Harris et al. did not evaluate or comment on what appears to be the strongest statistical association in their study: 3.7% (18/484) of those in the prayer group were discharged within 24 hours compared with only 0.9% (5/529) of those in the usual care group (P<.005 by \chi^2 test if observations are independent). Since these discharges occurred before the intervention began (mean±SE, 1.6±0.16 days after admission), we are concerned that the statistical methods used by Harris et al., which assume independence of the observations, may not be appropriate for their data. While their article states that "new patients were randomly assigned," it is not clear whether the same person who was readmitted for a new episode would have constituted a new patient; Figure 1 of their article does not indicate that readmissions of the same patient were excluded. Since patient assignment was based on an (odd or even) identification number that never changed, readmitted patients would remain in the same treatment group. For example, a patient with an even identification number who was admitted several times during the study period and tended to stay in the coronary care unit less than 24 hours (or to have a low Mid America Heart Institute Cardiac Care Unit MAHI-CCU score) would always be in the prayer group. Statistical methods that require independent observations would not treat the data from this individual correctly. Thus, we believe that the authors should comment on the following: Were multiple admissions from the same individuals included in their analysis? What were the potential reasons for...
the significantly higher (preintervention) 24-hour discharge rate of patients in the prayer group?

Even if the data of Harris et al 1 are taken at face value, we think their conclusions that the study "suggests that prayer may be an effective adjunct to standard medical care" is exaggerated, given the statistical weakness of the data and lack of a scientific basis for the hypothesis. Out of 40 "post-prayer" outcomes examined in Tables 3, 4, and 5 of their article, only 3 were significant at P<.05, and none was significant at P<.02. While these outcomes were positively correlated with each other, from a statistical standpoint this finding is not unexpected once multiple comparisons are considered.

References


I wish to share a few comments on the study of Harris et al 1 that appeared in the October 25, 1999, issue of the ARCHIVES. The purpose of this study was to see if there was any scientifically measurable effect of remote, intercessory prayer on the outcome of seriously ill patients in the coronary care unit.

The analysis of Harris et al seems to indicate that the main effect of intercessory prayer was on physicians and their medical decisions and not on patient outcome itself. That the same outcome was achieved with fewer controversial medical interventions in the prayer group is a bit sobering. In reading this study, I asked myself: Why should God allow the patients who
received the remote, intercessory prayer to do better than the control group? Does God love those for whom strangers pray more than those who were randomly assigned not to receive their prayers? I was taught that God is not capricious and that faith is not a matter of scientific proof.

Perhaps the irony of this study is that the outcome was the same despite fewer interventions in the prayer group. Perhaps the real conclusion is that God's grace is greater than our skills and immeasurable by our tools. Like many before them, the investigators may have missed the real message of their "study": that despite our arrogance, God's omnipotence is beyond our ability to add or detract.

As suggested by Harris et al, 1 effective remote, intercessory prayer could be explained by one of two mechanisms. It might represent a miracle: the intervention of God in the physical world by a supernatural force in ways that are incompatible with natural law. It might also represent a form of telekinesis: the movement (healing) of an object (human body) at a distance (remotely) with thought or will (prayer) by an unknown natural force. Miracle or telekinesis has never been shown to exist by credible, replicable scientific experimentation.

Harris et al state that their purpose is not to speculate on mechanisms, but rather to convey results. This approach seems to miss the heart of the issue. It is the very improbability of the mechanism that raises doubts concerning the validity of the results. Goodman 2 has cautioned against overreliance on P values in assessing the efficacy of studies. He emphasizes that P values must be evaluated within the context of the prestudy probability of efficacy. For years, skeptics have warned that extraordinary claims require extraordinary proof. This
is another way of stating Goodman's theme that results that are inconsistent with a well-validated scientific precedent (low prestudy probability of efficacy) require a higher burden of proof (lower P value). Within this context, the study of Harris et al actually suggests that remote, intercessory prayer has no effect on outcome.

Harris et al draw an analogy between their study and James Lind's scurvy trials. If Lind's studies had been subjected to statistical analysis, I suggest that the P value would have been far more impressive. Such a P value would have probably justified a reevaluation of the then current theories regarding the mechanism of scurvy. However, Harris et al are not merely testing the efficacy of a medication. On the basis of a P value of .04, Harris and his colleagues are suggesting the need to reassess 500 years of scientific advancement in our understanding of how the physical world is organized.

