

A META-ANALYSIS OF THE RELATIONSHIP BETWEEN

SPIRITUALITY AND QUALITY OF LIFE

by

Rick Sawatzky

B.Sc.N., Trinity Western University, 1999

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(School of Nursing)

We accept this thesis as conforming
to the required standard

.....

.....

.....

THE UNIVERSITY OF BRITISH COLUMBIA

September 2002

© Rick Sawatzky, 2002

Abstract

The relationship between spirituality and various dimensions of health and quality of life has been extensively examined during the past decade. Though several literature reviews have been conducted in an attempt to synthesize research findings pertaining to the relationship between spirituality and dimensions of health, a meta-analysis of studies examining spirituality in relation to quality of life has not been identified. The present study was therefore designed to: (a) determine whether there is empirical support for a relationship between spirituality and quality of life, (b) provide an estimate of the strength of this relationship, (c) hypothesize and examine the existence of any potential moderating variables affecting this relationship, and (d) contribute to the conceptualization of spirituality in relation to quality of life.

The research design followed methods for quantitative meta-synthesis as discussed by Lipsey and Wilson (2001), Cooper and Hedges (1994), and Hunter and Schmidt (1990). Potential moderating effects of several methodological differences and sample characteristics were examined using meta-analytic approaches to multivariate regression and analysis of variance. An extensive multidisciplinary literature search resulted in 3,040 published reports that were manually screened according to pre-established selection criteria. Subsequent to the selection process, 62 primary effect sizes from 51 studies were included in the final analysis.

A random effects model analysis of the bivariate correlation between spirituality and quality of life resulted in a moderate effect size ($r = 0.34$, 95% CI: 0.28 – 0.40), thereby providing support for the theoretical framework underlying this study wherein spirituality is depicted as a unique concept that stands in relationship to quality of life. Subsequent regression analyses indicated that differences between operational definitions of spirituality

and quality of life significantly affected the magnitude of this relationship ($R^2 = 0.27$). Other potential moderators, such as age, gender, ethnicity, religious affiliation and sampling method were also examined but the findings pertaining to these variables were inclusive due to limitations associated with the sample of primary studies. The implications of this study are mostly theoretical in nature and raise questions about the commonly assumed multidimensional conceptualization of quality of life.

Table of Contents

Abstract ii

Table of Contents iv

List of Figures viii

Acknowledgements ix

Introduction and Purpose 1

Conceptual and Theoretical Foundations.....3

 Conceptualization of Spirituality 3

 Definitions and attributes of spirituality. 4

 Organizational frameworks of spirituality..... 6

 Religion and spirituality..... 9

 Spirituality as a subjective experience..... 12

Measurement of Spirituality 14

Emerging Theoretical Relationships..... 17

 Defining quality of life..... 19

 Measuring quality of life..... 21

Theoretical Framework..... 23

Research Question and Objectives..... 24

Research Design.....25

 Criteria for Selection of Studies..... 26

 Substantive inclusion criteria..... 26

 Methodological inclusion criteria. 27

Exclusion criteria 28

Overview of the selection process 29

Search Strategy 33

 Computerized Databases Searches 34

 Overview and results of the search process 37

Coding and Classifying Study Variables 40

 Primary and secondary effect sizes 43

 Coding characteristics of operational definitions 45

 Issues of reliability throughout the coding process 47

Statistical Methods of Analysis 48

 Methods for Calculating Independent Effect Sizes 48

 Accounting for Variance in the Distribution of Effect Sizes 51

 Outlier analysis 51

 Analysis of homogeneity of variances pertaining to the distribution of effect sizes 51

Preliminary Analyses 54

 Analysis of Selection Process 54

 Analysis of distribution of effect sizes 56

 Outliers 57

 Non-significant unreported correlations 60

 Confidence intervals of original effect sizes 62

Results 64

 Sample Description 64

 Description of Operational Definitions 65

Mean Effect Size..... 67

Analysis of Moderator Variables 70

 Potential indicators of study quality. 70

 Moderating effects of sample characteristics..... 72

 Operational definitions as moderating variables..... 75

Multivariate regression analyses..... 80

Discussion.....83

 Magnitude of the Relationship between Spirituality and Quality of Life..... 83

 Factors Affecting the Relationship between Spirituality and Quality of Life 87

 Conceptualization and measurement of spirituality..... 88

 Operational characteristics of quality of life measurement. 94

 Limitations 95

 Bias in the selection process. 96

 The apples and oranges phenomenon. 97

 Lack of causal evidence..... 98

 The mechanism by which spirituality may affect quality of life. 99

 Generalizability of the results. 100

Conclusion100

References.....102

Appendix A: Selection Criteria Form.....115

Appendix B: Overview of Database Searches117

Appendix C: Code Book.....119

Appendix D: Reliability Coefficients for Spirituality and Quality of Life Instruments.....131

