A SYSTEMATIC REVIEW
OF THE QUALITY OF RESEARCH ON HANDS-ON AND DISTANCE HEALING:
CLINICAL AND LABORATORY STUDIES

Cindy C. Crawford, BA, Andrew G. Sparber, RN, MS, CS, and Wayne B. Jonas, MD

Cindy C. Crawford is an investigator at the Samueli Institute for Information Biology, Alexandria, Va and Department of Family Medicine Uniformed Services University of the Health Sciences, Bethesda, Md. Andrew G. Sparber is a consultant with Nursing and Patient Care Services, Clinical Center, National Institutes of Health, Bethesda, Md. Wayne B. Jonas is director of the Samueli Institute for Information Biology, Alexandria, Va, and associate professor, Department of Family Medicine, Uniformed Services University of the Health Sciences, Bethesda, Md.

Purpose • To systematically review the quality of published experimental clinical and laboratory research involving hands-on healing and distance healing between 1955 and 2001.

Data Sources • Studies were identified through comprehensive literature searches on spiritual healing in MEDLINE, PSYCHLIT, EMBASE, CISCOM, and the Cochrane Library from their inceptions to December 2001.

Study Selection • We selected published randomized, controlled trials of spiritual healing (hands-on healing and distance healing) done in clinical and laboratory settings, all of which had been peer reviewed.

Data Extraction • Independent quality assessment of internal validity was conducted on all included studies using the comprehensive Likelihood of Validity Evaluation scale. Clinical and laboratory studies were analyzed separately and then subdivided into hands-on healing or distance healing interventions.

Results • A total of 45 laboratory and 45 clinical studies published between 1956 and 2001 met the inclusion criteria. Of the clinical studies, 31 (70.5%) reported positive outcomes as did 28 (62%) of the laboratory studies; 4 (9%) of the clinical studies reported negative outcomes as did 15 (33%) of the laboratory studies. The mean percent overall internal validity for clinical studies was 69% (65% for hands-on healing and 75% for distance healing) and for laboratory studies 82% (82% for hands-on healing and 81% for distance healing). Major methodological problems of these studies included adequacy of blinding, dropped data in laboratory studies, reliability of outcome measures, rare use of power estimations and confidence intervals, and lack of independent replication.

Conclusions • When laboratory studies were compared to clinical studies in the area of hands-on healing and distance healing across the quality criteria for internal validity, distance healing studies scored better than hands-on healing studies, and laboratory studies fared better than clinical studies. Many studies of healing contained major problems that must be addressed in any future research. (Altern Ther Health Med. 2003;9(Suppl):96A-104A)

H
Healing, whether hands-on healing or from a distance, has been practiced worldwide since the dawn of antiquity. This type of healing is inherent in many cultures and is thought to work because of the belief of members of these societies and the expectation of healing. However, cultures that practice these methods claim that spiritual healing recruits forces beyond belief and expectation. If there is more to healing than belief, such effects should be able to be isolated by appropriate methods such as randomization, blinding, and control conditions. Collectively, these are called internal validity and reflect the quality of clinical and experimental research. Systematically evaluating the internal validity of research on healing can help to determine whether its claims are more than belief.

Systematic reviews of published research are also an important step in bringing about a consensus of thought in an area of study. Astin conducted a review of the clinical efficacy of distance healing. He found that 57% of the trials demonstrated a positive treatment effect. He used a simple tool designed for reporting quality, but not internal validity of these studies. Astin defined distance healing as including spiritual healing, prayer, any form of healing from a distance—all with a conscious act of attempting to benefit another

Reprint requests: InnoVision Communications, 169 Serrano Rd, Suite 104, Fountain, CA 93025; phone, (760) 633-3910 or (866) 828-2962; fax, (760) 633-3918; e-mail, alternative.therapies@innovision.com.
person. Examples of these methods include Therapeutic Touch and Reiki. He used vote count for summarizing results and the review was not comprehensive. Benor conducted a review of spiritual healing research in both clinical and laboratory settings. He found that nearly two thirds of all the experiments he examined reported significant effects. However, he did not systematically rate the quality of these studies, nor did he focus on the role of belief and study design effects.

