Healing as a Therapy for Human Disease:  
A Systematic Review

NEIL C. ABBOT, M.Sc., Ph.D.

ABSTRACT

Objective: To assess, from published clinical trials, the evidence for the use of healing as a complementary medical intervention in human disease.

Design: Limited to studies involving random assignment to a treatment group consisting of "healing," broadly defined, or to a concurrent control group. All randomized trials published up to the year 2000, were identified from MEDLINE, CINAHL, BIDS-EMBASE, the CISCOM complementary medicine databases and from bibliographic references of published articles. Copies of all published studies were obtained, data were extracted, and methodological quality (Jadad) scores were derived where possible.

Results: Fifty-nine randomized clinical trials (RCTs) were found comparing healing with a control intervention on human participants. In 37 of these, healing was used for existing diseases or symptoms (22 existed as fully accessible published reports, 10 as dissertation abstracts only, and 5 as "preliminary" investigations with limited evidential value).

The 22 full trials (10 reporting a "significant" effect of healing compared with control) constitute an extremely heterogeneous group, varying greatly in the method and duration of healing; the medical condition treated; the outcome measure employed; and the control intervention used. Many trials had a number of methodological shortcomings, including small sample sizes, and were inadequately reported. Only 8 studies (5 with a significant outcome for healing) had a maximum methodological quality score of 5, and in 10 studies this score was 3 or less. Two trials—both large scale and methodologically sound—were replicates, and each found a significant beneficial effect of intercessory prayer on the clinical progress of cardiac patients. Eleven of the 15 dissertation abstracts and pilot studies reported nonsignificant results for healing compared with control, a finding that probably reflects the relatively small sample sizes and the likelihood of type II errors.

The significant heterogeneity found in this group of trials makes categorization problematic and inhibits the pooling of results by meta-analysis or similar techniques to obtain a global estimate of the "treatment effect" of healing.

Conclusions: No firm conclusions about the efficacy or inefficacy of healing can be drawn from this diverse group of RCTs. Given the current emphasis on evidence-based medicine, future investigations should be adequately powered, appropriately controlled, and properly described. These future investigations would most usefully consist of: (1) pragmatic trials of healing for undifferentiated conditions on patients based in general practice and (2) larger RCTs of distant healing on large numbers of patients with well-defined measurable illness.
INTRODUCTION

There are several reasons why the anecdotes of self-selected patients or therapists represent poor evidence for the efficacy of healing. First, a number of "popular" therapies or diagnostic techniques have subsequently been proved ineffective for the purpose used; examples include laetrile as a complementary therapy for cancer and intravenous albumin as an orthodox intervention for critically ill patients. Second, animals—including humans—have amazing self-recuperative powers, even without the intervention of a healer, ensuring that most ailments are self-limiting. During World War II, Archie Cochrane—the de facto founder of evidence-based medicine—was the only physician in a prisoner of war camp in Salonica catering for some 20,000 prisoners (Cochrane 1984). The fact that only 4 deaths occurred, 3 of these from a "nonmedical" cause (i.e., shot by Germans), convinced Cochrane of the relative unimportance of therapy in comparison with the body's recuperative powers. Third, patient satisfaction per se gives no guarantee of efficacy. Although some of the "successes" seen by healers may indeed be based on the specific efficacy of healing—that is, the intentional channeling of energy through the healer from a source to a patient, which we are told is the crux of the healing encounter (Hodges and Scofield 1995)—some could be caused by other factors. These include the "Hallo-Bye" effect, in which politeness masquerades as improvement; the tendency for many ailments, such as low back pain, to resolve naturally over the short term; and the influence of additional elements (Fischer, 1971), such as counseling, inherent in the patient-therapist encounter.

