THE POWER OF NONSPECIFIC EFFECTS IN HEALING: IMPLICATIONS FOR PSYCHOSOCIAL AND BIOLOGICAL TREATMENTS

Alan H. Roberts
Scripps Clinic and Research Foundation

Donald G. Kewman
University of Michigan

Lisa Mercier
Mel Hovell
Graduate School of Public Health, San Diego State University

ABSTRACT. We evaluate the hypothesis that the power of nonspecific effects may account for as much as two thirds of successful treatment outcomes when both the healer and the patient believe in the efficacy of a treatment. Five medical and surgical treatments, once considered to be efficacious by their proponents but no longer considered effective based upon later controlled trials, were selected according to strict inclusion criteria. A search of the English literature was conducted for all studies published for each treatment area. The results of these studies were categorized, where possible, into excellent, good, and poor outcomes. For these five treatments combined, 40% excellent, 30% good, and 30% poor results were reported by proponents. We conclude that, under conditions of heightened expectations, the power of nonspecific effects far exceeds that commonly reported in the literature. The implications of these results in evaluating the relative efficacy of biological and psychosocial treatments is discussed.

The issue of specific and nonspecific effects in psychiatric and psychological interventions continues to be a matter of intense interest and debate. Controversies involve both biological (Fisher & Greenberg, 1989a; Margraf et al., 1991) and psychosocial (Beutler,
1991; Luborsky, Singer, & Luborsky, 1975) interventions. A recent history of psychosomatic illness (Shorter, 1992) documents how patients present and shift physical symptoms to correspond to the current models of disease promulgated by the medical community, changing their clinical presentations according to current theory and social beliefs. The controversies involve not only diagnosis and treatment, but also clinical and experimental research (Fisher & Greenberg, 1989a; Orne, 1962).

The placebo effect, among the most ubiquitous of phenomena in medicine and surgery, is well recognized and documented (Beecher, 1955; Benson & Epstein, 1975; Benson & McCallie, 1979; Shapiro & Morris, 1978; White, Tursky, & Schwartz, 1985). Over 12 years ago Turner, Gallimore, and Fox-Henning (1980) published an annotated bibliography of nearly 1,000 articles and books on placebo effects and related phenomena. A more recent extensive review of the area (White et al., 1985) cites dozens of additional publications.

In 1955, Beecher published his seminal and influential paper "The Powerful Placebo." That paper is still among the most frequently cited, and it was undoubtedly responsible for the double-blind design becoming a universal standard for medication evaluation studies. Despite its signal importance, that paper left many questions unanswered and these are still being debated. The placebo effect is a phenomenon still in search of a model or theory (Kojo, 1988).

Double-blind studies of both pharmacological and psychological treatments are being questioned for a variety of inadequacies (Fisher & Greenberg, 1989a; Margraf et al., 1991; Parloff, 1986). It is commonly understood that nonspecific effects are adequately controlled for by double-blind and prospective randomized controlled studies. This understanding is based upon the assumption that active and inactive treatments are effectively concealed from both the doctors and the patients. In many intended double-blind studies, true blindness may not occur, and in most trials there is no adequate way to estimate the degree to which this bias has inadvertently been introduced (Fisher & Greenberg, 1989b).

Another potential bias in double-blind studies is the expectancy and belief of the therapist and the patient that the treatment being administered is an effective treatment (Benson & McCallie, 1979; Frank & Frank, 1991; Plotkin, 1985). The optimism, strong conviction, and persuasive abilities of the therapist concerning anticipated positive responses to a drug or a treatment has long been recognized as an important aspect of healing. To date, however, it has not been possible to isolate the relative effects of these variables in clinical trials and double-blind studies.

To the extent that doctor-patient expectations affect outcome, this effect should be optimum when both the doctor and the patient believe strongly that the treatment is efficacious. To maximize positive expectancy in studies of treatment effectiveness, both the doctor and the patient would have to be naive concerning the use of a placebo, and the doctor would have to believe that the therapy is efficacious, as is most often the case in clinical practice. The independent variable in a study of these issues would be the presence or absence of positive expectations by both doctor and patient, while the dependent variable would be the relative efficacy of treatment. The complexity of a study such as this aside, it is probably unethical, would violate standards for the informed consent of human subjects, and likely could not be funded.

In the area of psychotherapy there are between 200 and 400 schools, systems, and theories of therapy. Most reviews and meta-analyses, however, have failed to find significant differences among treatments (Beutler, 1991; Luborsky, Singer, & Luborsky, 1975), although most also concede that psychotherapies are effective. This unsettling finding has led to a search for factors that are common across psychotherapeutic orientations. Some
have taken the position that failure to find differences among therapies reflects the huge diversity of the interactive variables that are involved (Beutler, 1991). Others have argued that everything a therapist does is therapy and that placebo controls do not serve as a primary test of the efficacy of psychotherapy (Parloff, 1986; Wilkins, 1984).

These problems raise again the question initially asked by Beecher (1955) about the potential power of nonspecific effects. If the contributions of expectancy, belief, and persuasion are on the low side, say one third of the variance as estimated by Beecher and as most commonly cited in the literature, then it may be quibbling to raise the question of uncontrolled nonspecific effects as a primary source of bias. If, on the other hand, nonspecific effects can generate high treatment success (say two thirds or more), then these nonspecific variables could exert profound effects upon controlled as well as uncontrolled trials of both biological and psychological treatments. If a high enough proportion of reliable variance can be accounted for by these and other nonspecific variables, then there may be little room for specific differences among treatments to become evident.

