

# **CRITICAL REVIEW**

## **Senior Honours**

**0107170**

**Morag D Nimmo**

**Supervisor: Dr S W Draper**

### **Placebo: Real, Imagined or Expected? A Critical Experimental Exploration**



**UNIVERSITY  
of  
GLASGOW**

12 January 2005

## ABSTRACT

Placebo literature has been profuse since the 1950's, with common held misconceptions shaping the future of medical, pharmaceutical, clinical and behavioural research. This review analyses research reviews of the seminal paper *The Powerful Placebo* by Beecher (1955), whose findings became scientific fact, convincing the scientific world that placebos were both ethical and scientifically necessary. Claims that one third of the population reacted to placebo have a huge impact on research outcomes. It took some 40 years for Beecher's research to be examined and Kienle and Kiene (1997) reported that the studies included in Beecher's original review could claim no such 'powerful placebo' effect at all. Kienle and Kiene's review along with subsequent meta-analyses are reviewed, with a view to identifying if there is such a thing as a placebo effect and what medical conditions do respond to it. Key to the debate is a lack of consensus in what constitutes a placebo as such. Experimenter and subject expectancies have been argued as a major confound in placebo and research in general and therefore Rosenthal (1966,1999) and Rosenthal and Jacobsons' (1968) findings and research into '*Expectancy Control*' is examined. Criticisms of research are offered and recommendations for the future of research methods in the furthering of basic understandings of placebos and their utility as a research tool are included.

## Introduction

Placebos have been used for centuries in the form of 'bread pills' in order to make patients feel better, but it was not until 1955 that an attempt was made by Henry Beecher to quantify the placebo, its uses and effects in many different therapeutic dimensions. Beecher stated that placebos had such a powerful effect that 35% of the population respond to placebo. Since then his seminal paper *The Powerful Placebo*, has been considered scientific fact and a benchmark made for future research in the area. Nordenberg (2000) explains that Beecher's claims of a 'Powerful Placebo Effect' resulted in all new drugs having to be shown to be significantly better than their placebo at improvement in symptoms or response. Indeed Evans (2003) claims "Beecher played a major role in persuading doctors that placebo controls were both ethical and scientifically necessary" (Evans, 2003: 10). The law surrounding the licensing of new drugs is argued by Wall (1999) to have perpetuated the separation of true versus imagined effects of placebo, thus giving a focus to placebo trials, rather than opportunities to refute the 'Powerful Placebo Effect' claims. Thus no published works appeared until the 1990's to this effect.

Once placebo controlled drug trials became a legal requirement in the licensing of new drugs, a plethora of research ensued into objective and subjective factors in placebo controlled treatment regimens. It was not until the 1990's that Beecher's claims were questioned or examined. It has since been argued that none of Beecher's original research claims have any basis and his methodology and selective reporting of results have been called into question.

A logical starting point would normally be to examine and define what a placebo is and indeed what a placebo effect is, however, placebo can have many forms and it will be argued that this is partly the reason for such ambiguity and mixed research findings leading up to the 1990's, when questions were beginning to form as to the basis of Beecher's 'Powerful Placebo Effect'. Listings in the Chambers Dictionary (2000) describes a placebo as one of four possibilities: vespers for the dead; a medicine given to humour or gratify a patient rather than to exercise any physical curative effect; a pharmacologically inactive substance administered as a drug, either in the treatment of psychological illness or in the course of drug trials; a sycophant. The Skeptic's Dictionary (2002) describes the placebo effect as 'a measurable

observable or felt improvement in health not attributable to treatment and that many believe the effect to be due to the placebo itself. Latin for 'I shall please'; placebo is a medicine or treatment believed by the administrator of the treatment to be inert or innocuous. However, placebos are now recognised to be more than sugar or starch pills: 'fake' surgery and 'fake' psychosurgery are also considered placebos' by some.

Despite the vast research in the area, very little is known of why an inert substance or 'fake' surgery or therapy would be effective.

One suggestion offered is that the placebo effect is psychological, a belief or expectancy of a theory or indeed to a 'subjective feeling' of improvement. These issues will be discussed in greater detail relating to how a persons hopes and beliefs combined with suggestibility may have a significant biochemical effect. As sensory experience and thoughts can affect neurochemistry, Carroll (2002) correctly states that our neurochemistry is affected by other systems including the hormonal systems, suggesting a possible link. Various authors have argued that some clinical conditions more than others respond to placebo (Kienle & Kiene, 1997; Hrobjartsson & Gotsche, 2001; Carroll, 2002; Nordernberg, 2000). Moving on from examining previous studies and combining their findings, Evans (2003) puts forward an argument for a testable model of why placebos should have an effect in certain disorders and not in others and openly welcomes people to disprove it. This would seem to be the first step in a new direction for research.

What is it then that caused Beecher to claim that 35% of people will improve with placebo treatment, then to find that no such evidence existed, whilst others claim that the figure is nearer 70%? Is expectation such a powerful thing as Nordenberg (2000) states that the more you believe you're going to benefit from a treatment the more likely it is you will experience a benefit? Why does the colour of the placebo pill influence its effect? Why does a blue pill have a depressant effect and red a stimulant effect? (Backwell, Bloomfield and Buncher, 1972) or why is blue better at making you go to sleep than orange? (Luchelli, Cattano and Zattoni, 1978).

The format of this review will examine the history and rise of placebo use and research, reviewing the seminal paper by Beecher in 1955 and how this impacted the

scientific community, together with recent questioning of its initial findings. The spread of these beliefs will be examined through examining scientific literature and research. Expectancy theory and belief effects will examine a more detailed explanation of Rosenthal's analysis of 'experimenter effects'. This comprehensive coverage of the theory of experimenter expectancy effects, explains how to manipulate and implement both experimenter and subject expectancy controls.

### **History and background of the Placebo**

The Chambers definition of placebo is echoed in the various historical writings in relation to placebo. Wall (1999) informs that Chaucer in 1340 referred to placebo with reference to the Psalms sung for the deceased, in mockery of the priests however, at the manipulation of money from people to assure prayers to be sung for the dead.

By the seventeenth century placebo's were adopted by doctors for the inactive medicines that greatly impressed their patients, however, even then some controversy surrounded placebo usage as Burton in 1628 claimed "there was no virtue in some remedies" other than "a strong conceit and opinion" (Wall, 1999:128).