Despite some controversy, it is becoming generally agreed that research into healing is both feasible and necessary. Though the core of healing is believed to be ineffable, mysterious, and indefinable, most healers accept that it ought to be possible, nevertheless, to measure by experiment the effect that healing has on clients. Similarly, although it is generally recognized that "evidence" has its limitations (Feinstein and Horwitz, 1997), hard evidence of effectiveness is increasingly required for therapeutic interventions. The establishment of the National Institute for Clinical Excellence (NICE) to oversee service-wide quality standards within the National Health Service, emphasizing clinical efficacy and cost-effectiveness of service provision, is likely to increase the pressure on therapies such as healing to expand their evidence base, preferably through randomized clinical trials.

In a previous review, Benor (1992) assessed the evidence for the effect of healing on living organisms. Few rigorous, controlled studies on human illness were available for inclusion, despite the fact that this category of information is of most concern to patients and health care providers. This study attempts to collate and review, to the year 2000, all the evidence available from randomized, clinical trials (RCTs) on healing as a therapy for human disease.

SYSTEMATIC REVIEW OF THE CURRENT EVIDENCE FROM RANDOMIZED CLINICAL TRIALS

Identification of studies

Searches were made of the MEDLINE, BIDS-EMBASE, and CINAHL databases for RCTs of healing on human subjects. In addition, the information contained in the specialist CISCOM database at the Research Council for Complementary Medicine (RCCM) was accessed. Copies of the original trial reports were obtained, and the reference lists from these reports were consulted for trials that might have been omitted from the databases. Other literature sources, such as the monograph by Benor (1992), were also consulted for references to possible RCTs.

After excluding duplication publications, a total of 59 separate randomized clinical trials of healing were identified using these methods. This total includes one trial described as "quasi-randomized" (Dixon, 1998).