Just how powerful are nonspecific effects when they include variables like expectancy, belief, and persuasion? The available literature is not responsive to this question. Beecher (1955) reported 26 studies with placebo response rates ranging from 15% to 58%. We have located only three other authors who have reported on the power of placebo effects independently of Beecher’s report.

Moerman (1983) reviewed 31 controlled trials of the treatment of acute ulcers with the drug cimetidine. Of a total of 916 patients treated with cimetidine, 76% were healed, but 48% of the 776 patients treated with placebo were also healed. For 18 of the 31 trials (more than half) there were no significant differences between drug and placebo outcomes. When the placebo trials alone were examined, the effectiveness of the placebo ranged from 10% to 90% with a mean of 48% effectiveness.

Evans (1974a, 1985, 1974b) has reported three times that placebo efficacy compared to analgesic efficacy is 56%. The data supporting this assertion regrettably have never been published in a refereed journal, and the supporting data (Evans, personal communication) are not available. Benson and McCallie (1979) cite variable figures in their review of placebo effects in angina pectoris. None of these reports addresses the important question of how powerful placebo affects may be when doctor and patient expectancies are maximized.

In this report, we reexamine the power of nonspecific effects under conditions where both the doctor and the patient believe that a treatment is effective even though it is not. Since data on noneffective psychosocial treatments are not available, we have chosen to review medical and surgical treatments that were thought to be effective at the time the treatments were reported in the literature. When the treatments were later demonstrated to be ineffective, results of the previous clinical trials provided a rough estimate of the power of nonspecific treatment effects when the beliefs and expectancies of both doctors and patients were maximized, compared to double-blind placebo-controlled studies. In the latter trials, where doctors and patients understand that some patients may be receiving ineffective treatments and where some doctors or patients feel able to break the code either accurately or inaccurately, the nonspecific variables in question are attenuated.

Our estimates will be coarse, but no cruder than previous estimates that do not enhance these variables. We attempt to compensate by selecting ineffective treatments methodically and in an unbiased fashion. We sample both surgical and nonsurgical treatments, recognizing that the examples we evaluate do not necessarily reflect a random sample of all possible ineffective treatments. To increase the reliability of our estimates we average the results across the treatments selected. We use only descriptive statistics and avoid inferential statistics in discussing our findings. Our assumption is that the estimates of
nonspecific effects derived from medical and surgical treatments will be applicable to psychosocial treatments if not actually underestimating them.

METHODS

Medical colleagues at Scripps Clinic and Research Foundation; the University of California Medical School, San Diego; and the University of Michigan Medical School were petitioned to provide us with examples of treatments, for any disorder, which were thought to be effective at the time they were reported but which later were determined to be ineffective. We received suggestions from 43 respondents. From these, a list of treatment modalities was compiled.

These suggestions were randomly assigned to investigators in California and Michigan. At each site, the suggested treatments were reviewed to determine if they met inclusion criteria for the study. In order for a treatment to be included in our study, it had to meet the following criteria: (a) strong positive reports by at least two groups of investigators; (b) at least one well-controlled negative report; (c) the treatment has been abandoned by the medical profession as ineffective; and (d) data from a major portion of positive studies are presented in a manner that permits reliable classification into categories of excellent, good, or poor outcomes. Treatments were rejected for study solely on the basis of not meeting one or more of these inclusion criteria and for no other reasons.

Examples of treatments reviewed but rejected for inclusion in this report include diethylstilbestrol for preventing spontaneous abortion and ultrasound for warts, both because data could only be dichotomized. Three treatments for multiple sclerosis were rejected: Decumarin and histamine treatments both lacked strong assertion studies, while low fat diet lacked a strong negation study and the assertion studies were reported by only one investigator (Swank, 1955). Other treatments evaluated were discarded for similar reasons.

For the treatments selected, a search was made for all clinical trials, controlled and uncontrolled. Each study was then classified according to strict criteria as “positive” or “negative.” The final list of treatments analyzed included glomectomy for treating asthma, levamisole for treating herpes simplex virus (HSV) lesions, photodynamic inactivation for treating HSV infections, organic solvents for treating HSV, and gastric freezing for duodenal ulcer.

The treatment results of each study were categorized and recorded in terms of excellent, good, and poor outcomes. When investigators reported more than three categories of outcome, they were reduced to three. For example, one report categorized results into “symptom free,” “excellent improvement,” “slight improvement,” “no improvement,” and “worse.” These were recategorized as excellent (symptom free and excellent improvement), good (slight improvement), and poor (no improvement and worse).

In some studies results were reported dichotomously as either improved or no change. Improved outcomes were then equally divided among excellent and good categories while the rest were categorized as poor. The rationale for this division into two equal-sized categories (excellent and good) is that the term improved covers the range of positive outcome from cured to slight improvement and could not be put reasonably into one category or the other.

When a significant number of patients were lost to outcome, the data were analyzed twice: once ignoring lost subjects and once including them as poor outcomes. Where the number of lost subjects was small they were classified as poor outcomes. All decision rules were in the conservative direction, in each instance reducing the number of positive outcomes and increasing the number of negative outcomes.