On a similar vein, President Jefferson in 1807 claimed it was a 'pious fraud' to use "bread pills, coloured water and powders of hickory" to debate over fraudulent placebo's and medicines that were at the time believed to be acting as a rational mechanism.

Traditionally placebos were used to test medical treatments that had recently been manufactured and used. This in itself has caused another debate, that of ethics (Nordenberg, 2000). Is such deception judged mentioned above ethically correct? Is the therapy judged placebo or true?

As Senn (2003) states, when a placebo group is added, we are actively placing one group at a disadvantage, thus it is open to ethical questioning. Many patients are entered into a clinical trial and have a treatment allocated at random. Senn, a medical statistician, states the ideal starting point would be 'equipoise', where the administrator is entirely ignorant to which treatment is better. Blinding has long since been used in placebo research and refers to an unawareness of which treatment group

is applicable. A single blind study refers to the subject being 'blind'; a double blind study refers to both subject and experimenter being blind to treatment group. However despite equipoise being the case in double blind control trials for strong drug trials, Senn accepts that at some point we will have a 'hunch' as to which treatment is better, thus we no longer hold the unbiased attitude.

Indeed a Placebo was used in the second MRC trial published in the BMJ in 1950 by Bradford Hill. This was a trial of antihistaminic drugs in treating the common cold (Day and Ederer, 2004). But not until 1955, was an attempt made by Henry Beecher to quantify the placebo and its uses and effects in many different therapeutic dimensions. Since then, his seminal paper *The Powerful Placebo* has been considered scientific fact (Kienle and Kiene, 1997).

### **Beecher's Considered Seminal Work**

Beecher's own interests in a placebo response came from his experiences as a field doctor during the War years, where analgesics were often in short supply. As analgesics ran out in the field hospitals, a placebo injection of saline solution was often given as a dummy treatment. Remarkable pain relief was noted from these patients. Similar results were being reported in medical literature, to pills and powders alike.

By 1955, Beecher decided it was time for an examination of the literature covering these pills and 'dummies' as they were also known, to distinguish pharmacological effect and the effect of suggestion, in what Beecher claimed was an 'unbiased assessment' of the result of experiments.

As it will be shown, this attempt was not as unbiased an assessment as it should have been. Beecher's original claims will be discussed fully initially, due to the dramatic effect they had on scientific and therapeutic practices for the forty plus years that were to follow, not to mention the lessons to be learned from his mistakes. This first comprehensive attempt to distinguish what, if any, were the responses to the use of placebo were argued to show that if the placebo had neither the reactivity nor the physical dimensions required of an effective drug, then what was their action on the reaction or processing component of suffering.

### ***Beechers' Claims***

Through this examination of 15 trials with different diseases, of the 1082 patient, the magnitude of the therapeutic effect of placebo was on average  $35.2 \pm 2.2\%$  and that the effect is on the reaction component of suffering. That is 35% were argued to have been satisfactorily relieved by placebo alone.

The studies chosen for analysis by Beecher were argued to be human ailments where subjective factors can enter, included those covering wound pain, angina pectoris, headache, nausea, cough and drug induced mood changes, anxiety and tension. The magnitude of the therapeutic effect of placebo was argued as evidence that drugs are capable of altering subjective responses and symptoms through their effect on the reactive component of suffering.

### ***Toxic Effects***

It was claimed that not only did placebo produce beneficial results but they had associated toxic effects also. A selection of the effects attributed to placebo were, dry mouth: 9%; nausea, 10%; sensation of heaviness, 18%, headache, 25% and drowsiness 50%. These reports were argued to have been recorded during double blinded trials. A further study describes 3 of 31 patients showing major reactions to placebo: weakness, rash and angioneurotic edema of the lips are a few of those mentioned which Beecher argued to be objective evidence that the reaction phase of suffering can produce physical change. Toxic effects were argued to show that placebos could set off adrenals and mimic drug action; indeed the more severe the disease state, the greater the effect.

From the conclusion offered by Beecher, it was obvious that placebos were a hotly debated issue at the time with details of controversies discussed at a conference where Fantus had argued that those with a lower IQ were more likely to show a placebo effect, whereas Diethelm posited that people react to suggestion, to the point that it becomes a reality. However, Beecher continues that if the Placebo was powerless, in that it was just beliefs etc. then an increase in the severity of pain would be noted along with stress in a situation, which he argues is clearly not the case at all.

These conclusions given by Beecher were accepted as scientific fact for over 40 years, in which time Beecher became the most frequently cited paper in placebo research.

However, 40 years on, it wasn't until evidence published claiming that the 'powerful placebo effect' was in many incidences much stronger than the 35% claimed by Beecher, but much higher at 70% in some incidences, the disparity was so high that Kienle became interested in the varied reports and beliefs surrounding placebo's. Kienle (1995) had previously completed a reanalysis of a classic German study carried out in 1986 by Netter et al, again showing the no placebo effects could be found. The resulting disparity of reports inspired them to go back to the original data Beecher analysed, to examine his claims and the data closer.

### **Kienle and Kiene's analysis of Beecher**

This research was carried out by Kienle and Kiene (1997), aimed to identify the answers to two questions. Firstly, did Beecher actually identify a true placebo effect in the 15 papers he looked at and secondly, if this was not the case, what were the factors (if any) that could have given a false impression of a placebo effect. The surprising results showed that no evidence was found at all in any of the studies cited by Beecher of a powerful placebo effect. Kienle and Kiene argued that there were many alternative explanations, other than placebo effect that could explain the results in each of the studies reported by Beecher. Kienle and Kiene have accused Beecher of 'sloppy methodological thinking', which has set the standard for research in this area in the intervening years, to the point that so much disparity is now seen in reported results. Beecher is accused of selective reporting of results and indeed misquoting the original authors in some cases where any effects were explicitly said not to have been due to a placebo effect. Kienle and Kiene urge that to fully understand the current placebo literature, we need to understand the mistakes and misinterpretations made by Beecher. These authors are commenting on literature claiming that 'today's' placebos were supposedly effective in almost every disease, with estimates going way beyond the initial claims of Beecher.