Developing working definitions in the field of healing research is a formidable task. Even defining what is “distance” may not be clear cut. An example is the categorization of methods, such as Healing Touch, where healing hands can be physically placed on the individual and/or can be raised several inches or even feet from the body.

After a review of the literature, we approximated a definition for inclusion of reviewed studies into hands-on healing or distance healing categories. We used the physical barrier of the room to distinguish these types of healing. We defined distance healing as any healing being transmitted from outside the treatment room or separated from the patient or healer by a door or mirror.

Rationale for Evaluating Hands-on Healing or Distance Healing in Clinical and Laboratory Studies

There is little debate that physical touch can create significant effects (note 1). Questions arise when healing touch or intention is not physical or even within a visible distance. Is it possible to evaluate the effect of healing both on the radish seed and the human subject? The purpose of our study is to evaluate the quantity and quality of clinical and laboratory research conducted on the effects of distance and hands-on healing. We were primarily interested in determining what types of bias exist and to what extent they exist in research on healing.

METHODS

Literature Search

A comprehensive literature search was conducted to identify studies of hands-on healing and distance healing. MEDLINE, PSYCHLIT, EMBASE, CISCOM, and the Cochrane Library were searched using the following key words: spiritual healing, intentionality, mental intention, energy medicine, subtle energies, faith healing, folk healing, prayer, therapeutic touch, healing touch, Reiki, healing, distance healing, hands-on healing, Johrei, psychic healing, and laying-on-of-hands. Review of bibliographies helped us assure that we had completed a thorough search of the literature. This search included studies published between 1956 and December 2001.

Study Selection

We selected studies that examined the impact of hands-on healing or distance healing on clinical conditions in humans, as well as in laboratory studies. Acceptable articles had to meet the following criteria for inclusion:

1. Random assignment between groups in the study;
2. Control interventions that used placebo, sham, or other “blindable” procedures;
3. Publication in peer-reviewed journals (no abstracts, theses, or unpublished articles);
4. Clinical condition being studied for the clinical set;
5. A healer being involved in the intervention;
6. English language; and

Assessment of Methodological Quality

The evaluation of study quality in systematic reviews has been extensively developed in clinical medicine, but is rarely used in laboratory research. We were primarily interested in determining to what extent belief and other biases were controlled for in studies on healing. We therefore used a scale that focused on internal validity called the Likelihood of Validity Evaluation (LOVE) scale. The LOVE scale was designed specifically to analyze complementary and alternative medicine (CAM) research. Its scoring sheet for evaluating the quality of clinical studies consists of 26 items (Checklist 1). Because of the differences between clinical and laboratory settings, the criteria scored for the laboratory studies consisted of a modified LOVE scale of 22 items (Checklist 2).

Scores for each study were established by evaluating how likely it was that the effects were due to the treatment. Quality criteria included the presence of controls, randomization, comparability, blinding, loss of data, outcomes measurements, statistical analysis, and reproducibility. We considered the percentage of affirmative answers as the raw score for internal validity. To assure interrater reliability, 2 reviewers independently evaluated a subset of studies, and their scores were compared using the Kappa statistic.

SPSS 10.0 statistical software (SPSS, Chicago, Ill) was used to analyze all data. Descriptive statistics were reported for each interval validity criterion for all studies and compared according to the category in which each study was placed.

RESULTS

Study Descriptions

We found 45 laboratory studies that met our inclusion criteria published in 12 journals between 1956 and 2001: of these, 6 focused specifically on parapsychology. A search for randomized controlled clinical trials recovered a total of 45 studies published in 35 different journals between 1956 and 2001. These 45 trials included a total of 8,455 study participants.

Study Quality

Laboratory

Quality scores for laboratory studies ranged from 73% to 95% out of a maximum of 22 quality variables, with a mean score of 18/22 (82%). The mean quality score for hands-on-healing in laboratory studies was 82% in a total of 25 studies, while the mean quality score for distance-healing studies was 81% in a total of 20 studies. These 2 interventions scored similarly according to overall internal validity quality.
**CHECKLIST Clinical LOVE scale**

Internal validity: How likely are the effects reported due to the treatment?