Each set of reports was evaluated independently by two investigators. If the initial
categorization was done in California the studies were sent to Michigan for independent analysis. Studies initially categorized in Michigan were independently retabulated in California. Where disagreement occurred, the studies were submitted to yet a third investigator. If the results of the third analysis agreed with one of the first two analyses, the agreed-upon results were used. If the third evaluation did not agree with either of the first two, then the most conservative of the three classifications was used.

For all treatment modalities, there was 100% agreement on the classification of papers as either positive or negative. For specific treatments, agreement among investigators concerning classification of outcome into excellent, good, or poor was very high, ranging from 83.3 to 100%.

Data were analyzed by a simple meta-analysis of the data using frequencies and weighted percents. The use of more extensive meta-analytic procedures is not warranted for data not randomly derived; inferential statistics were avoided. The analyses used were sufficient to evaluate the question being asked: Namely, what is the relative power of nonspecific effects when both healer and patient have increased expectations for a successful outcome?

RESULTS

Glomectomy as a Treatment for Bronchial Asthma

The removal of the carotid body or glomus to treat bronchial asthma was introduced by Nakayama in Japan and reported in the English literature in 1961 (Nakayama, 1961). Proposed for uncomplicated asthma, this surgical procedure was intended to correct a reflex disturbance in the bronchial system. The surgery exposed the carotid body, a small whitish mass at the bifurcation of the carotid artery. The carotid body was then dissected free from the adventitial mass of the external carotid artery.

Positive Studies. Nakayama initially reported on a series of 3,914 asthmatics (Nakayama, 1961). He reported that 2,535 or 65% of his patients showed moderate to marked improvement. Overholt (1963, 1962, 1961), who termed the operation glomectomy, felt that it was a cure for patients with intractable asthma. He performed the operation between 1958 and 1963 and reported on a series of 534 glomectomized patients. He observed “significant relief” in 47%, “some improvement” in 35%, and “no change” in 18% (Overholt, 1963).

Overholt’s report was followed by a number of uncontrolled clinical trials which reported positive response rates ranging from 51 to 95% (Halttunen, Laustella, & Tala, 1967; Kostelecky & Trnovec, 1971; O’Rourke & O’Rourke, 1964; Prusty, Thomas, Shankar, & Joseph, 1966; Sedwitz, Christoph, & Thomas, 1972; Winter, 1972). Table 1 summarizes the data from each of these studies which, together, reported on 5,976 glomectomized patients with 41% excellent, 30% good, and 29% poor outcomes (Combined Data I). Table 1 also summarizes the results of these clinical trials, omitting Nakayama’s 3,914 cases on the possibility that he might have exaggerated his outcomes (Combined Data II). Omitting his data increases rather than decreases the percent of excellent outcomes from 41 to 57%.

Negative Studies. Despite the overwhelmingly positive results reported in these studies, skeptics (Comroe, 1979; Segal & Dulfano, 1965) questioned the physiological basis of the procedure and raised concerns about possible adverse effects. Studies with varying degrees of design adequacy were published (Curran & Graham, 1971; Curran, Oser, Longfield, Broderick, & Culvahouse, 1966; Harjola & Kintner, 1968; Lynne-Davies, Mac-
TABLE 1. Glomectomy for Bronchial Asthma

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Excellent</th>
<th>Good</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nakayama (1961)</td>
<td>3,914</td>
<td>1,267.5 (32.4)</td>
<td>1,267.5 (32.4)</td>
<td>1,379 (35.2)</td>
</tr>
<tr>
<td>Overhold (1963)</td>
<td>534</td>
<td>251 (47.0)</td>
<td>187 (35.0)</td>
<td>96 (18.0)</td>
</tr>
<tr>
<td>O'Rourke &amp; O'Rourke (1963)</td>
<td>40</td>
<td>21 (52.5)</td>
<td>13 (32.5)</td>
<td>6 (15.0)</td>
</tr>
<tr>
<td>Prusty, Thomas, Shanker, &amp; Joseph (1966)</td>
<td>56</td>
<td>9 (16.1)</td>
<td>19 (33.9)</td>
<td>28 (50.0)</td>
</tr>
<tr>
<td>Haltunen, Laustella, &amp; Tala (1967)</td>
<td>29</td>
<td>12 (41.4)</td>
<td>6 (20.7)</td>
<td>11 (37.9)</td>
</tr>
<tr>
<td>Kostelecky &amp; Trnovec (1971)</td>
<td></td>
<td></td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Gp I</td>
<td>25</td>
<td>8 (32.0)</td>
<td>9 (36.0)</td>
<td>8 (32.0)</td>
</tr>
<tr>
<td>Gp II</td>
<td>181</td>
<td>41 (22.7)</td>
<td>90 (49.7)</td>
<td>50 (27.6)</td>
</tr>
<tr>
<td>Gp III</td>
<td>140</td>
<td>21 (15.0)</td>
<td>60 (42.9)</td>
<td>59 (42.1)</td>
</tr>
<tr>
<td>Winter (1972)</td>
<td>57</td>
<td>11 (19.3)</td>
<td>18 (31.6)</td>
<td>28 (49.1)</td>
</tr>
<tr>
<td>Sedwitz, Christoph, &amp; Thomas (1972)</td>
<td>1,000</td>
<td>810 (81.0)</td>
<td>140 (14.0)</td>
<td>50 (5.0)</td>
</tr>
<tr>
<td>Combined Data I</td>
<td>5,976</td>
<td>2,451.5 (41.0)</td>
<td>1,809.5 (30.2)</td>
<td>1,715 (28.8)</td>
</tr>
<tr>
<td>Combined Data II</td>
<td>2,062</td>
<td>1,184 (57.4)</td>
<td>542 (26.3)</td>
<td>336 (16.3)</td>
</tr>
</tbody>
</table>

Note. Combined Data II omits Nakayama's (1961) data; see text for explanation.