#### *Alternative explanations for reported placebo effect*

A wide range of alternative explanations that could be offered which do not support a placebo effect include: spontaneous recovery; additional treatment effects; fluctuation of symptoms; regression to the mean; conditional switching of placebo treatment; scaling bias; irrelevant or questionable response variables; answers of politeness; experimenter subordination; conditioned answers; neurotic or psychotic

misjudgements; psychosomatic phenomenon, misquotation and false assumptions of toxic effects. A few of these will be discussed in more detail as more recent literature also supports these views.

*Spontaneous recovery or improvement* Beecher claimed that in one study 35% of patients receiving placebo for the common cold recovered after 6 days. The original author stated that an improvement after 6 days would not be unusual, but this was not given credence by Beecher. Of the four trials on post-operative pain, Kienle and Kiene report that at least 2 could be explained through the patients' decreasing requests for analgesics and spontaneous improvement was a major factor in the clinical conditions covered by 10 of the 15 trials reported by Beecher. This error is argued to be widespread in literature still.

*Fluctuation of symptoms* Kienle and Kiene state that a feature of chronic diseases is that patients feel better one day and worse the next, but Beecher did not comment on such decreases and increases, instead reporting only the rate of improvement calling this the placebo effect. In a study of patients complaining of ulcers, migraine, muscle tension of headache who also suffered from anxiety and tension and were treated for eight, 2-week periods alternately with mephenesin and placebo, Beecher claimed a placebo effect of 30%, but Kienle and Kiene report that 20 – 30% got better and that 10-20% got worse, which was a net improvement of only 5-10%.

Senn (2003) points out that we should be interested in the difference between the effects on treatment and placebo groups and not the individual values of each.

*Conditional switching of treatments* In some of Beechers' reports, when patients felt better, they were excluded from the study until they felt ill again and could given placebo. In a study of postoperative pain, patients were only included if they had recovered sufficiently to receive oral medication and when they deteriorated, they were excluded until another improvement was seen.

*Irrelevant or questionable response variables* In an MS study, reported feelings of euphoria were taken as being a placebo response, which in actuality is one of the symptoms of MS itself. In a hypertension study, a placebo effect of 61% was noted against the drug veratum. However, all patients who had reported toxic effects when given the drug were then switched to placebo. The reported relief in the 61 of the

64% affected could be explained by the cessation of the veratum toxicity, with no reason to explain this due to a placebo effect.

*Experimenter subordination or conditioned answers* In the original studies, Kienle and Kiene report that the original experimenters claimed that the responses of the patients were often contrary to the physicians impression, possibly showing good manners and politeness, rather than true honest responses. Often patients are grateful for the time and attention offered by physicians, and respond over optimistically. Alternatively, patients can often say what is expected of them rather than the truthful response. To resolve conditioned answers, one needs to differentiate between the therapeutic placebo effect and conditioned effects. In clinical experiments, healing is not conditioned. Chronic diseases are more difficult to treat than acute first manifestations of illness. However, conditioning may be important when giving placebos as this can produce answers of politeness, verbal placebo effects or experimental subordination.

The main criticism of Beecher's findings by Kienle and Kiene, is that of selective reporting and sloppy methodological reporting, which has set a standard and marker for the research and reporting of future placebo studies. A key example of selective reporting was of 3 studies of angina reporting a 20% improvement nothing is mentioned of the 72% who deteriorated on placebo. Given the impact Beechers claims had on the scientific community, these standards have led to a mass misrepresentation in the ensuing 40 or so years since its publication.

### **The Hrobjartsson and Gotsche Meta-Analysis**

Similarly, Hrobjartsson and Gotsche (2001) also questioned the quality of evidence surrounding placebo treatments, claiming that previous findings had not been rigorously evaluated. They therefore claimed to have rectified this by conducting a systematic review of clinical trials in which patients were randomly assigned to either placebo or no treatment groups. Hrobjartsson and Gotsche, argue that the majority of Beecher's reports and others have estimated the effect of placebo as the difference from baseline in the condition of the patients of a randomised trial after treatment, and therefore cannot be distinguished from the natural course of the disease, the statistical regression to the mean and other factors. The reported large placebo effects could, in

part, be artefacts of inadequate research methods. A second aim of their study was identify whether the effect of placebo differed for objective and subjective outcomes.

The inclusion criteria set by Hrobjartsson and Gotsche for their meta-analysis appears to be strict and precise, despite a wide range of clinical conditions covered. Having conducted a literature search of on-line resources, studies were only included if patients were randomly assigned to groups and excluded if randomisation was predictable or unconcealed. No patients were to have volunteered, nor paid for their involvement. Drop out rates were below 50% and no other therapies were allowed that could possibly be construed as confounding results, e.g. movement techniques following operations. Studies were only included if they had three groups: treatment, placebo and no treatment. This, argue Hrobjartsson and Gotsche, is the only valid way to judge impartiality, if there is an effect of placebo as compared to a no treatment group.

More than 40 clinical conditions were investigated in the meta-analysis, covering a broad range: common cold, herpes simplex infection, asthma, marital discord, stress related to dental treatment and undiagnosed ailments. Binary and continuous outcomes were measured separately. Placebo treatments were judged so by the original authors, not Hrobjartsson and Gotsche.

Hrobjartsson and Gotsche claim in their conclusion to have found little evidence that in general, placebos have powerful effects, contrary to the claims of Beecher. They report small possible benefits on studies with continuous outcomes and also on the treatment of pain.

Kirsch (2002) however, one of the main critics of Hrobjartsson and Gotsche indicated that by comparing placebo groups to no treatment groups, this was not only unethical, but also not comparable. This is on the grounds that one group is actively denied treatment. One further criticism Kirsch made is also on the types of treatment claimed by Hrobjartsson and Gotsche to be placebo, in that one cannot evaluate the effects produced by medical placebos by examining studies in which the so called placebo uses irrelevant procedures, such as answering questions about one's favourite food, talking about pets, or reading short stories. Kirsch goes on to criticise Hrobjartsson

and Gotsche in that some of the placebos cited as treatment were included as control group treatments on other studies they also cite. In looking closer at the studies cited by Hrobjartsson and Gotsche, Kirsch offers that despite a general finding of no placebo response, there were in fact some which showed very strong responses to treatments of disorder whilst at the same time, others showed a weak response or no response at all.