In total, 22 of these studies were excluded from the systematic review, although they are referenced in the bibliography for completeness. This group included 15 trials in which healing was not performed on patient groups with identifiable treatable symptoms. These tri-
<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample size (no. of study groups)</th>
<th>Author's description of intervention or condition treated</th>
<th>No. of group (n)</th>
<th>Treatment duration (n sessions x length if known; duration)</th>
<th>Control (not known; duration)</th>
<th>Main outcome measure</th>
<th>Main outcome (treatment vs. control)</th>
<th>Were statistics (Y/N)</th>
<th>Jadad score (maximum: 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Byrd, 1988</td>
<td>393 (2)</td>
<td>IP Coronary disease</td>
<td>3-7</td>
<td>Daily prayers during hospital stay Daily over 28 days</td>
<td>Standard care (201)</td>
<td>Medical course to discharge</td>
<td>S</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Harris et al., 1999</td>
<td>990 (2)</td>
<td>IP Coronary disease</td>
<td>75</td>
<td>Daily for 10 weeks</td>
<td>&quot;Usual&quot; care (524)</td>
<td>Medical course to discharge</td>
<td>S</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Sicher et al., 1998</td>
<td>40 (2)</td>
<td>DH Advanced AIDS</td>
<td>40</td>
<td>Daily for 10 weeks</td>
<td>No distant healing (20)</td>
<td>AIDS-defining illness and severity</td>
<td>S</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Wirth et al., 1993</td>
<td>21 (1)</td>
<td>DH Postoperative gum pain</td>
<td>&quot;a group&quot; 6 x 15-20 min; over 6 h</td>
<td>No healer (21), crossover TT rub without intention (37); control rub (34)</td>
<td>VAS score for pain</td>
<td>Anxiety (STAI)</td>
<td>S</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Simington and Laing, 1993</td>
<td>105 (3)</td>
<td>TT Anxiety (institutionalized elderly)</td>
<td>1</td>
<td>1 x 3 min</td>
<td>TT rub without intention (37); control rub (34)</td>
<td>Anxiety (STAI)</td>
<td>S</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Abbot et al., 2000</td>
<td>120 (4)</td>
<td>H Chronic idiopathic pain</td>
<td>5</td>
<td>8 x 30 min; over 8 wk</td>
<td>Mimic healing (30); no healing (30)</td>
<td>Pain scores (MPQ)</td>
<td>NS</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Beutler et al., 1988</td>
<td>115 (3)</td>
<td>PH Essential hypertension</td>
<td>12</td>
<td>15 x 20 min; over 15 wk</td>
<td>&quot;Distant&quot; healing (37); no healing (38)</td>
<td>BP level</td>
<td>NS</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Harkness et al., 2000</td>
<td>84 (2)</td>
<td>DH Peripheral warts</td>
<td>10</td>
<td>Daily over 6 wk</td>
<td>No healing (43)</td>
<td>Presence and size of warts</td>
<td>NS</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>Joyce and Wheldon, 1965</td>
<td>38 (2)</td>
<td>IP Range of chronic conditions</td>
<td>19</td>
<td>Daily x 15 min; over 6 mo</td>
<td>Standard care (19)</td>
<td>General clinical state</td>
<td>NS</td>
<td>Y</td>
<td>4</td>
</tr>
<tr>
<td>O’Laore, 1997</td>
<td>406 (3)</td>
<td>IP Undifferentiated, conditions preoperative anxiety</td>
<td>90</td>
<td>Daily for 94 days</td>
<td>No IP (146)</td>
<td>Psychology</td>
<td>NS</td>
<td>Y</td>
<td>4</td>
</tr>
<tr>
