

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 reaction: Internal opiate secretion has been suggested as 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 we did in selecting studies for their review. The

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 JPG, 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 endogenous 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