Appendix E: Description and Classification of Operational Definitions of Spirituality139

Appendix F: Mean Effect Size Analyses of the Relationship between Spirituality and Quality
of Life.....147

Appendix G: Comprehensive Mixed Effects Model148

Appendix H: Trimmed Mixed Effects Model.....149

Appendix I: Trimmed Mixed Effects Model Not Adjusted for Instrument Reliability150

List of Figures

Figure 1	Relationship between Spirituality and Quality of Life	24
Figure 2	Histogram of the Distribution of Correlations.....	57
Figure 3	Histogram of the Distribution of Correlations.....	57
Figure 4	Comparison of Outliers to Regression Line	59
Figure 5	Distribution of Correlations after Adjustment for Outliers and Non-Significant Correlations that were Not Reported.....	61
Figure 6	Comparison of Adjusted Values to Regression Line	61
Figure 7	95% Confidence Interval for Unadjusted Correlations	63
Figure 8	Comparison of Ethnic Groups	75
Figure 9	95% CI Associated with Each Categorical Group of Spirituality Instruments.....	77
Figure 10	Box Plots of Collapsed QOL Categories.....	80
Figure 11	Venn-Diagram of Potential Conceptual Overlap	85
Figure 12	Combination of Operational Categories Represented by Each Effect Size	92
Figure 13	95% CI Associated with Each Categorical Group of QOL Instruments as Revealed by the ANOVA Analysis	95

Acknowledgements

This study was accomplished with the practical and encouraging support of friends, family members, professors and assistants. Their support has been both inspiring and humbling; a vivid reminder that few things in life are accomplished by the work of one person. I wish to especially express my gratitude to my wife, Helma, who was always there to support me even during the times that I was too tired and stressed to support her. I also wish to thank my advisors, Dr. Lyren Chiu and Dr. Pamela Ratner, and my third committee member, Dr. Joan Bottorff, who have gone far beyond my expectations in supporting me in both practical and affirmative ways. The entire process has been very enjoyable and encouraging; a testimony to a communal effort despite our diverse backgrounds and traditions. I am tremendously grateful for the assistance of Glenna Stewart, who spent endless hours reviewing abstracts and completing data collection forms, and who was always available to discuss and share ideas. This study would not have been completed in the same manner without her compassionate support. My gratitude also extends to Rubeena Mali, whose assistance in retrieving, photocopying and faxing hundreds of publications made the completion of this study so much easier. Finally, the completion of this meta-analysis would not have been possible without the extensive efforts of the librarians at the interlibrary loan department of Woodward library. I have no doubt that the completion of the project comes with a sigh of relief on their part.

At the heart of this study lies my own spirituality which is based on my personal relationship with Yahweh, the one who has given my life meaning, purpose and direction, the author of my faith.

Introduction and Purpose

I am interested in exploring the role of spirituality within the context of health care. Scholars, health practitioners, policy makers and recipients of health care increasingly recognize the influence of spirituality on health, as defined by the individual, and the importance of considering the spiritual dimension in the provision of health care. For example, qualitative studies on spirituality have focused on identifying defining attributes and conceptualizations of spirituality in relation to individually and culturally defined experiences of health and illness (Burkhardt, 1993; Chiu, 2000; Martinez, 1999; McGrath, 1997; Potts, 1996; Rehm, 1999; Tongprateep, 2000). Spirituality also has been examined from a quantitative perspective where various operationalizations of spirituality, such as specific spiritual or religious practices, measures of related concepts such as spiritual well-being, religiosity, and self-transcendence, have been examined in relation to diverse outcome measures including specific medical outcomes, functional outcomes, mental health outcomes and various measures of quality of life (Abeles et al., 1999; Larson, Sawyers, & McCullough, 1998; Thoresen, 1999). Although researchers conducting quantitative studies of spirituality initially and primarily used objectively identifiable indicators of spirituality, such as religious involvement and specific spiritual practices (George, Larson, Koenig, & McCullough, 2000), researchers have increasingly turned to using instruments that measure spirituality from a subjective perspective, as experienced by individuals who are situated in a particular cultural and historical context. Such instruments include measures of spiritual or existential well-being, measures of life purpose, and measures of self-transcendence. This orientation toward the subjective experience of spirituality is supported by an evolving collection of qualitative studies on spirituality.