Neil, Couves, & Sproule, 1970; MacGowan, 1967; Marschke, Beall, Stern, & Murray, 1965; O'Rourke & O'Rourke, 1964; Phillips & Kintner, 1970; Segal & Dulfano, 1965; Takino & Takino, 1965), all showing the treatment to be no better than sham procedures. A single-blind, nonrandomized study by O'Rourke and O'Rourke (1963) reported a success rate of 58% for 72 patients, but the control group (N = 17) obtained an even better success rate of 64%.

Double-blind studies (Curran & Graham, 1971; Curran et al., 1966; Marschke et al., 1965) found no significant differences between treated and control subjects. While these studies used relatively small numbers of treatment and control patients, it was impressive that two thirds of patients in each group felt that the treatment had made a significant difference in their illness. In 1968, the American Thoracic Society stated that glomectomy had not proven to be an effective treatment for asthma, and the procedure was discontinued in all major clinics in the United States.

**Levamisole as a Treatment for Herpes Simplex Virus (HSV)**

The immunomodulating drug levamisole has been used extensively as an anthelmintic (Jose & Minty, 1980). Kint and Verlinden (1974) hypothesized that levamisole influenced host defense mechanisms by restoring the functions of macrophages when those functions were deficient. They therefore attempted to treat recurrent relapsing HSV with levamisole, believing that a deficient immune response might play a contributory role in these recurrent infections. Although Russell et al. (Russell, Brisson, & Grace, 1978) noted that the different aspects of the cell-mediated immune response with respect to HSV infections were controversial, several investigators decided to treat recurrent herpes labialis and herpes genitalis with levamisole (Adno, 1978; Jose & Minty, 1980; Kint, Coucke, &
### TABLE 2. Levamisole for Herpes Simplex Virus

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Excellent</th>
<th>Good</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kint</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study I (Kint, Coucke, &amp; Verlinden, 1974)</td>
<td>15</td>
<td>4 (26.7)</td>
<td>8 (53.3)</td>
<td>3 (20.0)</td>
</tr>
<tr>
<td>Study II (Kint &amp; Verlinden, 1974)</td>
<td>33</td>
<td>20 (60.6)</td>
<td>9 (27.3)</td>
<td>4 (12.1)</td>
</tr>
<tr>
<td>Symoens &amp; Brugmans (1974)</td>
<td>25</td>
<td>12 (48.0)</td>
<td>12 (48.0)</td>
<td>1 (4.0)</td>
</tr>
<tr>
<td>Adno (1978)</td>
<td>14</td>
<td>9 (64.3)</td>
<td>5 (35.7)</td>
<td>0 (0.0)</td>
</tr>
<tr>
<td>Queiroz, Kinue, Poli, Nogueira, &amp; Guerrero (1976)</td>
<td>19</td>
<td>11 (57.9)</td>
<td>5 (26.3)</td>
<td>3 (15.8)</td>
</tr>
<tr>
<td>Jose &amp; Minty (1980)</td>
<td>33</td>
<td>11 (33.3)</td>
<td>12 (36.4)</td>
<td>10 (30.3)</td>
</tr>
<tr>
<td>Combined Data</td>
<td>139</td>
<td>67 (48.2)</td>
<td>51 (36.7)</td>
<td>21 (15.1)</td>
</tr>
</tbody>
</table>

Verlinden, 1974; Kint & Verlinden, 1974; Queiroz, Kinue, Poli, Nogueira, & Guerrero, 1976; Symoens & Brugmans, 1974).

**Positive Studies.** Kint, Couke, and Verlinden (1974) first reported an uncontrolled trial on 15 patients. They stated that 27% of their patients had no further recurrences or markedly improved, 53% had decreased frequency and severity, and 20% were unchanged or worse. Later that same year they submitted a formal report of 33 patients with 61% symptom-free or showing excellent improvement, 27% slight improvement, and 12% no change or worse (Kint & Verlinden, 1974). Symoens and Brugmans (1974) reported on 25 patients, 96% of whom had good or better response to therapy.

Other uncontrolled studies (Adno, 1978; Jose & Minty, 1980; Queiroz et al., 1976) reported another 66 patients. Table 2 summarizes a total of 139 patients from these studies with 48% excellent, 37% good, and 15% poor outcomes.

**Negative Studies.** There followed four randomized, double-blind studies (Chang & Finmarra, 1978; Mehr & Albano, 1977; Russell et al., 1978; Spruance, Krueger, MacCalman, Overall, & Klauber, 1979). These studies, with Ns ranging from 28 to 75 patients, all reported no differences between treatment and control groups. Russell et al. (1978) found that 81% of the control group and 83% of the treatment group in his study rated their response as moderate or excellent. A final study by Spruance et al. (1979) found that 24 patients who received levamisole obtained a decrease in lesion duration and pain severity compared to those receiving placebo, but also an increased lesion recurrence greater than the control group.

Overall, double-blind studies did not support the impressive results obtained in uncontrolled trials. As a result, the use of levamisole in treating HSV infections was discontinued.