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>Comparison/control group present</td>
</tr>
<tr>
<td></td>
<td>Parallel groups (data collected at the same time)</td>
</tr>
<tr>
<td>Randomization</td>
<td>Randomization done and method described</td>
</tr>
<tr>
<td></td>
<td>Subject assignment blind</td>
</tr>
<tr>
<td></td>
<td>Who screened, allocated, and enrolled subjects described?</td>
</tr>
<tr>
<td>Group baseline comparability</td>
<td>Group comparability checked/balanced (sex, age, main prognostic factors)</td>
</tr>
<tr>
<td>Blinding</td>
<td>Any done?</td>
</tr>
<tr>
<td></td>
<td>Checked?</td>
</tr>
<tr>
<td></td>
<td>All possible done (patients, therapies, evaluators, analysis)</td>
</tr>
<tr>
<td>Loss of data</td>
<td>Loss to follow-up &lt;20% for all endpoints?</td>
</tr>
<tr>
<td></td>
<td>Loss to follow-up &lt;20% for any endpoint?</td>
</tr>
<tr>
<td></td>
<td>Contamination (other unequal treatment provided to groups)</td>
</tr>
<tr>
<td></td>
<td>Compliance &gt;70% for all endpoints?</td>
</tr>
<tr>
<td></td>
<td>Compliance &gt;70% for any endpoint?</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Clearly defined and explicit?</td>
</tr>
<tr>
<td></td>
<td>Reliability established (by reference or measured in study)</td>
</tr>
<tr>
<td></td>
<td>Sensitivity of measurements assessed?</td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>Power calculation done and sample size achieved?</td>
</tr>
<tr>
<td></td>
<td>Stopping rules or interim analysis described</td>
</tr>
<tr>
<td></td>
<td>Sample size &gt;100 total or &gt;30 per variable</td>
</tr>
<tr>
<td></td>
<td>Descriptive statistics used, effect size reported?</td>
</tr>
<tr>
<td></td>
<td>P value statistics used?</td>
</tr>
<tr>
<td></td>
<td>Confidence intervals provided?</td>
</tr>
<tr>
<td></td>
<td>Intention-to-treat analysis done</td>
</tr>
<tr>
<td></td>
<td>Multiple outcomes corrected or not applicable</td>
</tr>
<tr>
<td>Sample</td>
<td>Subject selection criteria clear and focused</td>
</tr>
<tr>
<td>Reproducibility†</td>
<td>Results consistent with other reports (&gt;2 independently done)</td>
</tr>
<tr>
<td>Total scores</td>
<td>Possible answers are Yes (1 point scored), and no or unknown (0 points scored). Scores based on percentage of Yes answers.</td>
</tr>
</tbody>
</table>

*LOVE indicates Likelihood of Validity Evaluation scale.
†Used only for comparison with laboratory research set.

**Human Clinical Trials**

Quality scores for the human clinical trials ranged from 39% to 96% out of a maximum of 26 quality variables, with a mean score of 18/26 (69%, SD=13) and the median was also 69%. The mean quality score for hands-on-healing was 65% in a total of 27 studies, and the distance healing mean quality score was 75% in a total of 18 studies. Thus, distance healing studies scored statistically significantly better on internal validity than hands-on healing studies in clinical studies conditions ($t_{43}=2.7, P=.009$).

Interrater reliability of the LOVE scale was excellent, with a Kappa score for rater agreement beyond chance of .94 between 2 independent investigators on a subset of studies.

Because of advances in research methodology, one might assume that current research would be of better quality than earlier works. Astin’s review of clinical studies validated this assumption for reporting. However, our analysis did not show any such trend for internal validity.