Early researchers conducting studies on spirituality primarily tended to include concrete outcomes that reflected the modernistic values of the medical community, such as measures of recovery, morbidity, and mental health outcomes. More recently, researchers of spirituality have attempted to incorporate measures that are more meaningful to the holistic and subjective experience of health as operationalized by measures of perceived quality of life, wellbeing and life satisfaction.

Although several literature reviews have examined relationships between spirituality and health (Abeles et al., 1999; Aldridge, 1993; Elkins, Hedstrom, Hugher, Leaf, & Saunders, 1988; George et al., 2000; Larson et al., 1998; Levin, 1994; Mytko & Knight, 1999; Thoresen, 1999), a search of the literature did not reveal a meta-analysis of studies that explicitly examine subjective measures of spirituality in relation to perceived quality of life. Most reviews address important principles of theory development such as identifying defining characteristics of spirituality, addressing the measurement of spirituality, and identifying potential theoretical frameworks. However, proposed theoretical relationships have been derived primarily from qualitative syntheses of the existing literature and substantiated by quantitative findings of individual studies. Few attempts have been made to examine the important health-related correlates of spirituality by statistically aggregating diverse quantitative findings with the goal of explicating the relationship between spirituality and health while taking into consideration possible influences of extraneous variables. Substantiation of the relationships between spirituality and health-related concepts will help to incorporate the spiritual dimension of life in theoretical frameworks underlying the practice of health-care professionals and thereby provide direction to the provision of health-care services.

Conceptual and Theoretical Foundations¹

Familiarization with the theoretical and conceptual foundations of spirituality is necessary to identify the body of literature from which studies for a meta-analytic review can be drawn. The conceptualization of spirituality involves identifying attributes that are commonly associated with spirituality, such as transcendence, the existential and sacred, as well as distinguishing spirituality from related concepts such as religion and religiousness. For example, spirituality has been conceptualized as primarily religious in nature, as exclusively non-religious in nature or as a combination of religious and existential phenomena. Similarly, operational definitions of spirituality can be broadly dichotomized in terms of instruments that objectively measure the frequency of performing individual spiritual or religious activities and instruments that subjectively attempt to measure spiritual experiences and beliefs as well as the subjective meaning associated with spirituality. In addition, theoretical frameworks provide the basis for hypothetical relationships between spirituality and a wide variety of medical, psychological, social, and behavioural outcomes as well as more subjectively oriented measures of quality of life, wellbeing or life satisfaction. Consensus on conceptual and operational definitions and theoretical frameworks is necessary to foster continued empirical development and practical application of the concept of spirituality (George et al., 2000; Larson et al., 1998; Sussman, Nezami, & Mishra, 1997).

Conceptualization of Spirituality

Discussions about the conceptualization of spirituality reveal a variety of definitions and attributes of spirituality and include ongoing debates about the need to distinguish

¹ The section on conceptual and theoretical foundations is taken almost entirely from the following paper, which was written in preparation for this meta-analysis: "Spirituality and Outcome Measures: Examining the Feasibility of a Meta-Analysis of Existing Literature" (Sawatzky, 2002).

spirituality from religion or religiosity. In reviewing current spirituality literature an emerging consensus on defining spirituality as a subjective experience that may or may not be associated with institutionalized religion is becoming increasingly evident (Vaughan, Wittine, & Walsh, 1996).

Definitions and attributes of spirituality.

The challenge of defining spirituality becomes evident in reviewing the myriad of definitions and attributes ascribed to this concept. Clearly, consensus on how to define such a subjective and elusive phenomenon continues to be an ongoing struggle (Larson et al., 1998). Definitions of spirituality range from the perspective of organized religion with associated predefined beliefs and practices to seeing spirituality as a rather elusive and subjective human experience (Thoresen, 1999). Definitions include those provided by Vaughan et al. (1996), who defined spirituality as “a subjective experience that exists both within and outside of traditional religious systems” (p. 497) and by Sussman et al. (1997), who defined spirituality as “subjectively experiencing a life force” (p. 112). A slightly different approach to spirituality was taken by Emblen (1992) who conducted an analysis of existing definitions of spirituality to formulate the following definition: “Personal life principle [which] animates transcendent quality [of] relationship [with] God or god being” (p. 45). A collection of other definitions of spirituality cited by Aldridge (1993) exemplifies the various conceptualizations in relation to experiential, relational and existential phenomena and healing practices associated with spirituality (Thoresen, 1999).