**Photodynamic Inactivation for Treating HSV Infections**

The use of photodynamic inactivation, also called photodye inactivation, for the treatment of HSV infections gained more popularity and caused more controversy than almost
TABLE 3. Photodynamic Inactivation for Herpes Simplex Virus

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Excellent</th>
<th></th>
<th></th>
<th>Good</th>
<th></th>
<th></th>
<th>Poor</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Felber, Smith, Knox, Wallis, &amp; Melnick (1971)</td>
<td>60</td>
<td>26 (43.3)</td>
<td></td>
<td></td>
<td>26 (43.3)</td>
<td></td>
<td></td>
<td>8 (13.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Felber, Smith, Knox, Wallis, &amp; Melnick (1973)</td>
<td>20</td>
<td>8.5 (42.5)</td>
<td></td>
<td></td>
<td>8.5 (42.5)</td>
<td></td>
<td></td>
<td>3 (15.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedrich (1973)</td>
<td>30</td>
<td>24 (80.0)</td>
<td></td>
<td></td>
<td>6 (20.0)</td>
<td></td>
<td></td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaufman, Gardner, Brown, Wallis, Rawls, &amp; Melnick (1973)</td>
<td>48</td>
<td>36 (75.0)</td>
<td></td>
<td></td>
<td>5 (10.4)</td>
<td></td>
<td></td>
<td>7 (14.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lefebvre &amp; McNellis (1973)</td>
<td>11</td>
<td>5.5 (50.0)</td>
<td></td>
<td></td>
<td>5.5 (50.0)</td>
<td></td>
<td></td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Data</td>
<td>169</td>
<td>100 (59.2)</td>
<td></td>
<td></td>
<td>51 (30.2)</td>
<td></td>
<td></td>
<td>18 (10.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

any other recent dermatologic therapy (Nuss & Aeling, 1977). The treatment was based upon in vitro studies showing that the virus could be inactivated by applying heterocyclic dyes followed by exposure to fluorescent light. The dye becomes incorporated into the virus' DNA during replication. The dye functions as a light absorber for photochemical reactions in which molecular oxygen is used and the DNA molecules become disrupted (Nuss & Aeling, 1977).

In treatment, the lesion was ruptured and then painted with 0.1% neutral red or proflavine dye. Allowing 5 min for the dye to be absorbed, the area was then exposed to a 15-watt fluorescent light bulb at a distance of 6 in. for 15 min. The procedure was usually repeated in 1 to 6 hours.

Positive Studies. The technique as a treatment for HSV was first reported in 1971 (Felber, Smith, Knox, Wallis, & Melnick, 1971). Of 60 patients treated by Felber et al. (1971), 87% said the results were better than anything they had ever tried before in terms of symptom relief and rapidity of healing. That same year, three more uncontrolled clinical trials asserted the effectiveness of the treatments (Friedrich, 1973; Kaufman et al., 1973; Lefebvre & McNellis, 1973). The results of these trials, shown in Table 3, ranged from 85 to 100% positive results (good and excellent outcomes combined).

Felber, Smith, Knox, Wallis, & Melnick (1973) conducted a double-blind study and reported that 85% of their treatment group showed symptomatic improvement and/or reduced healing time. Their control group of 12 subjects also showed 75% with symptomatic improvement and 50% with greater reduction in healing time. They discounted this placebo response, however, and based on the outcome of the 20 experimental patients (shown in Table 3), they asserted the treatment to be new and promising and called for further study.

Summarizing the outcome data shown in Table 3, 169 patients participated in the combined case series. Of these, 59% were reported to have had excellent outcomes and 30% good outcomes. Only 11% obtained poor outcomes.

Negative Studies. In a nonblinded, nonrandomized study, Taylor and Doherty (1975) found that the surface area of lesions treated with proflavine remained larger significantly
The Power of Nonspecific Effects

### TABLE 4. Organic Solvents (Ether or Chloroform) for Treating Herpes Simplex Virus

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Excellent</th>
<th>Good</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nugent &amp; Chou (1973)</td>
<td>6</td>
<td>3 (50.0)</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Pasricha, Nayyar, &amp; Pasricha (1983)</td>
<td>11</td>
<td>4 (36.4)</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Sabin (1975)</td>
<td>20</td>
<td>5 (25.0)</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Farrell &amp; Nesland (1977)</td>
<td>4</td>
<td>1 (25.0)</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Takagaki &amp; Jidoi (1981)</td>
<td>8</td>
<td>6 (75.0)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Combined Data</td>
<td>49</td>
<td>19 (38.7)</td>
<td>16</td>
<td>14</td>
</tr>
</tbody>
</table>

*a* Includes 1 patient lost to follow-up. *b* Includes 7 patients lost to follow-up.

longer than the other two groups. They also reported that healing time was slower and HSV could be isolated for a larger period of time in the proflavine-treated group. Three randomized, double-blind studies (Kaufman, Adam, Mirkovic, Melnick, & Young, 1978; Myers, Oxman, Clark, & Arndt, 1975; Roome, Tinkler, Hilton, Montefiore, & Waller, 1975) then found no differences between treated and control groups.

Kaufman et al. (1978), in a well-designed study, compared 75 treated patients with 82 controls. They found no significant differences between treated and control groups. Interestingly, these are the same authors who were among those reporting impressive positive results in uncontrolled clinical trials shown in Table 3 (Kaufman et al., 1973).

