

COMMENTARY ON “EXPERIMENTS ON
DISTANT INTERCESSORY PRAYER” IN
ARCHIVES OF INTERNAL MEDICINE

BY J. E. KENNEDY

ABSTRACT: Chibnall, Jeral, and Cerullo argued that research on distant intercessory prayer assumes intervention by a supernatural being and does not meet the basic requirements for science of having explanatory relevance and being testable. They believed that this research will not be successful because it is “attempting to validate God through scientific methods.” They also proposed that the weak effects in the existing studies may be due to multiple analyses in exploratory research. From a parapsychological perspective, attempts to investigate healing by God have similar methodological issues as attempts to investigate survival of death. The effects could be due to psi by the research participants. Distant healing studies are greatly complicated by the fact that there is a background prayer effort and motivation for paranormal healing for virtually all severely ill persons. This research implies that certain prayers are better than others and opens the door to ethical abuse by unscrupulous or misguided persons. At the same time, the hypothesis of multiple exploratory analyses does not explain the results of the existing studies. Distant healing research appears to be subject to the same weak, unreliable effects and susceptibility to experimenter effects as other types of parapsychological studies.

The article “Experiments on Distant Intercessory Prayer: God, Science, and the Lesson of Massah” by Chibnall, Jeral, and Cerullo (2001) may be a pivotal commentary in the medical literature that precipitates a decline of interest in research on intercessory prayer. It may also have important implications for other types of research on distant healing.

Chibnall et al. originally intended to design a study to investigate possible healing effects of intercessory prayer. However, after delving into the methodology and existing findings, they concluded that this line of research will not be productive.

Their evaluation discussed several issues that had also caused me to lose interest in initiating research on possible healing effects of prayer. As background for discussing their article, it may be useful to summarize my previous rationale for not pursuing the topic of research on intercessory prayer.

The first issue was the implication that certain types of prayer or prayers by certain individuals are more effective than others. Virtually all severely ill patients are the recipient of prayers, their own and/or from others. The idea that a “treatment” group has greater healing than a control group implies that certain prayers are better than others. I believe it

is inevitable that some people and organizations would attempt to profit from the alleged ability to provide more effective prayers. When a patient does not recover, will family members have reason to feel guilty that they did not pray enough or pay enough for special prayers to get a different result? I did not want to start down this path.

The second issue is that research on prayer has the same basic methodological difficulties as research on survival of death. Intercessory prayer involves the concept that a divine or supernatural being will produce the desired results. A more parsimonious and therefore more scientifically testable hypothesis is that the person praying produces the effects directly, without the need for a nonphysical entity. For prayer research, the scientific issues are compounded by the strong profit potential noted above.

The third issue is that I came to view prayer studies as basically another type of parapsychological test method that will have the same problems of unreliability as other psi research techniques. In my early years in parapsychology, I believed that application of new technology, statistical methods, and special psychological situations might produce a breakthrough for reliable psi effects. However, after three decades of watching research methods come and go and learning about similar patterns in prior decades, I have come to believe that accepting and understanding the intrinsically unreliable nature of psi may be the best strategy for progress (Kennedy, 2001). Of course, this idea has been raised by others before (e.g., Beloff, 1994; Pratt, 1978).

Chibnall et al. discussed topics that relate to each of these issues. A summary of the key points in their article is given below, followed by some comments from a parapsychological perspective.

SUMMARY OF KEY POINTS

Chibnall et al. started their study design efforts by trying to identify factors that may possibly influence effective prayers. Their list included type, duration, frequency, and intensity of prayer; number of people praying; and the training, experience, and spiritual worthiness of those praying. When they looked into information from existing studies, they found great differences among studies and little guidance for designing a study. The significance of identifying effective prayer techniques was brought into focus when they noted that Catholics have weekly prayers for all the sick, which presumably provides a background level of prayer for everyone. Chibnall et al. then confronted a key issue:

If prayer is a metaphysical concept linked to a supernatural being or force, why would its efficacy vary according to parameters such as frequency, duration, type, or form? . . . Why, then, attempt to explicate it as if it were a controllable natural phenomenon? . . . there is no reasonable theoretical construct to which to link prayer because of, we would argue, its very nature. (p. 2530)

Chibnall et al. concluded that the hypothesis of intervention by a supernatural being does not meet the basic requirements for science of having explanatory relevance and being testable. They recognized that there are other options for research: "Clearly, there are alternatives to testing for God's intervention, alternatives that are fascinating, useful, and, above all, consistent with a scientific approach without being theologically untenable" (p. 2532). They objected to using God as the scientific explanation for "mysterious findings."

They particularly found problems with the fact that in existing studies intercessors have often been blind as to who they were praying for, and generally had little link with the prayer targets. They believed that this feature requires intervention by God for success, and therefore is testing God rather than testing people.

In a section on Statistical and Measurement Issues, Chibnall et al. brought these issues down to a technical level. For example, what is the null hypothesis in this research? They proposed that it is "Assuming that God cannot heal at the bequest of human intercessors, what is the probability of getting these results?" (p. 2533). Again the testing of God is prominent. Further, they pointed out, "It makes no sense to conduct a power analysis and set the alpha to a certain level if the laws of probability can be rescinded at any time" (p. 2533).

In a related point, they raised the question of why weak statistical effects are found in this research instead of striking nonstatistical beneficial effects from the supernatural power. For example, the results of the study by Harris et al. (1999) were $p = .04$ with 990 patients. The value and divine purpose for such a weak effect is highly questionable. Chibnall et al. suggested that the existing results may be due to the methodological "crud factor." Many of these studies had numerous outcome measures with only a few measures showing any possible effects. For example, Byrd's (1988) study had 29 outcome measures, and Harris et al.'s (1999) study had 35. Correction for multiple analyses and data fishing is always a problem under these circumstances. Chibnall et al. suggested that the many outcome variables indicate that these studies were exploratory rather than well formulated.

