minor effects of placebo, for example on pain. This question clearly needs further research.

A sixth type of reference is to several studies that seem to have been misquoted. For example, Schulz et al. (1994) did not show that blinding in clinical trials reduces the efficacy of placebo as Wickramasekera postulates. Furthermore, Kirsch and Scoboria are not correct when they claim that Shapiro & Shapiro (1982) used the same criteria as we did in selecting studies for their review. They

included studies in which “two or more groups received different psychological treatments, and another group (a no-treatment group, or failing that, a minimal-treatment group).” We included studies in which patients had been randomized to placebo or no placebo. A study by Traut & Passarelli (1957) is referred to by Kirsch and Scoboria when they claim that “placebo injections are more effective than placebo pills.” However, the study was not randomized and its authors cautiously emphasized: “One is justified in making only tentative assumptions from this small group of 39 patients.”

Conclusion

We agree with Brody and Weismantel that a large part of past empirical research on placebo is flawed, primarily because of lack of control groups. The interpretation of results is further complicated because different meanings are associated with the concept of “placebo effect.” However, since Beecher’s (1955) influential but mistaken paper “The Powerful Placebo,” the central example of “placebo effect” has been the causal association between receiving a placebo intervention, for example a dummy pill, and a considerable clinical improvement. We are aware that by failing to find such an improvement we challenge what Brody and Weismantel call the “very core” of the “belief system” of several people, some of whom “have based their intellectual approach ... on the idea that the placebo effect is real.” To prove there is no, or only minimal, effect of placebo interventions in all settings is impossible, even with a large number of heterogeneous trials such as the sample we collected. Despite our mostly negative findings, important effects of placebo interventions might exist, for example, in unidentified subgroups in the review, or in outcomes not included. However, the burden of proof now rests with those who claim there are important effects of placebo interventions. Claims of worthwhile effects should be based on reliable evidence, preferably rigorously conducted systematic reviews of randomized trials, not on beliefs.

References

- Beecher HK 1955 The powerful placebo. *JAMA* 159:1602–6
- Bergmann JF, Chassany O, Gandiol J, Deblois P, Kanis JA, Segrestaa JM, et al 1994 A randomised clinical trial of the effect of informed consent on the analgesic activity of placebo and naproxen in cancer pain. *Clinical Trials and Meta-Analysis* 29:41–47
- Berlin JA 1997 Does blinding of readers affect the results of meta-analyses? *Lancet* 350:185–186
- de Craen AJM, Tijssen JGP, de Gans J, Kleijnen J 2000 Placebo effect in the acute treatment of migraine: subcutaneous placebos are better than oral placebos. *J Neurology* 247:183–188
- de Craen AJM, Moerman DE, Heisterkamp SH, Fytgat GNJ, Tijssen JPC, Kleijnen J 2000 Placebo effect in the treatment of duodenal ulcer. *Brit J Clin Pharmacology* 48:853–860
- de Craen AJM, Roos PJ, de Vries AL, Kleijnen J 1996 Effect of colour of drugs: systematic review of perceived effect of drugs and of their effectiveness. *BMJ* 313:1624–1626
- Dahan R, Caulin C, Fiega L, Kanis JA, Caulin F, Segrestaa JM 1986 Does informed consent influence therapeutic outcome? A clinical trial of the hypnotic activity of placebo in patients admitted to hospital. *BMJ* 293:363–364
- Dush DM 1986 The placebo in psychosocial outcome evaluations. *Evaluation & Health Professions* 9:421–438
- Egger M, Smith GD, Schneider M, Minder C 1997 Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315:629–634
- Grenfell RF, Briggs AH, Holland WC 1961 Double blind study of the treatment of hypertension. *JAMA* 176:124–128
- Grissom RJ 1996 The magical number .7 +/- .2: meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *J Consult Clin Psychology* 64:973–982
- Gøtzsche PC 2000 Why we need a broad perspective on meta-analysis: It may be crucially important for patients. *BMJ* 321:585–586
- Gøtzsche PC 1994 Is there logic in the placebo? *Lancet* 344:925–926
- Hróbjartsson A, Gøtzsche PC 2001 Is the placebo powerless? An analysis of clinical trials comparing placebo treatment with no treatment. *New Eng J Med* 344:1594–1602
- Hróbjartsson A, Gøtzsche PC 1999 Placebo intervention compared with no treatment (Protocol for a Cochrane Review). In: *The Cochrane Library, Issue 3*. Oxford: Update Software
- Hróbjartsson A 1996 The uncontrollable placebo effect. *Eur J Clin Pharm* 50:345–348
- Kaptchuk TJ, Goldman P, Stone DA, Stason WB 2000 Do medical devices have enhanced placebo effects. *J Clin Epidemiology* 53:786–792
- Kirsch I, Sappirstein G 1998 Listening to prozac but hearing placebo: a meta-analysis of antidepressant medication. *Prevention and treatment* 1:1–17

Discussion

Hróbjartsson and Gøtzsche respond

- Kleijnen J, de Craen AJ, van Everdingen J, Krol L 1996 Placebo effect in double-blind clinical trials: a review of interactions with medication. *Lancet* 344:1347–1349
- Moerman DE 2000 Cultural variations in the placebo effect: ulcers, anxiety, and blood pressure. *Medical Anthropology Quarterly* 14:51–72
- Paul L 1967 Insight versus desensitization in psychotherapy two years after termination. *J Consulting Psychol* 31:333–348
- Paul L 1968 Two year follow-up of systematic desensitization in therapy groups. *J Abnormal Psychol* 73:119–130
- Roberts AH, Kewman DG, Mercier L, Hovell M 1993 The power of nonspecific effects in healing: implication for psychosocial and biological treatments. *Clinical Psychology Review* 13:375–391
- ter Riet G, de Craen AJM, de Boer A, Kessels AGH 1998 Is placebo analgesia mediated by endogeneous opioids? A systematic review. *Pain* 76:273–276
- Schulz KF, Chalmers I, Hayes RJ, Altman DG 1995 Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 273:408–412
- Shapiro AK, Shapiro E 1997 The placebo: is it much to do about nothing? In: Harrington A (ed). *The placebo effect. An interdisciplinary exploration*. Harvard University Press, Cambridge: p.12–36
- Shapiro DA, Shapiro D 1982 Meta-analysis of comparative therapy outcome studies: a replication and refinement. *Psychological Bulletin* 92:581–604
- Smith ML, Glass GV, Miller TI 1980 The benefits of psychotherapy. John Hopkins University Press, Baltimore
- Traut EF, Passarelli EW 1957 Placebos in the treatment of rheumatoid arthritis and other rheumatic conditions. *Ann Rheum Dis* 16:18–22
- Turner JA, Deyo RA, Loeser JD, Von Korff M, Fordyce WE 1994 The importance of placebo effects in pain treatment and research. *JAMA* 271:1609–1614