Despite the diversity in the conceptual definitions of spirituality, several common defining characteristics can be extracted. The conceptual definitions generally involve a relationship to something that lies beyond physiological, psychological or social human

perception or experience. This relational object of spirituality may be described as “divinity”, “a higher power”, a “divine being”, “ultimate reality”, “God” or “god-being” (Burkhardt, 1989; Dyson, Cobb, & Forman, 1997; Emblen, 1992; Larson et al., 1998; Reed, 1993; Thoresen, 1999). The experience associated with this relational aspect of spirituality can be described as transcendent or transpersonal in nature (Emblen, 1992; Reed, 1993; Vaughan et al., 1996).

Most definitions of spirituality support the notion that spirituality is associated with an existential search for meaning and purpose regarding complex questions about life (Larson et al., 1998; Thoresen, 1999). This search can be directed by religious institutions or by personal life experiences. However, Hill et al. (2000) caution that although the existential search appears to provide a fairly clear defining characteristic of spirituality, existentialism in and of itself is not equal to spirituality. There needs to be a criterion that distinguishes existential searching and ideology that is inherently spiritual in nature from that which is existential but not necessarily spiritual.

Based on an extensive review of literature on spirituality, several researchers suggest that the criterion distinguishing spirituality from other potentially meaningful existential pursuits, ideologies or life-giving practices is its orientation toward the sacred (George et al., 2000; Hill et al., 2000; Larson et al., 1998). Hill et al. (2000) state that "when the term 'spirituality' is invoked to describe ideologies or lifestyles that do not invoke notions of the sacred in one way or another, they are not spiritualities at all, just strongly held ideologies or highly elaborated lifestyles" (p. 64). This sacred core of spirituality is conceptualized by a large panel of researchers associated with the National Institute of Healthcare Research as “a divine being or Ultimate Reality or Ultimate Truth as perceived by the individual” (Larson et

al., 1998, p. 21). Similarly, Pargament (1999) explains that “the sacred encompasses concepts of God, the divine, and the transcendent, but it is not limited to notions of higher powers. It also includes objects, attributes, or qualities that become sanctified by virtue of their association with or representation of the holy” (p. 12). Sacred refers to those aspects of life that are either transcendent in nature or associated with a transcendent dimension. In addition, the sacred encompasses the transcendent and existential characteristics of spirituality discussed earlier. In consideration of the relational characteristic of spirituality, spirituality can then be defined as involving a relationship with the sacred.

Organizational frameworks of spirituality.

Unfortunately, the nature of how the sacred interacts with human existence is not clearly defined. This is problematic considering that the only way to measure such an elusive concept is by knowing how it is experienced or expressed. Two primary threads can be derived from spirituality definitions: (a) spirituality can be experienced through thoughts and feelings derived from a relationship with a transcendent reality as discussed earlier, or (b) spirituality can be expressed through behaviour and activities that are directed toward fostering the same relationship. The first thread, which may be termed the experiential dimension of spirituality, relates to Vaughan et al.’s (1996) definition of spirituality as “a subjective experience” (p. 497). This definition is based on the philosophical tenets of transpersonal psychology which has been described as follows:

“[Transpersonal psychology] allows for a vision of the human potential that explicitly includes spiritual experience. In the light of the perennial wisdom of spiritual teachings, it affirms the possibility of living in harmony with others and the environment, reducing fear and greed, and developing compassion and a sense of

meaning and purpose in life, regardless of the particular religious beliefs that may be espoused." (Vaughan et al., 1996, p. 485)

More specifically related to the conceptual attributes of spirituality mentioned earlier, spiritual experience can also be defined as "experiences that may or may not be part of daily life, and [that reflect] the perception of interacting and being involved with the sacred" (Larson et al., 1998, p. 26). The inherent notion that perception provides a link between the sacred and the experience of spirituality is consistent with the subjective nature of spirituality as defined by the transpersonal psychologists mentioned earlier. The main distinction is that for the NIHR research panel, spiritual experience is defined as merely one of many domains of spirituality, whereas Vaughan et al. (1996) see the subjective experience as encompassing spirituality as a whole (i.e., it appears from Vaughan's definition that spirituality does not exist unless it is subjectively experienced). Despite this apparent incongruence, it becomes clear from these definitions that there exists an experiential dimension of spirituality that is inherently subjective in nature, and that this experiential dimension of spirituality may or may not be expressed through predefined behaviors and practices.