The sweeping general statement over all clinical conditions included in the Hrobjartsson and Gotsche (2001) meta-analysis probably did not show any effect, but certain clinical conditions did in fact show some degree of placebo response. Kirsch therefore argues that placebo groups should be evaluated in reference to particular disorders, echoing the views of Kienle and Kiene (1997).

### **Hrobjartsson and Gotsche Update**

Hrobjartsson and Gotsche (2004) have since updated their research, by examining data from subsequent years 1999 – 2002, correcting some criticisms and investigated the effects of specific conditions. One of the inclusion criteria continued from their previous work is the inclusion of a third treatment group, that of no treatment, as compared against placebo and ‘active’ treatment. Hrobjartsson and Gotsche argue that without a no treatment group, no comparison can be made of a placebo intervention as distinguished from the natural course of the disease. This however, leads to a problem with double blinding procedures. Subjects in the no treatment group are well aware that they are not receiving treatment, however, placebo group members know at best they are receiving treatment, at worst placebo. Therefore, the thought patterns are not the same. With no double blinding procedures, the results are open to bias from various areas. Any studies that do not include randomised allocation of subjects to a placebo and to a no treatment group say Hrobjartsson and Gotsche could therefore be equally attributable to inadequacies or artefacts of research methods. Hrobjartsson and Gotsche noted that the patients’ in the studies self-reported estimated effects of continuous outcomes were three times higher than observed outcomes, thus they claim a reporting bias is recorded. Another disparity with unblinded patients is that they were more likely to seek treatment out with the study, therefore biasing results.

Their results did however examine conditions individually and only pain and phobias were found to respond to placebo on continuous outcomes, again, no binary outcomes at all were significant. Heterogeneity of results was better for trials in which concealed randomisation was used, i.e. the experimenter was not responsible for randomly allocating patients to treatment groups.

### **Defining Placebo**

It could be argued that at the root of the placebo debate, lies the varying descriptions and understandings of what exactly a placebo is. Indeed, without a shared common understanding and acceptance of definition of what is being discussed, then it is not surprising to find such a varied literature on reported placebo effects.

Beecher cites Goddum (1953) as stating the word placebo to mean a medicine given more to please than to benefit the patient. It is something intended to act through a psychological mechanism adds Beecher, accepting its use as an aid to therapeutic suggestion, whilst also being used as a tool to get at certain fundamental mechanisms of the action of drugs, especially those involving subjective responses. Therefore a placebo had to be given and the event had to have an effect on the placebo response, that wouldn't otherwise have occurred, and the event had to have some relevance to the medical condition (Kienle and Kiene, 1997).

Kienle and Kiene continue that placebos are imitations of specific treatments with the absence of specific therapeutic constituents. Under these criteria, many of the studies cited by Beecher and Hrobjartsson and Gotsche, are unsuitable for inclusion in the analysis of a placebo effect. However, even Hrobjartsson and Gotsche (2004) state that there is no formal definition of a placebo which all researchers and clinicians agree on.

The Skeptic's Dictionary, Carrol (2002) describes the placebo effect as a persons beliefs and hopes about a treatment, combined with suggestibility and may have a significant biochemical effect, in that sensory experience and thoughts can affect neurochemistry. Carrol raises the question is it more a case of mind over behaviour and not molecules? That is the behaviour of the sick is a learned therefore, to an extent we role-play when sick or hurt. Indeed it is further added that this behaviour is

socially and culturally based in the belief system, which has developed over a lifetime (Carroll, 2002).

### **Beliefs and Expectancies**

Some commentators argue that nature is merely taking its course, as those given no treatment and placebos do not do as well as active treatment and placebos (Nordenberg, 2000). The process of treatment in itself is argued to be enough to promote healing also by being shown attention, care or affection, in a process which is hopeful and encouraging may be sufficient to trigger physical reactions in the body.

Carroll (2002) asks if the placebo effect is just a measurement of the change of behaviour affected by the belief in the treatment. He continues by explaining behaviour change can include an attitude change, changes in what one says about how one feels along with how one acts. Experimental evidence from studies using placebo alone is offered: a brightly coloured dye was painted on warts and by the time the colour had faded, the warts had gone also; for pain relieving treatment following dental extractions, patients got relief despite the malfunctioning of the ultrasound equipment responsible for delivering the treatment, both dentist and patient believed the equipment to be in perfect order and therefore had a belief on its effect.

“Expectation is a powerful thing” (de Lap, in Nordenberg, 2000), the more you believe you are going to benefit from a treatment, the more likely you’ll experience a benefit. Wall (1999) however, argues that patient expectancies can in fact trigger the placebo effect. A 1995 study of patient admissions complaining of headaches, to an Accident and Emergency department based on the assumption that patients would expect to be given a strong medicine to help relieve symptoms. In this double blind study, one third of patients were given one of three possible treatments: aspirin; a narcotic; or saline, all administered by injection. Despite the pharmacological differences in the three treatments, all patients reported the same reduction in pain, regardless of treatments. This was argued to be a subtle but powerful expectancy effect. Any injection is expected to be powerful, more so than a pill (Wall, 1999). Rosenthal (1966) has carried out much research on experimenter expectancies, like those of Beecher, along with subject expectancies similar to those described by Wall and the various manifestations these can have on research results. His interest stems from an initial experience with such effects when completing his PhD and realised he

may in fact have had an effect on his subjects taking part in his research. Since then he has examined such effects and methods and makes several recommendations on avoiding such confounds (Rosenthal, 1999).

### **Experimenter Expectancy Effects**

To Rosenthal, the application of scientific method to the study of human behaviour has shown more precisely how complex human behaviour is and how much we still do not know or understand about it. The complexity of behaviour may derive from the psychological or behavioural experiment itself and thus the complexity may stem not from individual subjects themselves, but rather in the experimenter and the interaction between the experimenter and the subject. This offers important implications on how research is therefore carried out, conducted and assessed.

It is proposed that the double blind method of research is more than warranted, quoting research by Haas, Fink, and Hartfelder (1963) showing that when the experimenter does not know that the substance given to the subject is inert, he not only expects, but gets a better result.

