Prayer and health research: proxies, missed targets, and opportunities

As pesquisas em oração e saúde: substitutos, propósitos fracassados e oportunidades

Kevin L. Ladd[a], Bernard Spilka[b]

[a] Psicólogo da Religião, professor da Indiana University South Bend, South Bend - USA, e-mail: kladd@iusb.edu
[b] Psicólogo da Religião, professor da University of Denver - USA.

Abstract

The purpose of this article is to present a quick overview of the psychology of prayer and health research in the context of the United States and Europe. Its range is limited to primarily English literature. The conclusion is that the findings are thoroughly ambiguous. It is argued that this conclusion is fundamentally inevitable because, in most instances, the variable deemed prayer is actually a proxy for a wide constellation of beliefs and behavior. Suggestions are made for re-focusing the field on scientifically feasible, theologically sensitive studies of prayer.

Introduction

Academic papers often adopt the strategy of starting with their preferred position and concluding with caveats and problems. Though we have no explicit data to support our hunch, we believe this frequently means the latter sections are overlooked. To work against that possible trend, here we will begin with challenges that are confronting the prayer and health line of investigation and conclude with observations about the current state of knowledge. In this way, the state-of-the-art research will appear in bold relief against what remains to be done in order to advance our collective understandings.

General challenges and caveats

Writers of academic papers wrestle with a large number of critical decisions such as what material to include or exclude and how to interpret that background information. Since not every paper can cover the entire span of literature, authors make the attempt to include representative works to suggest the depth and breadth of extant subject knowledge.
Setting aside instances of explicit or implicit biases on the part of writers, it is fair to inquire about the potential for success of this enterprise. In general terms, one classic challenge is referred to as the *file drawer effect* (FDE) whereby investigations with null findings rarely find their way into publication. This causes an inherent bias in the literature available for review, potentially overestimating the strength and prevalence of findings. This form of bias is exacerbated when authors summarize the literature in terms of “pro” and “con” findings. This practice involves simply counting the number of studies that support versus fail to support any given hypothesis. The authors then declare the position with the most studies tallied as the intellectual winner.

At least in social psychology, recent years have shown investigators employing the tactic of reaching out via internet forums and with other “wide-net” approaches to solicit information about the extent of the FDE. The effectiveness of these efforts obviously depends on the willingness of the collective to provide these FDE data and the degree to which the request for information is circulated. It clearly has the benefit of representing a step in the right direction to dealing with this issue.

The latter problem of counting studies is arguably more difficult to overcome even though it has long been established as an inappropriate way to summarize a body of work (MEEHL, 1978, 1990). The challenge is that, on the surface, it appears to adhere to scientific and logical principles. This means that the influence of the presentation can be great, thereby introducing a significant amount of “noise” into the system, especially when appropriated by readers new to the area of study. The counterpoint to this troublesome practice is the employment of meta-analytic techniques. This manner of exploring a body of work involves not simply accepting authors’ conclusions, but looking critically at the methods and samples that constitute the backbone of potentially included studies. Once this vetting has occurred, statistical procedures are engaged to look at the data of the various studies with special attention given to overall effect sizes and similarly informative group-level indices. The practice is time consuming and requires a particular skill set; this means it does not happen with great frequency despite its great value.
Closely related to the FDE issue is the topic of replication. One of the hallmarks of scientific work is that findings of worth are relatively stable across time and should appear in multiple samples of similar composition. There are multiple ways to address the concept of replication. For instance, it is possible to perform an exact replication, repeating as closely as possible the precise conditions of the initial study. If the initial sample is sufficiently large, a quick and easy way to complete this sort of evaluation consists of randomly dividing the sample into multiple groups. Testing the same hypotheses in both groups offers an efficient way of demonstrating the consistency of findings. Alternatively, a conceptual replication is an option wherein the initial concepts are retained, but operationalized in a new fashion to add breadth to the understanding. Various factors work against replications, not the least of which is the premium placed on intellectual novelty. Often, an exact replication is disparaged as by deeming the effort to be a “mere” replication insinuating that the work is somehow intellectually substandard because the investigator could not develop her or his own thinking on the topic. Anyone who has attempted such work, however, can attest to the fact that sketchy method and procedure sections demand a great degree of insight and creativity in the attempt to re-create a research protocol.