The expressive dimension of spirituality involves a hodgepodge of behaviour, practices and interventions that are assumed to be spiritual in nature. In consideration of the proposed defining characteristics of spirituality, spiritual expressions would be distinguished by their association with the sacred. However, as is revealed in the definition of the sacred espoused by Pargament (1999), sacredness can be ascribed to almost any behaviour or expression. Spiritual expression is therefore not so much defined by predefined spiritual behaviour, practices or interventions, but rather by their association with the combined transcendent or existential qualities of spirituality. Although this distinction between spiritual

expression and non-spiritual expression may provide useful direction in theory development, distinguishing spiritual behaviour and practices from non-spiritual ones remains a difficult task that is subject to individual interpretation at the operational level.

Besides categorizing spirituality in terms of its experiential and expressive dimensions, several other organizational frameworks of spirituality have been suggested. For example, LaPierre (1994) structured spirituality according to the following spiritual categories: (a) “a search for meaning in life”, (b) “an encounter with transcendence”, (c) “a sense of community”, (d) “a search for ultimate truth, or highest value”, (e) “a respect and appreciation for the mystery of creation”, and (f) “a personal transformation” (p. 153). Though a comprehensive discussion of these categories goes beyond the scope of this thesis, the categories as a whole reveal a remarkable resemblance with the defining characteristics of spirituality discussed earlier. For example, the existential characteristic of spirituality is reflected in categories one and four and the relational characteristic of spirituality may be reflected in categories three and two. Category two also reflects the transcendent nature of spirituality and category four reflects the sacred core of spirituality as defined earlier. Finally, “a personal transformation” relates to the subjective nature of spirituality. By including “a sense of community” as an attribute of spirituality, LaPierre also suggests that spirituality involves a communal component that moves beyond the subjective toward a shared or common experience.

Other organizational frameworks are provided by extensive literature review initiatives supported by the National Institute for Health Research (Larson et al., 1998) and the Fetzer Institute and the National Institute on Aging (Abeles et al., 1999). In a report on “Scientific Research on Spirituality and Health,” Larson et al. (1998) identify the following

domains of spirituality to guide ongoing research: (a) “Religious/Spirituality Preference or Affiliation”, (b) “Religious/Spiritual History”, (c) “Religious/Spiritual Participation”, (d) “Religious/Spiritual Private Practices”, (e) “Religious/Spiritual Support”, (f) “Religious/Spiritual Coping”, (g) “Religious/Spiritual Beliefs and Values”, (h) “Religious/Spiritual Commitment”, (i) “Religious/Spiritual Motivation for Regulating and Reconciling Relationships”, and (j) “Religious/Spiritual Experiences” (pp. 24-26). Similar domains are defined by the research panel of the Fetzer Institute and the National Institute on Aging (Abeles et al., 1999). The panel members identified 12 domains of spirituality that relate to health-related outcomes. The domains are labeled as follows: “daily spiritual experience, meaning, values, beliefs, forgiveness, private religious practices, religious/spiritual coping, religious support, religious/spiritual history, commitment, organizational religiousness and religious preference” (p. 4). On the whole, these organizational frameworks provide a fairly concrete structure for examining spiritual phenomena in relation to a variety of health-related outcome measures. In addition, the defining characteristics of spirituality discussed earlier are consistent with these organizational frameworks. However, one pertinent area of ambiguity that becomes evident upon reviewing these organizational frameworks of spirituality is the apparent lack of distinction between the concepts of religion and spirituality. In preparing for a synthesis of existing empirical literature on spirituality, the conceptualization of spirituality underlying this synthesis must address this ambiguity.

Religion and spirituality.

In reviewing the debate on spirituality and religion, several points of discussion can be identified. First, an evolving body of empirical literature on spirituality reveals a general

drive to disassociate spirituality from religion where spirituality is increasingly defined as subjective experiences and religiosity is increasingly meant to describe institutionalized religious activity and participation. In North America, this drive can be traced to secular and individualistic movements during the second part of the twentieth century (Hill et al., 2000). However, as George et al. (2000), Hill et al. (2000), and Pargament (1999) point out, the conceptual distinction between spirituality and religion was virtually non-existent in research prior to this period. Therefore, the distinction between spirituality and religion should be seen as a fairly recent conceptual transformation which is occurring during a historical transformation from a religiously dominated spiritual world toward a humanistic and relativistic understanding of spirituality.