As science has advanced, we have actually become more confident that the earth is round, that lemons cure scurvy, that no miraculous forces suspend natural law, and that unknown forces do not move objects from a distance. Rather than doubting the fundamental nature of the scientific worldview, shouldn't we be questioning the meaning of a P value of .04? Is it not more likely that the results of the study conducted by Harris et al have occurred by chance (1 in 25) or by bias rather than postulating a mechanism that requires a seminal paradigm shift in physics? Do not their results suggest the need to reassess our statistical methods for judging efficacy rather than the need to reassess the fundamental theories of science?

The study by Harris et al is a wonderful example of a P value out of context and out of control. It is out of context because of the failure to properly adjust for mechanistic improbabilities. It is out of control because of its propensity to encourage much pseudoscientific mischief.

References


The literature on religious activity and health outcomes is fraught with methodological difficulties. Regrettably, the article by Harris et al on the impact of intercessory prayer in the coronary care unit (CCU) continues this tradition. Evidence for the conclusion that prayer has an impact on clinical course in the CCU and "may be an effective adjunct to standard medical care" is weak at best. Although the intercessors were instructed to pray for a speedy recovery, the prayer and control groups did not differ in length of stay in the CCU or in the hospital, nor did they differ on the Byrd scale. They only differed on the unvalidated Mid American Heart Institute–Cardiac Care Unit (MAHI–CCU) scale constructed for the purpose of this study. The lack of construct validity raises serious questions about this finding.

On both the unweighted and weighted scales, the prayer group showed a slightly but significantly better clinical course (ie, lower scores) than the control group. The unweighted scale is completely meaningless, as the authors' own example illustrates: a patient who dies in the CCU (1 event) has a lower unweighted score than one who requires antibiotics, arterial monitoring, and antianginal agents (3 events). The significance of the group difference on the weighted scale assumes that it has construct validity (eg, that the need for an electrophysiologic study is 3 times as bad as the need for antibiotics; as the scale indicates). This is by no means clear. High interrater agreement (96%) on the scores for 11 randomly selected cases is not a substitute for construct validity. Raters also will agree substantially on hair color, but that does not make it a meaningful clinical index.

Finally, there is the significant ethical issue raised by the conclusion that prayer should be added to the list of medical interventions. All intercessors in this study were Christian. Should only Christian prayer be recommended? Should we conduct studies to determine if Christian prayer is more effective than Jewish or Muslim prayer? Religion does not need medical science to validate its rituals. To attempt this trivializes religion.

As we have indicated elsewhere, there is little doubt that for many people, religion brings comfort when illness strikes. This does not, however, mean that
medicine should take on religious practices as adjunctive treatments. To do so flies in the face of the vast majority of empirical evidence and raises serious ethical issues.

References


Extraordinary claims require extraordinary evidence. However, is the evidence sufficient for the claim by Harris et al 1 that prayer may be an effective adjunct to standard medical care? Patients in the coronary care unit (CCU) who were randomized to receive remote, intercessory prayer (plus usual care) stayed as long in the CCU and in the hospital as patients who received usual care only. Furthermore, there were no differences between groups on 34 clinical outcome characteristics, but the prayer group had 11% lower scores on a new, unvalidated summary statistic describing clinical CCU course. The only alternative explanation that the authors discuss is chance, which they consider unlikely given one statistically significant (P=.04) difference between groups. The authors do not realize, however, that by making 34 comparisons using separate t tests with [alpha] set at.005 and another 3 with [alpha] set at.05, the chance of finding 1 significant difference is not 1 out of 25, but 1-(1-.05)^3+1-(1-.005)^34 =0.14+0.16 =0.30,

almost 1 out of 3. 2 Furthermore, with groups of more than 400, the smallest differences become statistically significant.
Finally, the authors fail to consider further alternative explanations for their findings. Consider the following: As a clairvoyant and telepath, I was aware (unlike the patients in the CCU and the staff involved) that this study was going on. Wanting to take advantage of the careful registration of CCU courses, I have subsequently used my telepathic powers to influence the CCU course of the experimental group. Admittedly, the effect was a little weaker than I anticipated, but that should be attributed to the fact that this was my first transatlantic telepathy work. My influence has worked quite satisfactorily in a recent European trial that some people think was investigating a new analgesic. I wonder whether Harris et al have convincing arguments favoring their interpretation of their data over mine. They might point to the fact that more people believe in prayer than in my clairvoyant and telepathic powers. There were times, however, that everyone believed that the earth was flat, and everyone was wrong.

Which will it be in this study—prayer, telepathy, or a summary statistic of uncertain validity? I am willing to reveal that I will settle for chance.

References