More and more attention has fallen on the fact that a preponderance of research is centered on Western, White samples and problems. As this systematic bias is explored, a common practice is to take existing scales and translate them into other languages for use in additional cultural settings. While this has the appearance of moving the field forward, there are multiple issues surrounding these efforts. In essence, this export/import business provides information about the extent to which people in other cultures think, act, and feel like Western, White people. This variety of comparison, if undertaken intentionally, obviously can provide important knowledge. More often, however, the intent of the research is argued to “validate” the scales in other languages and cultures. What the practice is actually doing is demonstrating that Western, White concepts often are understood by other sorts of people who can effectively group items related to those
notions. Arguably the more critical information consists of beginning in indigenous contexts to develop relevant measures that address that unique experience. After this sort of work has been undertaken, authentic comparisons can be explored.

Psychology of prayer and health challenges and caveats

The above observations are applicable to a wide range of work, including the psychology of prayer and health. Below, we turn our attention to providing comments on practices and trends in this subfield of inquiry.

The problem of rationale

At the outset, it is helpful to consider why anyone would think there should be any relation between health status and the practice of praying. One of the first contemporary models was developed by Levin (1996) who conceptualized prayer mechanisms along two dimensions: location and physicality. In his system, the effect of prayers may be local (specific) or non-local (diffused); particular people or events may be altered or the world might become more peace-filled. Additionally, these effects might take forms that are either physical or non-physical in nature; broken bones might be mended or emotions might be modified. This model is helpful in many ways to think about prayer, yet it assumes the very pragmatic position that prayer does something at a level that can be experienced by the practitioner.

This is not a problematic stance for most believers who pray from the standpoint of faith. What is interesting about the position is that the sacred texts of various major traditions are much less explicit about how and when prayer might “do something” in some fashion. In short, the official theological positions and the beliefs of typical adherents do not necessarily align with regard to the outcomes of prayer. A bedrock position of traditions might be better characterized
as an assurance that the deity toward whom the prayers are directed will hear the prayers. Yet, across the actual texts, this hearing does not always guarantee any demonstrable response, or, if there is a response, the positive or negative nature of the response. For each “prayer equates to health” passage that might be cited, it is possible to cite an alternative “the ways of God are beyond knowing” text. In short, expectations that prayer and health somehow should be related are highly ambiguous and cannot be taken for granted on the basis of explicit theological statements.

**Why study only costly health conditions?**

Even a quick survey of the literature dealing with health and religion in general, or prayer more specifically, reveals an emphasis on relatively extreme health care concerns. Dominating the research are inquiries into high impact, high cost physical problems such as heart disease, cancer. Largely missing from the discussion is sustained consideration of less dramatic, but far more common health issues such as colds, minor cuts and abrasions, and so forth. The implication is that only expensive, challenging health issues are worthy of investigation in relation to faith. One could argue that this is because the relatively quickly passing ailments typically resolve with no medical intervention. But that argument only represents the one side of the equation. Just because these low-level, transitory health problems do not require high-cost interventions or the extended attention of physicians does not mean that people are not praying about them or otherwise engaging religious or spiritual resources at some level.

This focus, however, is not an accurate reflection of how many practitioners engage their beliefs. Among the most intense believers, their faith stance is not limited to only “major” life situations, but to “all” life situations. For these individuals, it makes little sense to suppose that prayer would be of use only in relation to life-threatening physical conditions; prayer and faith are applicable to events both large and small. From this perspective, the literature is artificially restricted and it is an
open question as to why this is the case. In order to adequately represent the full domain of the health–faith spectrum as it is lived by actual people, research will need to include more than complex, expensive, and chronic conditions that often do not respond to even the most currently advanced medical interventions. To limit investigations in that way suggests that faith can be adopted as a treatment strategy when conventional scientific approaches are not successful. That sort of co-opting of faith for pragmatic application is at odds with the reasons underlying religious and spiritual practices.