In statements that perhaps summarize their overall position, Chibnall et al. wrote: "Science does not deny God, miracles, and the like, it merely neglects them . . . Science cannot actualize our spirituality, so why do we ask this of it?" (p. 2535).

COMMENTS

From the perspective of parapsychology, the main theme of Chibnall et al.'s article is addressing the same issues that occur with the study of survival of death. Attempts to investigate God and attempts to investigate spirit survival both try to infer the presence of nonphysical entities from observations of physical events or reports of mental events. However, the

results could be produced by psi from the people in the study. Very difficult problems result from attempts to justify a hypothesis that is not the most parsimonious explanation for the data. It is likely that future discussions of distant healing research will be conceptualized more in terms of psi rather than prayer.

Conceptualizing the results as psi is more scientifically testable, but many of the problematic issues remain. There is still a background effort and motivation for paranormal healing that must be considered. Problems in specifying how the healing effect is guided also remain. Identifying the source and mechanism of psi are pervasive problems, with experimenter effects perhaps providing the most parsimonious explanation and the clearest motivation for distinguishing the treatment and control groups in healing studies. In a related issue, the characterization of effective psi efforts (duration of effort, number of people, people with special abilities, etc.) remains a dilemma with significant ethical implications in healing research. At a minimum, cautions about the limitations of exploratory research findings should be prominently discussed to reduce the potential misuse of the findings.

In addition, from my perspective, Chibnall et al. took an overly black-and-white approach to scientific investigation of paranormal aspects of spirituality. They had an underlying assumption that divine phenomena would not have patterns that could be discovered with science. Of course, another spiritual perspective is that the world is a manifestation of God, and therefore the established physical laws of nature demonstrate that consistent patterns are the norm rather than the exception for divine manifestations. Chibnall et al. appeared to have a theological position that there is a great separation between humans and the divine. An alternative position has a more intertwined relationship between humans and the divine.

I have recently suggested that the slow, perhaps negligible, progress in parapsychology, combined with the evidence that psi sometimes seems to guide people (rather than people guiding psi), suggests that it may be time to consider the possibility that psi is at least sometimes guided by a "transcendent" aspect (Kennedy, in press). This proposal is based on the recognition that the failure to make progress with the most parsimonious explanation is justification for exploring hypotheses that are more difficult to investigate. Progress with this proposal assumes that there are patterns or relationships that could be identified as meaningful from a more holistic perspective. Ultimately, this is an empirical question.

The problems with multiple analyses in exploratory studies are well known to parapsychologists. However, the number of outcome measures in the prayer studies is larger than is typically found in parapsychological research and suggests extra caution.

At the same time, the possibility that the results are some type of "mysterious finding" may deserve greater consideration than the methodological crud factor suggested by Chibnall et al. My evaluation of the data in

Byrd's (1988) study indicates that the results for two of the outcome measures are significant at the .05 level even after conservatively correcting for 29 multiple analyses.¹ Further, Harris et al. (1999) described their single primary overall outcome measure as being predefined.

It may also be worth noting that the .04 significance level with 990 patients in the study by Harris et al. (1999) is exactly what would be expected with efficient, goal-oriented experimenter effects that produce the desired experimental outcome with minimal paranormal influence (Kennedy, 1994, 2001).

REFERENCES

- BELOFF, J. (1994). Lessons of history. *Journal of the American Society for Psychological Research*, **88**, 7–22.
- BYRD, R. J. (1988). Positive therapeutic effects of intercessory prayer in a coronary care unit population. *Southern Medical Journal*, **81**, 826–829.
- CHIBNALL, J. T., JERAL, J. M., & CERULLO, M. A. (2001). Experiments on distant intercessory prayer: God, science, and the lesson of Massah. *Archives of Internal Medicine*, **161**, 2529–2536.
- HARRIS, W. S., GOWDA, M., KOLB, J. W., STRYCHACZ, C. P., VACEK, J. L., JONES, P. G., ET AL. (1999). A randomized, controlled trial of the effects of remote, intercessory prayer on outcomes in patients admitted to a coronary care unit. *Archives of Internal Medicine*, **159**, 2273–2278.
- KENNEDY, J. E. (1994). Exploring the limits of science and beyond: Research strategy and status. *Journal of Parapsychology*, **58**, 59–77.
- KENNEDY, J. E. (2001). Why is psi so elusive? A review and proposed model. *Journal of Parapsychology*, **65**, 219–245.
- KENNEDY, J. E. (in press). Do people guide psi or does psi guide people: Evidence and implications from life and lab. *Journal of the American Society for Psychological Research*.
- PRATT, J. G. (1978). Prologue to a debate: Some assumptions relevant to research in parapsychology. *Journal of the American Society for Psychological Research*, **72**, 127–139.

¹ Using the Bonferroni method to correct for 29 analysis, the probability for an outcome must be $.05/29 = .00172$. Table 2 in Byrd's (1988) article reported that "intubation/ventilation" occurred for 0 of the 192 treatment patients and 12 of the 201 control patients. Fisher's exact test for this difference gives $p = .0003$, one-tailed, which is below the Bonferroni criteria. Similarly, "antibiotics" were given to 3 of the treatment patients and 17 of the control patients, which has $p = .0014$. The article reported a stepwise logistic regression and a rating scale that combined measures. However, the stepwise regression did not adjust for all the measures in the study, and the rating scale was done at the suggestion of a publication reviewer after the researchers knew which measures had the greatest differences (Gunnard Modin, data analyst for the study, personal communication, June 8, 1994). Therefore, the adjustment for multiple analyses noted here is a good indication of the overall significance of the results.