### Organic Solvents (Ethyl Ether and Chloroform) for Treating HSV

Another failed treatment for HSV is the topical application of organic solvents such as chloroform and ether. Studies in vitro had revealed the lipid-containing membrane around the HSV to be ether-labile. It was hypothesized that the membrane could be removed or rendered vulnerable to natural host defense mechanisms by applying ether.

The treatment consisted of the topical application of full strength ether or chloroform to the lesions. It was usually applied with a piece of cotton and left on for 5 min. The treatment was repeated several times during a 24-h period.

**Positive Studies.** Nugent and Chou (1973) first described their experiences treating themselves and a few colleagues in an anecdotal letter to the editor. They reported 100% success (N = 6) in preventing lesions from progressing beyond early stages and in decreasing the number of recurrences. Pasricha et al. (Pasricha, Nayyar, & Pasricha, 1983) treated 11 patients and reported that 4 had no further recurrences, 3 obtained a significant increase in the interval between episodes, 1 had a single recurrence, 2 experienced no change, and 1 was not followed.

In the ensuing 8 years, three more uncontrolled trials asserted the efficacy of this treatment (Farrell & Nesland, 1977; Sabin, 1975; Takagaki & Jidoi, 1981). These trials reported success rates of good or better ranging from 71 to 100%. Combining the results of all five assertion reports in Table 4, we obtain a total of 49 patients with 39% excellent results, 33% good results, and 29% poor results. If seven patients lost to follow-up are
excluded from the totals shown in Table 4, 45% of 42 patients studied were reported to have had excellent results, 38% good results, and 17% poor outcomes.

**Negative Studies.** Three negative studies were reported between 1977 and 1980 (Corey et al., 1978; Guinan, MacCalman, Kern, Overall, & Spruance, 1980; Taylor, Hendley, Greer, & Gwatney, 1977). Taylor et al. (1977) observed 41 patients in a single-blind study and found no significant differences in healing time between treatment and control groups. Corey et al. (1978), in a nonblind, randomized trial \(N = 71\), found that topical ether did not shorten the duration of lesions and did not prevent or delay recurrent episodes significantly more than control procedures.

Guinan et al. (1980) reported a double-blind study with no significant differences between treatment \(N = 25\) and control \(N = 26\) procedures. In addition, subjective assessment revealed that 76% of the patients in both treated and control groups were convinced of the efficacy of the treatment they received.

**Gastric Freezing for Duodenal Ulcer**

Wangensteen introduced gastric freezing for the treatment of intractable duodenal ulcer in 1962 (Wangensteen, Peter, Bernstein, et al., 1962; Wangensteen, Peter, Nicholoff, et al., 1962). The procedure was an outgrowth of gastric cooling, which had been used earlier in the treatment of patients with hemorrhage of gastric or duodenal ulcers (Wangensteen, Root, Jenson, Imamogha, & Salmon, 1958). While temporary reduction of chemical activity in peptic acid secretions achieved by the cooling procedure was accepted as a physiologically sound principle, the physiologic basis for gastric freezing has never been clarified (Hitchcock, Ruiz, Sutherland, & Bitter, 1966). Wangensteen, in his 1962 reports, asserted that “physiologic gastrectomy,” suppression of gastric secretions, could be achieved by a temporary period of gastric freezing. He based this conclusion on studies of both dogs and humans.

The apparatus used for gastric freezing was a low temperature gastrohypothermia machine. Cold absolute ethyl alcohol was circulated in a stomach-shaped balloon. The deflated balloon was passed through the mouth into the patient’s stomach with the patient in a sitting position. Once in place, the balloon was partially inflated with cooled absolute ethyl alcohol. When the patient reported the experience of fullness in his stomach (between 400 and 1100 cc), the balloon was left in place for approximately one hour. The balloon was then slowly emptied and then left in place until it had separated from the gastric wall. Additional details of the procedure are described in Peter et al. (1962).

**Positive Studies.** Wangensteen and his colleagues at Minnesota published numerous preliminary, interim, and follow-up reports between 1962 and 1966 (Alcock, Carpenter, Bernstein, Peter, & Wangensteen, 1963; Bernstein et al., 1964; Bernstein, McFee, Goodale, Madsen, & Wangensteen, 1963; Peter et al., 1962; Ritchie, Edlich, Breen, Molina, & Wangensteen, 1966; Wangensteen, Peter, Bernstein, et al., 1962; Wangensteen, Peter, Nicholoff, et al., 1962). The initial reports (Peter et al., 1962; Wangensteen, Peter, Bernstein, et al., 1962; Wangensteen, Peter, Nicholoff, et al., 1962) indicated that between 98 and 100% of patients experienced marked or complete relief immediately following treatment and for up to 6 weeks following freezing.