Rosenthal (1966) describes various factors that may be affected by experimenter effects including those that although the experimenter does not affect the subject responses themselves, may affect the results and conclusions made. Alternatively, when the experimenter serves as an observer of the subjects' behaviour and the data is recorded, subsequently analysed and interpreted, s/he may err on the significant side, even though not altering the subject responses.

This is in contrast to the situation when the experimenter interacts with the subject, thus his own attributes, attitudes and expectancies may prove to be significant determinants of subjects behaviour in an experiment. One of the most familiar experimenter effects of this type, known to all psychology undergraduates is that of 'Clever Hans' reported by Pfungst (1911). Mr van Osten his owner trained him to spell, read, subtract, multiply divide and 'solve problems of musical harmony'. By inadvertently changing his poise just as the horse came to the correct response, Hans picked up on the physical cue given by the trainer and stopped, offering the illusion of knowing the correct answer.

In a clinical setting, one famous situation is reported by Greenblatt (1964). He describes a patient suffering from advanced cancer. When most treatments had been exhausted, the doctor explained there was a new drug that was believed to be extremely effective, and so the patient was given the drug. The patient soon showed a remarkable improvement and was discharged on the drug. Shortly afterwards, it was reported that the new drug had been proven to be ineffective and the patient relapsed and returned to hospital. The doctor knowing there was nothing else to try, convinced the patient that the drug was effective for him, gave an injection of saline and a similar improvement was shown. Months later, the American Medical Association denied any value in the drug and the patient relapsed once again returning to hospital, lost all hope and died within 48 hours.

In these two cases, the beliefs of the experimenter are argued to have affected the responses of the subjects, especially in the second case where the doctor convinced the patient that the drug was effective and despite receiving a saline injection, the patient went in to remission. Transference of belief to patient was intentional however, but shows the power of subject expectancies.

### **Expectancies and the Pygmalion Effect**

Rosenthal and Jacobson (1968) explain experimenter effects in terms of the 'Pygmalion phenomenon'. This was resulting from experiments from 'Oak School', where teachers were led to expect a certain group of children to have enhanced performances. In fact no such characteristics were seen in the group, but they were randomly selected from the other class members. It showed that the teacher expectancies did in fact show enhanced performance from these groups of children, often twice as much as the others. Improvement was measured on an IQ performance type test, measured at several time points throughout the school year: pre-test, 4 months, 8 months and at 20 months. The maximum overall effect was at 8 months; however, the effect was still present at 20 months. In short, when teachers expect students to do well and show intellectual growth, they do; when teachers do not have these expectations, then the performance or growth of the students is not as great (Rhem, 1999). Rosenthal and Jacobson (1968) describe many similar studies suggesting that our expectations can strongly influence others around us, in any

setting and not just the classroom. These same effects are those argued that experimenter can have in behavioural research.

Rosenthal and Jacobson (1968) have related these phenomena to the placebo effect, in that it can be partially understood in terms of the healer's expectation for the efficacy of the preparation. Indeed the healer's self-fulfilling prophecy can be an important component of the placebo effect. Double-blinded placebo experiments are argued to be the antidote to this. Quoting another paper by Beecher comparing the effects of morphine to the effects of a placebo saline solution, no difference in pain relief was found between the two. Beecher had employed a double blind design and therefore Rosenthal argues that investigators were not affected by self fulfilling prophecy. The double blind methodology is not always used though, therefore leaving the results open to bias from experimenter effects.

### **Controlling Experimenter Expectancies**

Rosenthal (1966) describes various studies he and his colleagues carried out manipulating the expectancies of experimenters and showed various effects on results, but also offers many strategies to minimise such expectancy effects. Implementing such expectancy controls highlights ethical considerations as many involve deception, however, the deception in this instance is involving experimenters and data collectors. Placebo and double-blinded deceptions have shown themselves to be warranted for use by the greater knowledge, now gained from the drug actions. Rosenthal's own research has involved much use of experimental and expectancy effects employing deception and has encountered no harmful effects. Any factually erroneous information can and has been given to data collectors, quickly and cognitively, correcting earlier information and has been met with no hostility from the personnel involved, once given an explanation of why it was necessary. Hostility can be evoked on subjects; however this tends to be due to the manner and the personalised nature in which deception was introduced. Rosenthal offers that so long as the subjects and data collectors see that the deception has a rational motivation (e.g. for the sake of science), then they react with an appreciation for its need. Controlling strategies include randomising and calibrating for experimenter effect which would give a reference point for definition of accuracy for future results, sampling experimenters instead of data collectors as this would assess experimenter's accuracy whilst also

assessing bias and consistency. Sampling data collectors would help especially if large numbers of subjects are involved, but also allows the assessment of influence of the data collector on the results. If however, there is a low sample size, so small that one experimenter could collect all the data, this is all the more reason to sample experimenters.

### *Sampling Experimenters*

Sampling experimenters offers many benefits argues Rosenthal as it would reduce the number of subjects each experimenter must contact, thus reducing any potential biasing effects of the experimenter. Each experiment would need to be subdivided with respect to various concepts.

*Learning to bias* Experimenters can learn from repeated responses given by subjects and could therefore unintentionally influence subsequent responses. If fewer subjects were seen per experimenter, then this would offer fewer opportunities to learn from responses.

*Maintaining blindness* The more subjects seen, the more opportunities there are to guess groupings of subjects and then 'crack the code' so to speak, thus breaking down the blinded procedure.

*Early returns* If more experimenters were involved, and they could all be collecting data simultaneously, then data is gathered more rapidly, reducing the need for early returns to examine early data for effects, thus not allowing any bias to be put on these early returns.

If each experimenter was to run only one subject from each condition, this eliminates early returns and bias, however, it could also have its drawbacks, explains Rosenthal. Logistically, each experimenter would have to be trained in every condition and experimental procedure, however, the utility of the procedure, may outweigh the increased cost. That is, if the information gained was going to be of great importance, then, any extra cost, time or otherwise is surely justified.

### **Implementing expectancy controls**

In drug trials, one expects a difference between the outcomes of two groups, i.e. experimental group and control group. Rosenthal reasons that one therefore needs to understand the treatment effects unconfounded by experimenter effects

Given that experimenter expectancies is preconfounded with their experimental and control conditions, the challenge is how to create a counter expectancy. Rosenthal suggests various methods for achieving this, however, most necessitate deceiving experimenters through withholding information from them or giving them false information.