<td>Quinn, 1989</td>
<td>153 (3)</td>
<td>TT Preoperative anxiety</td>
<td>1</td>
<td>1 x 5 min</td>
<td>Mimic TT (NK); no treatment (NK)</td>
<td>Anxiety (STAI), BP, heart rate</td>
<td>NS</td>
<td>Y</td>
<td>4</td>
</tr>
<tr>
<td>Gagne and Tote, 1994</td>
<td>31 (2)</td>
<td>TT Anxiety (psychiatric elderly)</td>
<td>NK</td>
<td>2 x 15 min; over 2 days</td>
<td>Relaxation therapy (12); mimic TT (9)</td>
<td>Anxiety (STAI), motor activity</td>
<td>NK</td>
<td>Y</td>
<td>4</td>
</tr>
<tr>
<td>Gordon et al., 1998</td>
<td>25 (3)</td>
<td>TT Osteoarthritis of knee</td>
<td>1</td>
<td>6 (NK); over 6 wk</td>
<td>Mimic TT (11); standard care (8)</td>
<td>Pain score (MPI)</td>
<td>S</td>
<td>Y</td>
<td>3</td>
</tr>
<tr>
<td>Turner et al., 1998</td>
<td>99 (2)</td>
<td>TT Pain and anxiety</td>
<td>3</td>
<td>5 x 5-20 min; over 5 days</td>
<td>Mimic TT (37)</td>
<td>Pain score (MPQ)</td>
<td>S</td>
<td>Y</td>
<td>3</td>
</tr>
<tr>
<td>Dixon, 1998b</td>
<td>57 (2)</td>
<td>H &quot;Chronic symptoms&quot;</td>
<td>1</td>
<td>10 x 40 min; over 10 wk</td>
<td>Waiting list (24)</td>
<td>General symptoms scores</td>
<td>S</td>
<td>Y</td>
<td>2</td>
</tr>
<tr>
<td>Authors</td>
<td>Sample size (no. of study groups)</td>
<td>Author's description of interventiona</td>
<td>Condition treated</td>
<td>No. of healers</td>
<td>Treatment duration (n sessions × length if known; duration)</td>
<td>Control group (n)</td>
<td>Outcome measure</td>
<td>Main outcome treatment vs. control</td>
<td>Were statistics described?</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>-----------------------------------------------------------</td>
<td>-------------------</td>
<td>-----------------</td>
<td>----------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Keller and Bzdak, 1986</td>
<td>60 (2)</td>
<td>TT</td>
<td>Tension headache</td>
<td>1</td>
<td>1 session only</td>
<td>Simulated TT (30)</td>
<td>Pain scores</td>
<td>S</td>
<td>Y</td>
</tr>
<tr>
<td>Miller, 1982</td>
<td>96 (2)</td>
<td>RMH</td>
<td>Hypertension</td>
<td>8</td>
<td>NK; 6 × 20 min; over 5 or 6 wk</td>
<td>No healing (NK)</td>
<td>Systolic BP</td>
<td>S</td>
<td>Y</td>
</tr>
<tr>
<td>Peck, 1997</td>
<td>82 (2)</td>
<td>TT</td>
<td>Chronic arthritic pain</td>
<td>5</td>
<td>No healing (NK); Progressive muscle relaxation (37)</td>
<td>VAS score for pain</td>
<td>NS</td>
<td>NS</td>
<td>Y</td>
</tr>
<tr>
<td>Meehan, 1993</td>
<td>108 (3)</td>
<td>TT</td>
<td>Acute post-operative pain</td>
<td>1</td>
<td>1 × 5 min</td>
<td>Mimic TT (36); injection (36)</td>
<td>VAS scores for pain</td>
<td>NS</td>
<td>Y</td>
</tr>
<tr>
<td>Sundblom et al., 1994</td>
<td>24 (2)</td>
<td>SH</td>
<td>Chronic idiopathic pain</td>
<td>1</td>
<td>3–8 × 40 min</td>
<td>No active treatment (12)</td>
<td>Pain scores and VAS</td>
<td>NS</td>
<td>Y</td>
</tr>
<tr>
<td>Castronova and Oleson, 1991</td>
<td>37 (3)</td>
<td>H</td>
<td>Chronic back pain</td>
<td>NK</td>
<td>8 × 50 min; over 8 wk</td>
<td>Psychotherapy (13); no treatment (12)</td>
<td>VAS score for pain</td>
<td>NS</td>
<td>Y</td>
</tr>
<tr>
<td>Collip, 1969</td>
<td>18 (2)</td>
<td>IP</td>
<td>&quot;Leukemic&quot; children</td>
<td>10 families</td>
<td>Daily over 15 months</td>
<td>Standard care (8)</td>
<td>Survival</td>
<td>NS</td>
<td>Y</td>
</tr>
</tbody>
</table>