An interim report on 185 patients (Bernstein et al., 1964) with follow-up from a few weeks to a little over a year stated that 69% were asymptomatic. The final report of the Minnesota group (Ritchie et al., 1966) included 557 patients treated between 1961 and 1966. The outcomes of these patients are summarized in Table 5. Twenty-four percent
TABLE 5. Gastric Freezing for Duodenal Ulcer

<table>
<thead>
<tr>
<th>Study</th>
<th>Sample size</th>
<th>Excellent</th>
<th>Good</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ritchie, Edlich, Breen, Molina, &amp; Wangensteen (1966)</td>
<td>557</td>
<td>131 (23.5)</td>
<td>106 (19.0)</td>
<td>320 (57.5)</td>
</tr>
<tr>
<td>Gibbons (1966)</td>
<td>41</td>
<td>15.5 (37.8)</td>
<td>15.5 (37.8)</td>
<td>10 (21.4)</td>
</tr>
<tr>
<td>Combined Data</td>
<td>598</td>
<td>146.5 (24.5)</td>
<td>121.5 (20.3)</td>
<td>330 (55.2)</td>
</tr>
</tbody>
</table>

were asymptomatic and another 19% had minimal symptoms. The remaining 57% were symptomatic (N = 247), decreased (N = 23) or lost to follow up (N = 5). The authors concluded that gastric freezing represented a useful tool for the control of peptic ulcer disease.

Gibbons (1966) reported that 76% of his patients were "generally well" at follow-up, but stressed that some had been treated too recently for definitive assessment. Of the 20 patients who had been followed for more than one year, 70% had not undergone surgery and were reported as being well. The remaining 6 patients had either undergone surgery or were dissatisfied. Combining Gibbons' findings with those of the Minnesota group (Table 5) provides an overall sample size of 598 with 24% excellent, 20% good, and 55% poor outcomes.1

Negative Studies. Of the seven negative studies found in the literature, three were uncontrolled clinical trials (Barner, Collins, Jones, & Garlick, 1966; Hitchcock et al., 1966; McIntyre, Jalil, & Deitel, 1969), one (Lubos, Viril, & Klotz, 1966) was a nonrandom, nonblind, controlled study of 188 patients distributed into a treatment group (N = 120) and two control groups (Ns = 17 and 18), and three were controlled, double-blind studies (Harrell, Rose, Fordtran, & Friedman, 1967; Ruffin et al., 1969; Zikria, De Jesus, Cunnick, & Ferrer, 1967).

Ruffin et al. (1969) found no significant differences between treatment (N = 82) and control groups (N = 78) at the end of a 6-week follow-up. Dependent variables included pain relief, secretory suppression, number and severity of recurrences, or end point outcomes (perforation, obstruction, or surgery). They concluded that gastric freezing was no better than sham procedure. Harrell et al. (1967) compared 28 treated subjects with 24 controls who were followed for 18 months. Although there were statistically

1Since this review and analysis was completed, we became aware of a review article which cited 11 additional reports on gastric freezing that had escaped our literature search (Miao, 1977). Of these, 6 would have met our inclusion criteria for assertion studies (Artz, McFarland, & Barnett, 1964; Heineken, Rich, Griefinger, Stoll, O'Grady, & Manowitz, 1963; Karacag & Klotz, 1964; Manlove, 1965; Rose, Fordtran, Harrell, & Friedman, 1964; Scott, O'Neill, Snyder, & Shull, 1965). These six reported a total of 314 patients, with 187 classified as having excellent or good results. Had these reports been included in Table 5, the total of good and excellent results shown in that table would increase from 45 to 65%, further strengthening our hypothesis. The total percentages shown in Table 6 for all treatments combined, however, would remain essentially unchanged.
TABLE 6. Outcomes for all Positive Studies Combined

<table>
<thead>
<tr>
<th></th>
<th>Excellent</th>
<th>Good</th>
<th>Poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>2,784</td>
<td>2,049</td>
<td>2,098</td>
</tr>
<tr>
<td></td>
<td>(40.2)</td>
<td>(29.6)</td>
<td>(30.3)</td>
</tr>
</tbody>
</table>

significant differences between the two groups at 3 months, the benefits were short-lived and the need for subsequent surgery was not obviated.

Finally, Zikria and his colleagues (1967) reported a small double-blind study comparing eight patients treated with gastric freezing and eight patients treated with a sham procedure. While all 16 were asymptomatic initially, only one patient (a control patient) had not experienced recurrence of pain at the end of an 18- to 24-month follow-up. The employment of gastric freezing as a treatment for duodenal ulcer has been discontinued.

The Placebo Effect Under Conditions of Heightened Expectations

The data so far presented (Tables 1-5) summarize the outcomes of primarily uncontrolled clinical trials (all positive studies) when both the doctors and the patients have heightened expectations that the treatment will be effective. These five treatments have all been abandoned as ineffective following controlled trials. Despite our current understanding of the inadequacy of these five treatments, enthusiastic adherents reported good or better outcomes for patients ranging from 45% (see footnote 1) for gastric freezing ($N = 598$) to 89% for photodynamic inactivation ($N = 169$).

The combined results for all five treatments are summarized in Table 6. For a total of 6,931 patients treated by these five methods, 2,784 (40%) were reported to have had excellent outcomes, 2,049 (30%) good outcomes, and 2,098 (30%) poor outcomes. These summary data should be compared with our earlier hypothesis (Roberts, 1983) of one-third excellent results, one-third good results, and one-third poor results under these conditions of heightened expectations, and to the common belief that nonspecific effects will generate an average of only one-third positive outcomes.

DISCUSSION

A major purpose of this analysis was to estimate the power of nonspecific effects when the expectations of patients and healers were maximized. In contrast to controlled studies of efficacy, the examples selected for our estimates are based upon uncontrolled clinical trials where both therapists and patients had beliefs in and expectations of a successful outcome. This kind of situation more closely resembles what occurs in the day-to-day treatment of both psychological and medical problems.