This again raises ethical issues, questioning whether deception is warranted by the importance of the results obtained as a result of the expectancy-controlled experiment. Rosenthal states that indeed most research carried out in the behavioural sciences should be expectancy controlled as virtually no scientific research can be described as too trivial to warrant adequate controls. When one examines the details, one either deceives the data collectors or risk producing results subject to serious error. Appropriate controls on experimenter expectancies, could therefore safeguard the integrity of research results. Expectancy control groups however, have greater implications on time and resources in the management and procedures in experiments.

In a simple two treatment between subjects experiment, to fulfil Rosenthal's recommendations, four groups are required and not just two. Table 1 below shows the general confounding of experimental treatment conditions with the experimenter's expectancy (Rosenthal, 1966).

**Table 1**

**Confounding of Treatments with Experimenter Expectancy**

		<b>Treatment Conditions</b>	
		<b>Experimental</b>	<b>Control</b>
<b>EXPECTANCY</b>	<b>Occurrence</b>	<b>A</b>	<b>B</b>
	<b>Non occurrence</b>	<b>C</b>	<b>D</b>

Cell A represents the condition in which the experimental treatment is administered to subjects by a data collector who expects occurrence of the treatment effect. Cell D represents the condition in which the data collector expects the non-occurrence of the treatment effect. Rosenthal argues that normally the investigator is interested in the treatment effects unconfounded by experimenter expectancy therefore the addition of

appropriate control groups will permit the evaluation of the treatment effect separately from the expectancy effect. A 'complete expectancy control' required the addition of both B and C, whereas a 'partial expectancy control' required only B or C. Subjects in cell B are those who will not receive the experimental treatment but who will be expected by the experimenter to show a treatment effect. Subjects in cell C however, receive the experimental treatment, but will be contacted by an experimenter who does not expect any treatment effect.

There are varying hypothetical outcomes from this type of analysis. A main effect attributable to the experimental treatment, a main effect attributable to the experimenter expectancy, or an interaction of them both, and can therefore be either, significant and large, significant and small or insignificant and almost zero.

Rosenthal (1968) explains seven different methods of inducing experimenter expectancies, all based on the model in table one above, having effects on the combination of cells A, B, C and D. These include: ascribe subject characteristics; ascribe experimental conditions; disparagement of treatment effectiveness; theory reversal; intentional influence; unintentional communication and subject responses. A description of some include:

*Ascribing subject characteristics* Subjects would be described to their experimenters as having certain characteristics, such as their response would be similar to those in cell A, similarly, subjects in cell C are described as sharing similar characteristics to those in cell D.

*Ascribing experimental conditions*, one can label the treatment conditions, where the experimenter does not administer the experimental treatment s/he can be told group B will receive the experimental treatment and C the control.

*Intentional influence* is one of the few that does not require deception. Experimenters instead are asked to intentionally influence the subjects in groups B and C, but still giving the same instructions as those to groups A and D respectively. However, this can lead to a lack in asymmetry between groups B and C with cells A and D. I.e. only two cells are being intentionally influenced and not all 4.

*Subject responses* are another method for creating expectancy control, but it doesn't involve creating expectancies in experimenters, but is derived from studies of the early data returns, as mentioned earlier. Half of the cell A experimenters are provided

with results that disprove an effect, and thus appear more like cell C. Half the experimenters for cell D are similarly provided with disconfirming data, thus making them more like cell B experimenters.

#### *Varying Experimenter Numbers*

If only the one experimenter is available, we can then employ him in cells A and B, or C and D as appropriate. Incidentally, diagonal comparisons are shown by Rosenthal to be more misleading in their indications and therefore this is why A and B or C and D comparisons are preferred to A and D or B and C. (The precise details are not necessary for the purposes of this review). However, the initial assignment of experimenters is also of importance as a control measure.

#### **Experimenter Assignment**

Rosenthal's ideal is to use expectancy groups and would include taking a large and random sample of experimenters and assign them to the various sub-conditions of the experiment mentioned above. This would utilise the advantages of a large number of experimenters. However, it is still possible to make use of expectancy-controlled trials in the absence of such a large experimenter pool.

If only one experimenter is available then subjects in conditions A and D would be contacted together and then B and C together. Certain control methods of those named in the last section would then be employed to each experimental condition. If only two experimenters are available, several possibilities exist for dividing the cells. Where more than two experimenters are available, then a combination of these divisions mentioned above would offer the best strategy.

In any experimental situation, expectancy controls can therefore be applied by the investigator, with the ultimate design being determined by the nature of the research question and due consideration of the resources available. However, as we have already noted, subject expectancies can also manipulate research results. Therefore, controls have been devised for this phenomenon also.

#### **Controlling for Subject Expectancies**

The experimental example given by Rosenthal for controlling subject expectancies would not generally stand today, due to more complex research in this area in the

intervening years. Alcohol expectancies are more complex, built on individual experience and reaction to experience. However, taken at face value the method can be understood.

In alcohol consumption, subjects and experimenters alike can share common beliefs in alcohol consumption and its effects on verbal learning and performance. Half the subjects are randomly sampled to an alcohol drink condition and the other half are randomly sampled and allocated to a control or soft drink group.

Experimenters can still be controlled leading them to expect that half the subjects in each treatment condition are in the control group and vice versa. (I.e. cells (B and C) or (A and D) in table 1).

However, because subject expectancies confound the design, the ABCD design is open to several effects and interactions and therefore needs to be extended further. The double confounding of both experimenter and subject expectancies can be seen clearly from table 2 below.

**Table 2**

**Double Confounding of Treatments with Experimenter and Subject Expectancy.**

	EXPECTANCY		Treatment Conditions	
	Experimenter	Subject	Experimental	Control
Occurrence	Occurrence	Occurrence	A	B1
		Non occurrence	A1	B
Non occurrence	Occurrence	Occurrence	C	D1
		Non occurrence	C1	D

To fulfil the subject expectancy control measures, cells B1 and D1 involve a non-alcoholic beverage that has an alcoholic taste, and an alcoholic beverage that has a non-alcoholic taste is used in A1 and C1. The differing tastes offer different cue to subjects about the alcohol content, but verbal statements from experimenters can also help to vary performance expectancies.