AH, absent/distant healing; AIDS, acquired immunodeficiency syndrome; BP, blood pressure; DH, distant healing; H, "healing"; IP, intercessory prayer; MPI, multidimensional pain inventory; MPQ, McGill pain questionnaire; NK, key information not reported; NS, treatment group not significantly improved compared with control, as reported by the authors; PH, paranormal healing; RMH, remote mental healing; S, treatment group significantly improved compared with control, as reported by the authors; SH, spiritual healing; STAI, state-trait anxiety inventory; VAS, visual analogue scale.

aAll trials, except Collip 1969, gave some form of description of the method followed by the healers.

bStudy described as "quasi-randomized."
<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample size (no. of study groups)</th>
<th>Author’s description of intervention</th>
<th>Disease/symptom</th>
<th>Control group (n)</th>
<th>Outcome measure</th>
<th>Outcome (treatment vs. control)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dissertation Abstract Only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attevelt, 1988</td>
<td>90 (3)</td>
<td>H and DH</td>
<td>Asthmatic conditions</td>
<td>Distant healing (30); no treatment (30); Mimic TT (8); Anxiety scores (NS)</td>
<td>Asthma symptoms</td>
<td></td>
</tr>
<tr>
<td>Bowers, 1993</td>
<td>20 (2)</td>
<td>TT</td>
<td>Preoperative anxiety</td>
<td>Mimic TT (NK); Anxiety scores (NS); routine care (NK)</td>
<td>Anxiety scores</td>
<td>NS</td>
</tr>
<tr>
<td>Hale, 1986</td>
<td>48 (3)</td>
<td>TT</td>
<td>In-patient anxiety</td>
<td>NK; Headache; BP; Anxiety scores (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MacNeil, 1995</td>
<td>10 (2)</td>
<td>TT</td>
<td>Tension headaches</td>
<td>Relaxation (NK); Psychologic parameters (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mersmann, 1994</td>
<td>90 (3)</td>
<td>&quot;Nonprocedural touch&quot;</td>
<td>Preoperative stress</td>
<td>standard care (NK); Quantity of milk (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parkes, 1986</td>
<td>18 crossover</td>
<td>TT</td>
<td>Lactation impairment</td>
<td>Mimic TT (18); No treatment (18) crossover; Anxiety scores (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quinn, 1978</td>
<td>60 (3)</td>
<td>TT</td>
<td>In-patient anxiety</td>
<td>Simulation with (20) and without (20) intent; Anxiety level (S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robinson, 1996</td>
<td>60 (2)</td>
<td>TT</td>
<td>In-patient anxiety</td>
<td>&quot;No TT&quot; control (30); Grief inventory (NK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sodergren, 1994</td>
<td>22 (2)</td>
<td>TT</td>
<td>Grief experience</td>
<td>Mimic TT (11); Symptom scores (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Preliminary Pilot Studies</strong></td>
<td></td>
<td></td>
<td>Post-chemotherapy sickness</td>
<td>Information group (NK); No treatment (NK)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland, 1998</td>
<td>80 (3)</td>
<td>TT</td>
<td>HIV-infected children</td>
<td>Mimic TT</td>
<td>Anxiety scores</td>
<td>NS</td>
</tr>
<tr>
<td>Olson et al., 1997</td>
<td>20 (2)</td>
<td>TT</td>
<td>High stress</td>
<td>No healing (11); Immunologic parameters (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olson and Sneed, 1995</td>
<td>40 (2)</td>
<td>TT</td>
<td>General anxiety</td>
<td>Control (NK); Anxiety scores (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walker et al., 1997</td>
<td>40 (2)</td>
<td>IP</td>
<td>Alcohol dependency</td>
<td>No prayer (18); Alcohol consumption (NS)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wirth and Mitchell, 1994</td>
<td>16 (1)</td>
<td>TT and IP</td>
<td>Diabetes mellitus</td>
<td>No healing (16), crossover; Insulin usage (NS)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BP, blood pressure; DH, ditant healing; H, "healing"; HIV, human immunodeficiency virus; NK, key information not reported; NS, treatment group not significantly improved compared with control, as reported by the authors; S, treatment group significantly improved compared to control, as reported by the authors; TT, therapeutic touch.
als were mainly on healthy volunteers in a laboratory or experimental setting. This group comprised a series of 5 replicated studies of healing of experimental dermal wounds (an overview of this series, which resulted in 2 positive and 3 negative outcomes for healing, is given by Wirth [1995]), and a further 10 trials with a variety of rationales and outcomes (Collins, 1983; Randolph, 1980; Hinze, 1988; Post, 1990; Van Wijk et al., 1991; Wirth and Cram, 1993, 1994, 1997; Wirth et al., 1997; Wirth et al., 1996). Also included in this group were 7 trials for which the abstract reports contained information too rudimentary for conclusions to be drawn and for which the original reports were unobtainable (Glasson, 1996; Green, 1993; Kemp, 1996; Kramer, 1990; Silva, 1996; Snyder et al., 1995; Woods et al., 1996).

The remaining 37 trials included 22 full trials for which a published paper was available in the scientific literature. The main characteristics of each of these reports are shown in Table 1. Ten additional trials had been performed as part of doctoral or master's degree dissertations and had not been subsequently published in the general scientific literature. Although the original theses are held in the universities of origin, informative abstract reports for these investigations were available from the RCCM via the Dissertation Abstracts or Masters Abstracts International service. Main details extracted from these abstracts are shown in Table 2. A final five trials were described by their authors as "preliminary" or "pilot" studies. These are also shown separately, in Table 2, although their evidential value is poor because their subject numbers were very small.