**Victim blaming and the prosperity gospel**

Another dilemma facing the current literature is the phenomenon referred to as “blaming the victim.” When interpretations are advanced that religion or prayer is influential in advancing (or retarding) healthcare for individuals, this raises the question of how those effects are operating. In essence, why are the forms of healing (or not) so divergent among people of ostensibly similar levels of faith? For instance, if the data show that “on average” those who pray recover more rapidly following invasive surgical procedures, what does this mean at the personal level for a particular person who prays devoutly, but does not recover quickly and perhaps even dies following the surgery? Was there a problem with the form or sincerity of praying? This is an awkward interpretive challenge that is typically sidestepped in papers that claim positive relations between health and faith. These interpretive moves are reminiscent of what the position of the controversial “prosperity gospel” that claims more pure spirituality literally generates more income, health, and happiness. Those who do not experience these specific forms of “blessings” are regarded (both implicitly and explicitly) as less spiritually attuned. Such understandings clearly have a great potential for harm. Far more careful attention needs to be devoted to setting findings into larger theological contexts, not to promote a specifically religious psychology of religion, but in order to properly set the psychological research in the participant’s contexts of belief.
What is the goal?

It is also a fair question to ask about the goal of health and faith research. If the intention is to provide strategies to increase health among people, how would this strategy proceed? It is clearly not feasible to assign individuals to conditions of belief as part of a healthcare initiative. It is even doubtful that counseling someone to deepen or intensify her or his religious or spiritual practice is a viable option. The difficulty is that the engagement of religious and spiritual disciplines is not a purely “rational” decision for the majority of people. More often, these activities are driven by intuitive forces that set themselves in distinct opposition to rational forces. Attempting to promote prayer for its potential health benefits is clearly impractical (or even impossible) and it disregards the essential impetus behind praying, substantially misrepresenting the practice.

Exploring the Literature

Having raised a number of concerns regarding the conceptualization of the health-religion-prayer nexus, we turn our attention to the English-language literature for a brief review of the representative topics and approaches. Readers interested in a more in-depth review may wish to consult Ladd and Spilka (2012).

Simply put, prayer serves many purposes and takes many forms. Despite its complexity, the dominant reason for praying is the occurrence of adversity, most often when ill health or physical problems occur (Bell, 2005). Prayers are offered by the affected person, family members, friends, and others such as fellow believers. Depending on the condition, polls indicate that up to 95% of Americans believe that God answers prayers and 90% assume that such prayers improve their health (Spilka; Ladd, 2013). Sloan and Bagiella (2000) who strongly advocate a hard scientific approach examined 266 research studies and found that only 45 or 17% supported the popular stance. In a later work, Sloan (2006) was highly critical of deficiencies in the design and analysis of most of the studies claiming health benefits for prayer.
Religion and medicine: immediate concerns

One does not usually consider traditional medicine as favorably disposed toward religion and prayer. Recent years have, however, witnessed a surprisingly strong movement by physicians toward the acceptance and utilization of religious and spiritual practices in their dealings with patients. This is evident in the development of what is termed “Complementary and Alternative Medicine,” a movement with its own journals, support from the National Institutes of Health, and the funding of research, workshops, and clinics to the tune of many hundreds of millions of dollars. Studies conducted within the last few decades have demonstrated a considerable readiness on the part of physicians to see potentially desirable roles and effects of prayer plus a willingness to be prayerfully involved with their patients (FRANCIS; ASTLEY, 2001; MONROE et al., 2003). A CBS Poll found that 63% of respondents felt “doctors should pray if asked” (CBS, 2009).

Prayers concerning health fall into two large categories: one’s own situation or those of others. The latter are subsumed under the rubric of intercessory prayer. The widespread prevalence of such personal and social needs and activities explains why exploration of this domain has been extremely popular for the last half century. Efforts to summarize a massive amount of research have been extremely frustrating as it evidences every conceivable relationship to health from negative associations to independence and finally some positive results (POWELL; SHABABI; THORESEN, 2003); this profound ambiguity of findings is in line with the above noted problem of rationale. In addition to that problem, serious questions regarding every aspect of study definition, design, and analysis plague this body of work.

Prayer and health by the affected

Regardless of the negative physical or mental state that one is experiencing, the likelihood is overwhelming that prayers will be offered to alleviate it. The research literature in this realm is extensive but, at best, weakly supportive of the notion that prayer may influence health.
Prayer and the cardiovascular domain

More prayer research has been reported relative to cardiovascular conditions than any other form of ill health. In part this may be due to the fact that cardiac problems are the primary cause of death in the United States (U. S. CENSUS BUREAU, 2008). Blood pressure (BP) is a significant part of this area and has been the object of research with regard to prayer.

A major concern, too often ignored by those tying religious considerations to cardiovascular issues, is the likelihood of measurement errors plus defective instruments particularly with BP. These are detailed in Spilka and Ladd (2013). There is also surprisingly little researcher agreement concerning the methodologies employed.