This would lead to a 2x2x2 analysis of variance, but the model also allows for flexibility of analysis of all eight conditions and cannot be practically managed if too few subjects are available. It can in fact be cut in half. Rows 1 and 2 can be run and compared, where all experimenters expect effects, rows 3 and 4 showing none would expect effects at all of alcohol. Any one of these would be helpful argues Rosenthal, but none would permit a comparison of experimenter vs subject expectancy, only rows 2 and 3 would allow this comparison.

Therefore depending on the purpose of the research, this format gives great flexibility in order to decide which groups to run to give optimal results required. Rows 1 and 4 would allow generalisations of real life social drinking, rather than its chemical effects. This particular comparison would include the original A and D comparison in a standard non-expectancy controlled experiment.

If on the other hand only two groups could be run, due to subject numbers for example, any row could be used, although some are better than others, depending on the point of interest. Rows 1 and 4 would be preferred to 2 and 3 for ecological validity and even then, row 1 preferred over row 4, purely as this would be more practical.

### **Summary of Rosenthal's Control measures**

Rosenthal has therefore offered many methods of control, and how they can be combined.

By limiting the contact between experimenters and subjects in expectancy-controlled studies, we limit the opportunity for transference of the experimenter's expectancies onto subjects. Therefore by combining blind contact and expectancy control groups, this has special advantages states Rosenthal. It allows the assessment of the success of blinding of the contact i.e. affirming if the blinding procedures effective.

If blinding has been successfully maintained at minimal or low, there will have been no experimenter expectancies or any interaction involving experimenter expectancy. If any such effects were found, then the blinding process will have been ineffective, thus questioning the methodology.

### **Rosenthal's evaluation of Costs versus Utility**

By implementing the full range of experimental controls, the issue of deception and having to withhold information from data collectors and experimenters is far outweighed by the strength of the data that can be assuredly gathered.

Surprisingly, the costs would appear to be minimal also. No extra subjects are required, no extra time per subject is required and often no extra experimenters are required. However, the creation of expectancy control groups takes additional time and planning. If more experimenters are used, the training stage will be longer. It is argued that this extra time is measured in hours and minutes, rather than weeks or months and is argued to prove no real obstacle. Any financial costs would be relatively small as the same time to run subjects is taken, whether one experimenter or eight, therefore, the more experimenters, the fewer subjects seen by each one. Only the training phase would incur more costs with respect to more experimenters. As mentioned previously, utility can also be made of the shorter time taken to gather data if multiple experimenters can work simultaneously.

In all, the total cost of conducting expectancy-controlled experiments seems trivial, as compared to the increased knowledge gained from their use.

Why then, is there a lack of experimental evidence for the uptake of these methods in behavioural experiments? It is very difficult to find any published research in which one can see all these control methods in action.

### **Conclusion**

We have seen a progression from placebo as an experimental effect, an experimental tool and its use is shown to involve biasing effects, both of the experimenter and subject.

Many conclusions seem evident from the various reviews carried out by Beecher (1955), Kienle and Kiene (1997) and Hrobjartsson and Gotsche (2001, 2004). The main criticism must be that there is no consensus on a definition of placebo or a placebo effect. The overall impression is of an orchestra playing a symphony, but all the instruments are tuned to different keys, therefore no harmony is found between

any of the parts. Beecher is tuned with the violins to 'D', Kienle and Kiene to 'F' and Hrobjartsson and Gotsche to many keys simultaneously.

One of the key lessons to be learned is that of Beecher and how his results became regarded as scientific fact for so long. Indeed, why did it take 40 years to complete a review of his analysis? Looking at his citation list on Web of Science (07.12.05), over 750 published articles are listed as citing his paper. It would seem from examining the titles of available online-journals that favour did not start to waver from his view until well into the 1990's. Only then did titles start speaking of the 'dark side' of placebos, or the 'mystery' of placebos.

As Kienle and Kiene (1997) indicate, that it was not until so many results were being published that placebo effects were shown in up to 70% of the population in some cases, that serious doubt was cast on the reliability or validity of Beecher's analysis.

Beecher obviously became so enthralled by his experiences with the 'placebo' effects shown in the field hospitals during wartime that his own 'experimenter expectancies' clouded his judgement in analysing the 15 study results included in his analysis. Rosenthal would probably criticise the selective reporting of results to 'erring on the significant' (Rosenthal, 1966), when he selected the results that supported his own hypothesis. Little justification can be given in the case of cerebral infarction when 21% were reported as improving warranted reporting, but no mention was made of the 53% who died on placebo treatment. He was correct however, in being the first to identify pain as especially responsive to placebo. As Kienle and Kiene (1997) comment, some of the conditions were more significant than others. They continue to recommend that as the results were so mixed, future analysis should examine the effects of specific clinical conditions to placebo, rather than a meta-analysis of various conditions with one singular value given.

Hrobjartsson and Gotsche (2001) cited Kiene and Kiene's analysis of Beecher, however, did not consider the clinical conditions individually as recommended. It took them to their 2004 analysis to consider this and did in fact find significant effects for both pain and phobias.

It could be argued that too much attention has been paid to low quality studies. In both of the Hrobjartsson and Gotsche papers, criticisms are made to the data they had available to them. Initially, one could argue that the inclusion criteria they set for the analyses is fairly robust, however, they do not state a singular definition of placebo and therefore leave their analyses open to bias and artefact, which they themselves recognise.

Dependence has been on published results from other sources, where it could be argued that better practice would be to join ranks or become associated with a largish research centre and become involved in the design and analysis of a definitive study comparing double blinded measures, incorporating Rosenthal's expectancy controls. The range of treatments considered as placebos would therefore need to be narrowed and defined as one particular form. Perhaps several similar studies could be run, each with a different definition of placebo to satisfy all 'camps'.

On an ethical issue, the widespread use of placebo control groups in pharmacological research suggest there are no harmful effects of the deception necessary in double blinded studies, with the benefits far outweighing the costs. This is warranted by the greater knowledge of the drug action they have given, especially in placebo and double blind studies as patients and experimenters have reported alike.