The debate surrounding the polarization of spirituality and religion reflects a change in how these two concepts are defined (Emblen, 1992; Hill et al., 2000; Pargament, 1999; Slanter, Hall, & Edwards, 2001; Zinnbauer, Pargament, & Scott, 1999). Zinnbauer et al. (1997) argue that the study of religion originally encompassed everything that is now deemed spiritual, thereby suggesting that the differentiation between spirituality and religion occurred in response to secular ideology and “a popular disillusionment with religious institutions” (p. 550). This shift is exemplified in definitions in statements from transpersonal psychologists such as Vaughan et al. (1996) who suggest that "... spirituality, unlike religion, does not require obedience to a particular set of beliefs or prescribed dogma" (p. 500). Similarly, Sussman et al. (1997) suggest that “spirituality involves transcendental processes that supersede ordinary existence, whereas religion involves subscription to a set of beliefs which are organized and institutionalized” (p. 122). These distinctions are consistent with Emblen’s (1992) analysis of spirituality and religion as discussed in the nursing literature. Emblen’s

analysis reveals that religion is primarily defined as a “system of organized beliefs and worship [which the] person practices,” whereas spirituality more commonly refers to “a personal life principle [which] animates transcendent quality [of] relationship [with] God or god being” (p. 45). Consequently, the distinction between spirituality and religion appears to primarily refer to the degree of association with institutionalized or organized religion. Yet, as Zinnbauer et al. (1997) point out, institutionalized religion is only one of the many domains of the overarching theoretical construct of religion. The theoretical construct of religion traditionally applied to a large body of empirical research on religion encompassed subjective and experiential domains that were not necessarily associated with institutionalized religion. Thus, there appears to be a shift in terminology where, according to these authors, the original construct of religion encompasses what is now called spirituality.

There appears to be a general consensus that spirituality and religion must, to a large extent, be overlapping or closely related concepts that are not easily examined in isolation of one another (George et al., 2000; Larson et al., 1998; Zinnbauer et al., 1997). For example, even though theorists with a background in transpersonal psychology argue that spiritual experiences can occur in isolation of institutionalized religion, they also recognize that spiritual experiences can be mediated through religion (Elkins et al., 1988; Vaughan et al., 1996). Certainly it must be recognized that spiritual experiences can occur within the context of organized religion since spirituality lies at the heart of religious purposes. If this is the case, then an overarching conceptualization of spirituality must also encompass religious phenomena.

Several attempts have been made to reach a consensus on the spirituality versus religion debate at the conceptual level. Based on their review of historical and empirical

literature George et al. (2000) conclude that there seems to be a general consensus that religion and spirituality are distinguished by the “collective or institutional context” (p. 103) that is seen as a defining characteristic of religion but not necessarily of spirituality. From this perspective, spirituality is subjectively defined by personal experiences, beliefs and practices that may be regulated by an overarching organized social context (Elkins et al., 1988; Larson et al., 1998; Vaughan et al., 1996). This notion is consistent with the distinction between spirituality and religion espoused by the NIHR research panel which suggests that spirituality can be defined as “the feelings, thoughts, experiences, and behaviors that arise from a search for the sacred” (Larson et al., 1998, p. 21). Religion can be seen as encompassing the definition of spirituality with the added criterion of being part of an organized social context (Hill & Hood, 1999; Larson et al., 1998) or “system of organized beliefs” (Emblen, 1992, p. 45).

Spirituality as a subjective experience.