**Extraction of data from included studies**

Table 1 shows the data extracted from the individual studies. The "type of healing" and "medical condition treated" are presented as they were described by the authors of each paper. Because most of the studies did not designate a primary outcome measure before the start of the trial, Table 1 shows the "main outcome measures" (i.e., those of most relevance for the particular patient group, and those on which the statistics were reported).

The methodological quality of each study was rated according to the method described by Jadad et al. (1996), one of several possible methods that can be used to assess trial quality. By this method, 1 point is allocated for each of five methodological features relevant to good-quality clinical trial reports, namely, (1) the study was described by the authors as randomized; (2) the allocation procedure was described and was appropriate; (3) the study was described as "double-blind," defined for this review as patient and evaluator/assessor blind; (4) the procedure to ensure double-blinding was described and was appropriate; and (5) there was a description of withdrawals and dropouts from the study. The maximum score for an individual trial report is 5, and 1 point is deducted if the randomization method was inappropriate for the study or if the method of double-blinding was inappropriate. This score, though essentially crude, gives some indication of the consideration given by the authors to methodological issues.

During the extraction of data, the statement in the abstract concerning direction of outcome of each study—medical condition significantly ($P < 0.05$) or nonsignificantly ($P > 0.05$) improved by healing compared to a control intervention—was checked against the data in the relevant results section. Where there was a discrepancy, the outcome suggested by the results section was used and is presented in Table 1. Such a discrepancy was seen in two studies. In O’Laoire (1997), the "significance" reported in the abstract referred to changes in outcome measure from baseline rather than to differences between treatment and a control intervention (which were nonsignificant for the main outcome). In Gagne and Toye (1994), a positive result was suggested by the abstract when, in fact, no conclusion could be drawn about the effect of healing per se. These discrepancies emphasize the undesirability of relying on conclusions obtained from reading the abstract of a paper alone without referring to the data in the results section. This is particularly relevant for interpretation of the dissertation data in Table 2, which have been derived solely from the authors’ published abstracts.
Description of studies

The 22 studies shown in Table 1 (10 with a significant outcome, 11 with a nonsignificant outcome, and 1 with an undetermined outcome for healing) form an extremely heterogeneous group of trials. They varied greatly in number of healing treatments, their duration (from one 5-minute session only to one 15-minute session daily for 6 months), and the mode of application of healing, precluding estimates of dose equivalence or estimates of the dose effect across studies. Also, there was a large variation in medical conditions treated, and hence in outcome measures used. Pain, whether chronic or acute, was the single most commonly treated symptom (9 trials, 4 reporting a significant effect of healing). The range of “control” interventions was broad. Only 7 studies (Abbot et al., 2000; Gordon et al., 1998; Keller and Bzdok, 1986; Meehan, 1993; Quinn, 1989; Simington and Laing, 1993; Turner et al., 1998) used a “mimic healing” intervention, and 4 of these reported a significant effect of healing. Others used waiting-list controls or “comparison” interventions such as relaxation or psychotherapy, so the results seen could be compounded by nonspecific effects.

Some of the studies also exhibited methodological limitations. Table 1 shows that for the 22 studies for which a Jadad score could be derived, 8 had a maximum score of 5 points, 4 studies scored 4 points, and the remaining 10 studies scored 3 points or less. Of the 8 studies scoring the maximum of 5 points on the Jadad scale, 5 used distant healing or prayer (4 with a positive and 1 with a negative result). Two of these higher-quality (or more adequately reported) studies were replicates, a rare phenomenon in complementary medicine. Harris et al. (1999) was designed as a replication of the famous positive report by Byrd (1988), and it found a similar (although not identical), significantly positive effect of healing over control in a large group of cardiac patients. Of the remaining 14 studies, 10 used “nondistant” healing or therapeutic touch (5 with a significant outcome).