Rosenthal (1966, 1999) and Rosenthal and Jacobson (1968) have offered much in the explanation of expectancies, whether teacher, experimenter or subjects themselves. The control measures of Rosenthal (1966) appear robust and comprehensive, however, are not practiced in general research. As he stated himself, convincing experimenters that they are capable and indeed do have expectancies and pass them on is an entirely implicit effect. Perhaps this is why this methodology has not been widely utilised. From the social situation where the experimenter meets the subject through to the gathering of data and how this is analysed, all these factors can be controlled in order to find the true measure of the experimental conditions. The difficulty in convincing researchers of the need to use them is that they are unaware of the influences themselves and how these are physically or otherwise transferred from person to person, by whatever means. Despite some believing themselves to be

entirely correct in their own beliefs and hypotheses, few of us like to think of ourselves as capable of such basic confounding behaviour.

The persistence of the idea Beecher's Powerful Placebo, even to this day in some cases, has meant that the double blind trial has been shown to be the most favoured research methodology in placebo-controlled trials.

Nitzan and Lichtenberg (2004) recently performed a questionnaire study on the use of placebo with community staff, physicians and GPs. 60% actively use placebos, with a similar number prescribing placebo once a month. 68% of these lead patients to believe they are receiving actual medication, with 94% finding placebos generally or occasionally effective. In such a case, with concerns growing for over prescribing of drugs, this perhaps is not such a shocking thought as it seems. The ethical question, however, is pertinent. How would one feel if one had gained relief when prescribed a certain 'drug', only to realise the doctor had prescribed a 'sugar pill' instead?

As to the future of research into the placebo itself, perhaps further research especially in the field of Evan's proposal that the placebo response is in some way linked to the immune system and to the release of endorphins, a similar reaction will take place and a common linkage be found between all the varying descriptions of placebo and common ground can be found in research. However, how scientific would this approach be? Better that focussed and robust trials showing integrity, advance the placebo phenomenon and how best to explore its responses with possible uses to advance scientific and behavioural research.

## REFERENCES

- Beecher, H.K. (1955) The Powerful placebo. *Journal of American Medical Association* 159, 1602 – 1606.
- Blackwell, B., Bloomfield, S.S., & Buncher, C.R. (1972) Demonstration to medical students of placebo responses and non drug factors. *The Lancet* 19, 1279 – 1282 in Kirsch 2002
- Carrol, R.T. (2002) The Placebo effect. <http://skeptical.com/placebo.html> accessed 31.05.04
- Carrol, R.T. (2003) Experimenter Effect. <http://skeptical.com/experimenteffect.html> accessed 07.12.05
- Chambers Dictionary* (2000). Edinburgh: Chambers Harrap.
- Day and Ederer (2004) in Machin, Day, Green (eds) Textbook of Clinical Trials, Wiley.
- Evans, D. (2003) *Placebo. The Belief Effect*. London: HarperCollins.
- Goddum, J.H. (1953) Walter Ernest Dixon Memorial Lecture: Clinical Pharmacology, Proc. Roy.Soc.Med. 47, 195-204 in Beecher, H.K. (1955) The Powerful placebo. *Journal of American Medical Association* 159, 1602 – 1606.
- Greenblatt, M. (1964) Controls in Clinical Research. Unpublished Paper. Tufts University School of medicine.
- Haas, H., Fink, H., & Hartfelder, G. (1963) The placebo problem. *Psychopharmacol. Serv. Cent. Bul.*, 2, 1-65, in Rosenthal, R. (1966) *Experimenter Effects in Behaviour Research*. New York: Merideth Publishing Company
- Hrobjartsson, A., & Gotsche, P.C. (2001) Is the Placebo Powerless? Analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 344 (21), 1594 – 1602.
- Hrobjartsson, A., & Gotsche, P.C. (2004) Is the Placebo Powerless? Update of a systematic review with 52 new randomised trials comparing placebo with new treatment. *Journal of International Medicine* 256 (2), 91 – 100.
- Kienle, G.S. (1995) Der Sogenannte Placebo Effekt; Illusion, Fakten, Realitat. Stuttgart, New York: Schattayer Verlag GmbH, in Kienle, G.S. & Kiene, H. (1997) The Powerful Placebo: Fact of Fiction? *Journal of Clinical Epidemiology* 50 (12), 1311 – 1318.
- Kienle, G.S. & Kiene, H. (1997) The Powerful Placebo: Fact of Fiction? *Journal of Clinical Epidemiology* 50 (12), 1311 – 1318.

Kirsch, I. (2002) Yes, there IS a placebo effect, but is there a powerful antidepressant drug effect? *Prevention and Treatment* 5, 22.

<http://www.journals.apa.org/prevention/colume5/pre005002i.html> accessed 03.12.04

Luchelli, P.E., Callaneo, A.D. & Zattoni, J. (1978) *European Journal of Clinical Pharmacology* 13, 153 – 155, in Kirsch, I. (2002) Yes, there IS a placebo effect, but is there a powerful antidepressant drug effect? *Prevention and Treatment* 5, 22.

<http://www.journals.apa.org/prevention/colume5/pre005002i.html> accessed 03.12.04

Nitzan, U & Lichtenberg, P. (2004) Questionnaire study on the use of placebo. *British Medical Journal* 329 (7472), 944-946.

Nordenberg, T. (2000) The Healing power of Placebos. FDA Consumer magazine

[http://www.fda.gov/fdac/features/2000/100\\_heal.html](http://www.fda.gov/fdac/features/2000/100_heal.html) accessed 31.05.04

Pfungst, O. (1911) *Clever Hans (the horse of Mr von Osten): A contribution to experimental, animal and human psychology*. Translated by C. L. Rahn.). New York: Holt.

Rhem, J. (1999) Pygmalion in the classroom. *The National Teaching and Learning Forum* Vol 8 (2). Accessed 07.12.05

Rosenthal, R. (1966) *Experimenter Effects in Behaviour Research*. New York: Merideth Publishing Company

Rosenthal, R. (1999) Covert Communication in Classrooms, Clinics and Courtrooms. *Eye on Psi Chi*. 3(1), 18-22.

Rosenthal, R. & Jacobson, L. (1968) *Pygmalion in the Classroom: Teacher expectation and Pupils' Intellectual Development*. New York: Holt, Rinehart and Winston, Inc.

Senn, S. (2003) *Dicing with Death: Chance, risk and health*. Cambridge University Press.

Wall, P.D. (1999) *Pain, The Science of Suffering*. Weidenfield & Nicolson.