Findings specifically relating to prayer and worship either seem trivial or confounded with other factors promoting more healthful behavior on the part of religionists such as reduced likelihood of smoking, drinking, drug use, and so forth. Mormons clearly reflect such an orientation with associated positive life-enhancing behaviors (ENSTROM, 1989).

Two interesting studies illustrate certain problems in this research. Koenig and his associates (1998) studied almost 4000 older respondents and the BP observations favored a combination of public worshippers, pray- ers, and Bible readers. The findings could be seen as statistically “significant” but are quite weak at the practical level. In another large sample, Buck, et al. (2009) obtained results with diastolic BP opposite to what was anticipated. Employing large samples increases the likelihood of observing statistical significance but this is not the entire story. As a start, examination pre-study power analyses might stimulate thinking regarding over-powering sample sizes. Reporting of effect sizes will also provide much more detailed information concerning the nature of the findings than it is possible to discern from standard significance tests.

Shifting from BP to cardiovascular disease, we encounter the fact that intra-reader reliability and inter-reader agreement tend to be low (WESTON; BETT; OVER, 1976). In addition, cardiac stresses from lifestyle, jobs, home life and seasonal change have barely been examined. As already noted, those who are religiously active are more socially involved and less apt to engage in unhealthy behaviors (BRUMMETT et al., 2001;
UCHINO, 2004). This is largely ignored. Sloan (2006) and Sloan and Bagiella (2000) claim that none of 12 pertinent studies identified by Luskin (2000) meet criteria for acceptable research. In like manner, only four of 39 studies cited by Koenig et al. (1998) were considered passable.

Of relevance is the observation that 96% of patients awaiting cardiac surgery employed prayer and 70% felt it was their most helpful activity (SAUDIA et al., 1991). Depression is a concomitant of such surgery and is regarded as predictive of future coronaries and death (CARNEY et al., 1988). Ai and her colleagues (1996, 1998, 2000) observed two-thirds of these surgical patients used prayer. It was viewed as a therapeutic suppressor variable.

In sum, research in the cardiovascular area leaves more questions than answers. The role and place of prayer is far from clear and shortcomings in research are abundantly evident.

Prayer and cancer

In early stage cancer, of 11 coping strategies, prayer ranked ninth with only 9.4% of respondents indicating its use. Among advanced stage patients, it ranked first with use by 35.3% employing prayers (GOTAY, 1984). A survey of 17 studies reporting mixed results claimed that when hope is low and potential outcomes are poor, prayer is much more common (THUNE-BOYLE et al., 2006). Feher and Maly (1999) noted that 70% of their breast cancer patients valued their own and others' prayers. As already observed, religious groups such as the Seventh Day Adventists and Mormons who espouse healthful living practices show low cancer rates (HOFF et al., 2008).

Other conditions

The essential nature of research on prayer relative to ill-health is well illustrated by the foregoing ventures into the cardiovascular and cancer realms. Needless to say, similar efforts have been directed at HIV/aids, Multiple Sclerosis, Arthritis, renal disease, pregnancy, disabled children, old age, and other conditions with studies and findings similar to those
already mentioned. Theory, research design and analysis are becoming more sophisticated but there remains much room for development.

**Intercessory Prayer (IP)**

As common as it is for people with ill health to pray for themselves, some of the most interesting and controversial work has been undertaken on others who pray for those with medical problems. The modern era began with Galton (1983) who hypothesized that the common public prayers in England for the sovereign should endow her or him with an especially long life. When compared with 11 other privileged groups, he reported that the former “are literally the shortest lived of all who have the advantage of affluence” (GALTON, 1883, p. 282).

**Cardiac problems**

Except for a few small studies which produced negative results, work in this area was dormant until the seemingly supportive study of Byrd was reported in 1988. Before long challenging questions were phrased and the supporting edifice rapidly began to crumble. We provide a bit more detail concerning this study in order to dispel popular myths regarding its findings because, although deeply flawed, it is often cited without regard for the systemic shortcomings.

The general pattern in this type of research was to select a group of ill persons, randomly split it into experimental and control groups. The former would be prayed for by a number of pray-ers who would only know the first names of those for whom they offered prayers. The control group simply received the standard treatment for their condition. This was also provided to the experimental subjects. The “blind” conditions meant that the supervising physicians, the patients and the person conducting the research did not know who for whom the prayers were offered.