Clinical and laboratory studies were compared across quality scores to determine if there was a difference in the strength of the research. Because of previous modification of the LOVE scale for the laboratory analysis, only criteria that overlapped both scales were used. We had a total of 14 overlapping criteria. In addition, we compared hands-on healing versus distance healing studies according to whether the study was clinical or laboratory.
<table>
<thead>
<tr>
<th>Criteria</th>
<th>Evaluation</th>
<th>Yes</th>
<th>No</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls</td>
<td>Comparison/control group present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parallel groups (data collected at the same time)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reasonable and convincing control procedure (2 pts mock, 1 pt no mock, 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>nothing)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomization</td>
<td>Randomization done and method described</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group baseline comparability</td>
<td>Group comparability checked/balanced (eg, dose, animal size, age,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>growth or expression rates in controls)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding</td>
<td>Sample allocation (cells/animals) done blind?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Adequacy checked</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Analysis done blind?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss of data</td>
<td>Sample loss &lt;20%?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clear distinction between pilot and confirmatory data sets?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intervention</td>
<td>Clearly detailed description (dose, time, procedure)?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>Clearly defined and explicit?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reliability established (by reference or measured in study)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sensitivity of measurements assessed?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Objective measurement?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical analysis</td>
<td>Power calculation done and sample size achieved?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Descriptive statistics used, effect size reported?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$P$ value statistics used?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Confidence intervals provided?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple outcomes corrected or not applicable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reproducibility</td>
<td>More than one therapist within study</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Is this study an independent replication of a previous study by a different</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>investigator?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total scores</td>
<td>Possible answers are Yes (1 point scored), and no or unknown (0 points</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>scored). Scores based on percentage of Yes answers.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*LOVE indicates Likelihood of Validity Evaluation scale.

**Quality Assessment Across Study Design**

Table 1 displays a comparison of hands-on healing versus distance healing studies according to whether the study was clinical or laboratory based, while the Figure displays a direct comparison between laboratory and clinical studies. The Appendix lists the articles that we included in our analysis.

Controls

All studies had a control group present because this was one of our inclusion criteria for the study.

Randomization

All studies were randomized because this was one of our inclusion criteria for the study.

Group Baseline Comparability

We observed that group baseline comparability (GBC) was more frequently reported in laboratory (98%) than in clinical studies (75%) and that distance healing was more likely to report GBC (92%) than hands-on healing studies (81%).

**Sample Blinding**

It was more common for distance healing studies (87%) to report sample blinding (SB) than hands-on healing studies (65%), and laboratory studies utilized SB (79%) more often than clinical studies (74%).

Adequacy of Blinding

Checking for adequacy of blinding scored relatively low for both clinical and laboratory studies and was less likely to be conducted in clinical and hands-on healing studies (22%) than laboratory and distance healing studies (55%).

**Dropout Rates**

Laboratory studies more frequently had dropout rates below 20% for hands-on healing than distance healing studies, whereas
TABLE 1 Comparisons of hands-on healing and distance healing studies for clinical and laboratory studies according to internal validity criteria.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Hands-on Healing Studies, No. (%)</th>
<th>Distance Healing Studies, No. (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinical</td>
<td>Laboratory</td>
</tr>
<tr>
<td>Control group</td>
<td>27 (100)</td>
<td>25 (100)</td>
</tr>
<tr>
<td>Randomized</td>
<td>27 (100)</td>
<td>25 (100)</td>
</tr>
<tr>
<td>Group baseline comparability</td>
<td>18 (67)</td>
<td>24 (96)</td>
</tr>
<tr>
<td>Sample blinded</td>
<td>17 (63)</td>
<td>17 (68)</td>
</tr>
<tr>
<td>Adequacy of blinding checked</td>
<td>6 (22)</td>
<td>17 (68)</td>
</tr>
<tr>
<td>Sample loss &lt;20%</td>
<td>25 (93)</td>
<td>21 (84)</td>
</tr>
<tr>
<td>Outcomes clearly defined and explicit</td>
<td>25 (93)</td>
<td>25 (100)</td>
</tr>
<tr>
<td>Reliability established</td>
<td>16 (59)</td>
<td>18 (72)</td>
</tr>
<tr>
<td>Sensitivity of measurements assessed</td>
<td>19 (70)</td>
<td>19 (76)</td>
</tr>
<tr>
<td>Power calculation/size met</td>
<td>4 (15)</td>
<td>5 (20)</td>
</tr>
<tr>
<td>P value statistics</td>
<td>26 (96)</td>
<td>24 (96)</td>
</tr>
<tr>
<td>Confidence interval provided</td>
<td>2 (7)</td>
<td>16 (64)</td>
</tr>
<tr>
<td>Multiple outcomes corrected or not applicable</td>
<td>19 (70)</td>
<td>25 (100)</td>
</tr>
</tbody>
</table>

in just about all clinical trials, dropout rates were below 20%.

Outcomes
Almost all studies had clearly defined outcomes with laboratory studies (100%) more often defining single primary outcomes than clinical studies (94%).