Our examples are not a representative sample of the domain of noneffective treatments. While our method of selecting studies to evaluate was not biased to select treatments demonstrating higher than average nonspecific effects, the clinical problems and treatments we evaluated were not randomly sampled. An additional bias is the fact that clinical trials not showing positive outcomes are less likely to be reported.

Additional sources of potential bias include the subjective nature of some of the symptoms used as outcome measures or lack of rigor in the collection or analysis of data in these kinds of studies. Such biases, of course, are subsumed under the heading of nonspecific effects in all uncontrolled studies and may also influence better controlled trials.
In the current study, nonspecific effects including, but certainly not limited to, positive expectations and beliefs as well as the persuasive powers of the doctor and the situation accounted for good to excellent improvement in almost 70% of patients treated (Table 6). The data from our sample of 6,931 patients, who underwent five noneffective treatments, clusters around one-third excellent results, one-third good results, and one-third poor results when conditions maximize the expectations of both doctors and patients. We conclude that under typical treatment conditions of clinical practice, where the doctor and the patient usually have some faith in the efficacy of the treatment, nonspecific effects exert considerably more influence than commonly believed and reported in many controlled research studies.

We do not assert that the data presented reflect solely the influence of expectations, enthusiasm, or faith. Rather, the treatment outcomes observed result from the operation of a variety of uncontrolled nonspecific effects, only one of which is the belief in the efficacy of a treatment. Some nonspecific effects, such as the expectations of doctors and patients, which are emphasized in this report, enhance response to a treatment whether specific or nonspecific. Other situations, such as the double-blind study, may serve to inhibit response to both active and inactive treatments by reducing positive expectations in the healer and the patient. This is true for both psychosocial and biological treatments.

Some important nonspecific effects appear to act independently of treatment condition (i.e., active treatment or nonspecific treatment) to influence outcomes—for example, the natural history of a particular disease (i.e., spontaneous remission). While spontaneous remission is an important contributor to improvement in the studies we examined, it does not affect positive and negative studies differentially; it only becomes more apparent in controlled studies. Diseases such as HSV infection, which have high rates of spontaneous remission and long asymptomatic periods, occurring unpredictably and independent of treatment, are seen to have the highest rates of positive response to nonspecific treatments. On the other hand, some conditions, such as asthma, may be particularly vulnerable to the influence of psychological variables occurring independently in the patient's life.

The length of time between publication of initial positive studies and better controlled negative studies in the treatments we sampled ranged from approximately 4 years for levamisole and organic solvent treatments for HSV to approximately 7 years for glomectomy. These time intervals may reflect the normal time needed to conduct and publish controlled clinical trials. They may also reflect partly the length of time newly introduced treatments can be sustained by enthusiasm in the absence of specific treatment effects. It is not possible to know how many unreported patients were treated over the years by means of glomectomy, “physiologic gastrectomy,” or other nonspecific treatments.

Our data do not directly address psychosocial treatments but serve as a model for them. It is likely that the power of nonspecific effects in psychosocial treatments, which include expectations, beliefs, and persuasion, will not be less than the approximate one-third excellent results, one-third good results, and one-third poor results reported here. If this assumption is correct, then it may be almost impossible to demonstrate reliably significant differences when comparing one psychosocial treatment with another, with a "placebo" treatment, or with an active biological treatment.

The data reported here do not argue for or against the effectiveness of psychological treatments, any more than they are an argument for or against the effectiveness of specific medical or surgical treatments. Rather, they indicate why it has been so difficult to demonstrate the specific effectiveness of one psychological treatment compared to another. The large nonspecific variables may overshadow any specific differences that might exist.

A treatment outcome is not due solely to nonspecific placebo effects or to specific effects. A treatment outcome is always due to some interactive combination of specific
and nonspecific treatment effects. From a scientific perspective, however, it may be time to re-evaluate our approaches to the study of psychosocial treatments.

Dennis Rousseau (1992) recently described three characteristics of what he termed "pathological science": (a) The effect being studied is often at the limits of detectability or has a very low level of significance so that it is difficult to design experiments that test the effect reliably; (b) a readiness to discard prevailing ideas and theories; and (c) the near impossibility of devising and conducting experiments which can provide definitive answers to the questions being asked. These three characteristics are uncomfortably close to what has been occurring in the study of the relative efficacy of psychosocial treatments. The scientific study of whether successful psychotherapeutic outcome is a matter of technical skills, as specific factors, or interpersonal skills, as nonspecific factors, may be an example of pathological science. The results of the current study point in this direction.

In summary, under the treatment conditions described in this report, the average power of nonspecific effects appears to be considerably greater than the one third routinely reported in both the psychological and the medical literature. The treatment conditions leading to such high levels of nonspecific effects in some medical and surgical treatments are routinely the case for most psychosocial treatments. Such reasoning leads to the conclusion that there is little, if any, room for demonstrating differential effects among psychosocial treatments when positive expectations are shared by practitioners and patients. This observation may extend also to treatments provided by many unconventional practitioners. A major block to conducting research to resolve some of the issues raised by this report is the absence of a theory or model encompassing the phenomena of nonspecific effects. Such a theory or model is sorely needed.

REFERENCES


An article in the Minneapolis Star (Thorkelson, 1975) describing faith healing in Catholic and Protestant churches quoted the Rev. Francis MacNutt, a prominent Harvard-educated Dominican priest, as follows: "The priest said that about a third of the people he ministers to are healed, another third are 'noticeably improved' and the other third are unchanged."