Few of the reports in Tables 1 and 2 mentioned sample size calculation in relation to a designated outcome measure, so it is not possible to assess whether, on the whole, the number of patients in a particular trial was adequate for a treatment effect to be seen. Indeed, only 3 of the 37 trials had a treatment group with more than 60 people. All of the studies in Table 2 and some in Table 1 (Colipp, 1969; Gagne and Toye, 1994; Joyce and Whelldon, 1965; Sundblom et al., 1994) had very small numbers of patients and were therefore subject to type II errors (obtaining a falsely negative result for a treatment that is, in fact, effective). It is therefore unsurprising that 11 of the 15 studies in Table 2 reported healing to have no significant effect, 3 did not state a clear result, and only 1 reported a positive outcome. Overall, there was no relation between group sample size and the direction of outcome of the studies.

Conclusions of systematic review

The total number of RCTs of healing for human disease found in this study was small. The MEDLINE and CISCOM databases combined contained approximately 10.5 million entries. Approximately 3,700 of these were RCTs of the complementary therapies (Barnes et al., 1999), including 455 RCTs of acupuncture. Therefore, the 37 trials obtained (or 59, if all RCTs on humans are considered) represent a sparse evidence base given the relative popularity of this therapy. Eisenberg et al. (1998) reported that healing, broadly defined, had been accessed by up to 7% of the U.S. population.

The 37 RCTs reviewed here constitute a very heterogeneous group of trials. They differ greatly in type and duration of healing (i.e., treatment “dose”), in number of healers used and their method (i.e., treatment application), and in medical conditions treated. They also suffer from methodological inadequacies, such as small sample sizes and inappropriate designs that do not allow for consideration of nonspecific effects. This significant heterogeneity, which prevents meaningful categorization of trials, also precludes any overall conclusion about the efficacy or ineffectiveness of healing as a therapy, and it certainly inhibits the estimation of overall treatment effects by pooling techniques such as meta-analysis (Naylor, 1997).
Considering the 22 studies for which full scientific reports are available (as opposed to dissertation abstracts, the conclusions of which are unverifiable), the overall “tally” of 10 significant, 11 nonsignificant, and 1 indeterminate outcome for healing has little meaning in the absence of sample size calculations, predefined outcome measures, “optimal treatment” from a range of healers, and the explicit use of procedures for randomization and double-binding.

Although the number of RCTs is slowly accumulating, the overall conclusion in 2000 is similar to that found by Benor in 1992 and Dossey in 1993: Despite some intriguing observations, no firm conclusions about the efficacy or inefficacy of healing can be made from the evidence contained in the RCTs currently accessible in the scientific literature.

**IMPLICATIONS FOR HEALING AND FUTURE HEALING RESEARCH**

The inconclusiveness of the evidence from this systematic review is the norm for trawls of the published evidence in the complementary therapies, such as hypnotherapy for smoking cessation (Abbot et al., 1998) or acupuncture for low back pain (Tulder, 1999), and comes about because of the low priority given to such therapies by funding bodies and by orthodox scientists who have special skills to offer. Too little research has been done, and that which is published is too often ill-conceived, ill-reported, and ill-performed, often by experimenters with more enthusiasm than expertise.

It can take years of painstaking work to establish the effectiveness (or efficacy) of a truly “effective” therapy. A famous example is aspirin for the reduction of subsequent heart attack in patients with myocardial infarction. After the first RCT in 1974, which showed a “negative” result albeit associated with a trend in favor of aspirin, a further five RCTs were conducted between 1974 and 1980 using the same hard outcome measure (i.e., death). These too were “negative” individually, and it was only after a weighted overall effect from all six studies (representing 10,859 patients) was calculated that a reasonable estimate of the “true” effect of aspirin—a 23% reduction in death from heart attack—was established (Elwood, 1997). Because, at present, the “unknowns” involved in the healing encounter exceed those involved in swallowing an aspirin, it may be many years before the specific efficacy or otherwise of healing under particular conditions is established.