Reliability of Outcome Measures
Establishing reliable outcome measures was far more common in laboratory (84%) than in clinical (52%) studies, but quite similar between distance and hands-on healing studies.

Sensitivity of Measurements
Hands-on healing and distance healing studies scored similarly on whether sensitivity of measurements were assessed. While distance healing clinical studies (50%) addressed this considerably less often than did laboratory distance healing studies (100%), hands-on healing laboratory studies (76%) addressed this only slightly more than hands-on healing clinical studies (70%).

Statistics analysis
Power Calculations. Power calculations were conducted in fewer than 30% of these studies overall. However, distance healing studies (34%) estimated power more often than hands-on healing (17%) studies, and clinical (38%) studies more so than laboratory (15%) studies.

P Values. Almost all studies reviewed used P values.

Confidence Intervals. Confidence intervals were used more often in laboratory studies (60%) than in clinical (12%) studies, but there was no difference between hands-on healing and distance healing studies (35% versus 37%).

Multiple Outcomes. Corrections for multiple outcomes were used for most studies when applicable. Laboratory studies (98%) scored higher than clinical studies (71%) in reporting this, and there was no difference between hands-on healing and distance healing studies (85% versus 84%).

Independent Replication. Independent replication is lacking in the majority of these studies and is needed to strengthen the quality of research.

COMMENT
In our systematic review of 90 randomized controlled stud-
ies, distance healing scored higher in quality than hands-on healing, and laboratory studies scored higher than clinical studies. In general, laboratory studies scored higher than clinical studies, reflecting better experimental controls. Trials with humans may contribute to lower internal validity scores. While it is more straightforward to conduct a laboratory study than a clinical study involving humans, it is easier to alter the results of laboratory studies and drop negative data as pilot data. Blinding issues are also much easier in distance healing studies than hands-on healing studies, which made them better able to control for expectation and belief.

This review has a number of limitations. We restricted our review to randomized controlled group studies. This made the overall quality of these studies high and perhaps not reflective of the overall wide range of the literature. Another problem was determining how to categorize the healing interventions. Authors in this field have their own individual definitions of healing, which confounded accurate categorization into hands-on healing and distance healing.

Previous reviews have not reviewed the literature using standard quality criteria for internal validity as we have done here. Astin used the JADAD scale to score the quality of reporting of some studies involved in his review. This scoring system involves only 3 criteria; randomization, double blinding, and withdrawals and dropouts. The LOVE scale is designed specifically for CAM research and covers many more internal validity quality items than the JADAD. We scored the clinical studies according to the JADAD scale and came up with essentially the same score as Astin's study (68% versus 72%). The JADAD scoring system is not designed for laboratory studies and so was not used in this study. Our search was more comprehensive than Astin's and we also included laboratory studies in our review. Benor reviewed a comprehensive set of studies, both clinical and laboratory-based; however, he did not systematically evaluate their quality.

Conclusion

Healing techniques have been practiced worldwide since the origin of human culture. Research on these practices is still in the beginning stages and needs to be expanded. Standard definitions of the healing techniques are needed, and quality should be improved. The current body of research on healing shows that studies of distance healing have higher quality than hands-on healing, and laboratory studies are better conducted than clinical studies in terms of comparability, blinding, reliability, sensitivity, confidence intervals, outcome measures, and replication. The main deficiencies in the field are the lack of
independent replication, inadequacy of blinding, dropped data in laboratory studies, reliability of outcome measures, and rare use of power estimations and confidence intervals. Attention to these criteria can improve our understanding of healing and its relationship to expectation and belief.

Note
1. The views, opinions, and assertions expressed in this article are those of the authors and do not reflect official policy of the Department of Defense, the Department of Health and Human Services, the National Institutes of Health, or the United States Government.

Acknowledgements
This project was supported by the LSR/Samueli Program on Energy Medicine Research, Samueli Institute for Information Biology. We would like to thank the deputy director, Ronald A. Chez, MD, for his helpful review of the manuscript.

References


Laboratory Papers

Only 29 publications are listed here because some had several experiments that were reviewed as separate studies in one paper.


21. Tedder W, Monty M. Exploration of long-distance psychokinesis: a conceptual replication of the influence on a


