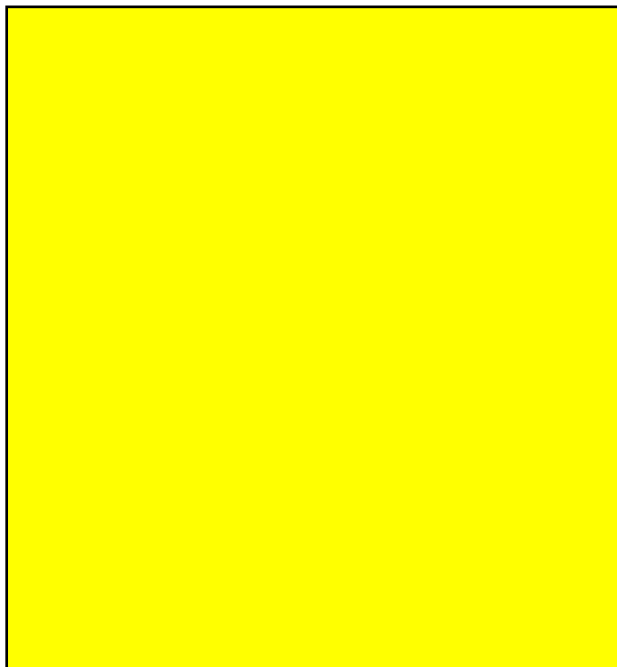


Discussion

Greene et al.



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

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

* 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.

** 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
- Greenfell 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