Research into healing is said to be complicated by difficulties not usually found in the orthodox therapies. These difficulties fall into three groups:

- those that arise because healing is not a “treatment” in the conventional sense (i.e., the patient gets the healing that is “needed,” at a variety of levels, mental and spiritual as well as physical, complicating outcome measurement)
- those concerning “delivery” of therapy (i.e., everyone has the capacity to heal, complicating the choice of “placebo” intervention)
- those concerning the involvement required of the patient (i.e., the patient must want to get well at a deep as well as a superficial level, complicating patient selection, and possibly, randomization to groups).

Although there are a number of arguments against the appropriateness of RCTs to test healing (Targ, 1997a, 1997b), most of the points raised are also problematic for orthodox medicine (Kleijnen et al., 1997; Mant, 1999). It is becoming increasingly recognized that none of these objections are insurmountable because current methodologies can be adapted to take account of most healer concerns. It is important, however, that healers and researchers agree on appropriate designs. The two “bottom lines” for evidence-based researchers are randomization of patients and blinding of participants (Kleijnen et al., 1997), and there are a range of possible designs that retain both without compromising “uniqueness” of healing.

There are two possible, and promising, areas for healing research. The first concerns the application of face-to-face healing/therapeutic touch for chronically ill patients under the care of the general practitioner. The cohort study by Brown (1995) and the subsequent quasi-randomised trial reported by Dixon in 1998 both showed impressive positive effects for the healing encounter. Indeed, general practice re-
search has many advantages as a test bed for healing. Because some 40% of general practice consultations involve “watchful waiting” (Mant, 1999), there is no shortage of willing subjects who require an intervention, often for undifferentiated symptoms and multiple conditions (comorbidity). Because healing is sometimes called an “undifferentiated” therapy that claims to act at many levels simultaneously, it seems particularly appropriate for this setting. Patients are also more likely to be representative of the population at large than those who volunteer for hospital-based clinical trials. The logical next investigative steps would involve true randomization, and would include comparisons with psychologic interventions (done without healing intent) as well as with waiting-list controls in order to assess the “pragmatic efficacy” (Gottzsche, 1994), and possibly the cost–benefit ratio, of healing in a population setting.

The other potentially fruitful area for research, as indicated by this review, is distant healing of large numbers of patients with a specific single medical condition. This design is also amenable to randomization and acceptable controlling. The study by Byrd (1988) and the replication of it by Harris et al. (1999) with a similar positive result, as well as that by Sicher et al. (1998) on patients with the acquired immunodeficiency syndrome, indicate that it may be possible to build up a large body of evidence using this model. Such randomized trials, with independent monitors and with analysis performed by those without a vested interest in any particular outcome, could provide powerful evidence of specific or “fastidious” efficacy (Gottzsche, 1994) if such an effect exists.

In conclusion, there are good reasons why healing needs evidence in the form of clinical trials to back up its claims. A systematic review of currently available RCTs presents an unclear picture and fails to provide convincing evidence for or against the efficacy of healing as a therapy for human disease or symptoms. Two possibilities for future healing research involve (1) pragmatic trials of healing for undifferentiated conditions on patients based in general practice and (2) larger RCTs of distant healing on large numbers of patients with well-defined measurable illness.

ACKNOWLEDGMENTS

The author wishes to acknowledge the many healers who took part in his clinical trials of healing at the University of Exeter. This review has been done in recognition of their efforts.

REFERENCES


Address reprint requests to:
Neil C. Abbot, M.Sc., Ph.D.
Department of Epidemiology and Public Health
University of Leicester
24-28 Princess Road West
Leicester, LE1 6TP
United Kingdom

E-mail: nca2@le.ac.uk