Byrd’s experimental and control groups had 192 and 201 cardiac patients. He initially compared them on 33 variables and reported no
significant differences. After the pray-ers performed their duties for individuals over a 10 month period, comparisons yielded claimed statistical significance on six measures, all in favor of those for whom prayer was offered. Closer inspection reveals multiple errors in this interpretation. For example, despite the large numbers of subjects, overall, the majority of variables pre and post contained less than 10 patients; some had none. Second, a Bonferroni analysis indicated that the claimed .05 level of significance for the 29 measures was actually the far more stringent .0017 and none of the differences were truly statistically meaningful. One of the claimed differences was the use of antibiotics with the control group showing greater use. Reference physicians indicated that this group simply contained sicker participants. Additional problems are detailed in Spilka and Ladd (2013).

The Byrd work stimulated a number of follow-up, similarly structured studies none of which yielded findings truly supportive or definitive relative to the effectiveness of IP. The definitive rejection of Byrd’s claims came from the work of Benson and his colleagues (BENSON et al., 2006; DUSEK et al., 2002). Working with 1802 patients, a more complex system of three groups was formed. Group 1 was uncertain they would receive IP but all did; group 2 was also uncertain about receiving IP and none received it; group 3 was certain they would receive IP and did. There were no significant differences between groups 1 and 2 and when 1 and 3 were compared relative to the certainty of intercession, results were in the opposite direction to what was expected. In other words, as the authors stated “intercessory prayer had no effect,” and those who were certain of intercessors praying for them had more complications (BENSON et al., 2006, p. 941).

Continuing research in this area, Roberts et al. (2010) overviewed 7,800 patients in 10 studies and with obvious reluctance concluded that IP had no effect on their clinical status, death rate, re-hospitalization rate and readmission to a coronary care unit. In sum, IP demonstrated essentially no positive effects in the cardiovascular domain.

As expected the promise of IP stimulated research on a variety of physical problems such as rheumatoid arthritis, kidney dialysis, aids, neurosurgical pituitary difficulties plus mental conditions such as anxiety, depression, alcoholism etc. Again, findings were disappointing.
A variety of errors in research design and analysis further negated hints of positive effects (SPILKA; LADD, 2013). The bottom line is that efforts to find mechanisms for IP have not supported it.

Though one may accept the standard traditional view of science and after examining the research literature, reject the idea of intercessory prayer, a very sophisticated scientific scholar with expertise in both the mathematics of measurement and the psychology of religion, Richard Gorsuch (2008) sees alternative possibilities which, it could be argued may be in the borderland between the philosophy of science and theology. He dealt with the issue of miracles noting such may be present, but they would not be scientifically identified or testable since they are by definition one-time events. In a more general sense, Dossey and Hufford (2005, p. 115) assert that “prayer not be dishonored or degraded through research”.

Conclusions

Our brief review of research caveats, challenges, and findings concerning religion, prayer and health suggests that there is much room for the refinement of investigations in this arena. The deepest dilemma is that many of the research efforts have not fully taken into account the manner in which prayer is actually practiced. This lack of sensitivity to the context from which prayer arises inevitably results in operationalizations that are mischaracterizations of the lived phenomena. These problems are then amplified because they lead to research designs that are inappropriate to the essence of the behavior.

The fact that 90 percent of the population or more, including non-religious people, suggests prayer is a psychologically useful feature of the human behavioral armamentarium. It is our intent in this paper to emphasize an evaluation and critique of how research in this area has been conceptualized and undertaken. There is obviously a strong need for the creation of solid testable theory along with utilization of more sensitive and sophisticated research designs and data analyses. This is especially true as we begin to bring English-language literature into
contact with other cultures (ESPERANDIO; LADD, 2013; LADD; LADD, in press; LENTINE et al., 2013) because there is no guarantee that what has worked (or failed to work) is not confounded in some fashion by culturally relevant variables. By beginning from a position of contextual sensibilities, we are hopeful that the depth and breadth of the psychology of prayer will continue to be better understood on a global level.

References


CBS. **CBS Poll**: prayer can heal. 11 Feb. 2009.


Received: 10/21/2013
*Recebido*: 21/10/2013

Approved: 11/18/2013
*Aprovado*: 18/11/2013