On the whole, my discussion thus far reveals an emerging conceptualization of spirituality as a subjective experience that reflects the previously identified common threads of transcendent relatedness and existential searching, which are both associated with the sacred and which may or may not be mediated by religion. In consideration of extensive conceptual debates in the literature, it seems that the definition of spirituality provided by the NIHR research panel provides useful direction for ongoing research and theory development by distinguishing spirituality from a myriad of other phenomena that, albeit meaningful and existential in nature, are not necessarily spiritual. In addition, the definition provides a useful distinction between spirituality and religion that recognizes recent trends in spirituality research while incorporating previous theoretical development surrounding the concept of

religion. Furthermore, the definition of the NIHR research panel is consistent with a large variety of definitions espoused by researchers from diverse disciplinary and theoretical backgrounds. The complete definition of spirituality suggested by the NIHR research panel is as follows:

The feelings, thoughts, experiences, and behaviors that arise from a search for the sacred. The term “search” refers to attempts to identify, articulate, maintain, or transform. The term “sacred” refers to a divine being or Ultimate Reality or Ultimate Truth as perceived by the individual. (Larson et al., 1998, p. 21)

Two criteria are added to this definition to distinguish spirituality from religion: (a) Religion may or may not encompass “a search for non-sacred goals (such as identify, belongingness, meaning, health, or wellness) in a context that has as its primary goal the facilitation of [the previous definition]” and (b) religion refers to “the means and methods (i.e., rituals or prescribed behaviors) of the search that receive validation and support from within an identifiable group of people” (Larson et al., 1998, p. 21). According to this definition, religion does not occur in isolation from spirituality because spirituality encompasses religion. A useful distinguishing feature of this definition is that those practices or behaviours that resemble religious expression but that do not necessarily involve the sacred are not considered to be spirituality. This distinction recognizes the mutual concerns of people with backgrounds in transpersonal psychology, humanism, existentialism and formal religion that religious beliefs, expression and behaviors can be abused for forms of personal gain in ways that ignore the sacred.

Measurement of Spirituality

Challenges in measuring spirituality follow the same concerns and trends as difficulties in conceptualizing this concept. In the beginning stages, research was almost exclusively focused on the measurement of religion (George et al., 2000; Larson et al., 1998; Sussman et al., 1997). However, during the past few decades numerous measures of spirituality and religion have been added to the repertoire of available instruments. Reviews of spirituality instruments reveal an overwhelming diversity of operational definitions (Hill & Hood, 1999; Larson et al., 1998; MacDonald, Friedman, & Kuentzel, 1999; MacDonald, Kuentzel, & Friedman, 1999; MacDonald, LeClair, Holland, Alter, & Friedman, 1995). Larson et al. (1998) describe a limited collection of instruments in relation to each of their spiritual domains described earlier. However, the empirical utility of this type of operational taxonomy is constrained by the difficulties in providing empirical validation of the different categories. In addition, different instruments often are derived from different theoretical backgrounds and typically measure diverse aspects of spirituality.

MacDonald (2000) sought to address the diversity of empirical and theoretical developments pertaining to spirituality by conducting a factor analysis of 11 instruments measuring spirituality with the purpose of “[developing and measuring] a descriptive organizational model of spirituality that could be used as a framework for structuring existing scientific knowledge and as a basis for guiding future research” (p. 156). By using this approach MacDonald was able to identify seven distinct operational dimensions of spirituality. These dimensions were used in the construction and validation of a new scale, the “Expressions of Spirituality Inventory” (MacDonald, 2000). Five of the original dimensions were retained in a subsequent the factor analysis of this new scale. These

dimensions were labeled as follows: (a) “cognitive orientation towards spirituality,” (b) “experiential/phenomenological dimension of spirituality,” (c) “existential well-being,” (d) “paranormal beliefs,” and (e) “religiousness” (p. 169). Unfortunately, though the operational framework developed by MacDonald (2000) does provide an operational structure for a pre-defined selection of instruments used to measure various aspects of spirituality, a large number of spirituality instruments were systematically excluded from his analysis thereby constraining the intent to develop an operational framework that is representative of the entire spiritual domain.² The problem pertaining to the diversity of instruments that are considered to be representative of the spiritual domain is therefore not truly addressed, and the difficulty of distinguishing instruments that measure spirituality from those that measure related but distinct concepts remains.

In preparation for a meta-analysis on spirituality it is necessary to identify a collection of instruments that can be meaningfully combined without systematically excluding instruments that are representative of the spiritual domain. The selection of such instruments should take direction from the conceptualization of spirituality underlying this meta-analysis as discussed earlier. Operational definitions that reflect this conceptualization of spirituality include instruments that use subjective ratings of spiritual experiences resulting from a search for the sacred, where spiritual experiences are those experiences associated with (a) an existential search for meaning and (b) a relationship with a transcendent reality. Excluded are instruments that exclusively measure the frequency of religious or spiritual practices or

² Though MacDonald conducted several extensive reviews of 70 instruments measuring spirituality (as cited earlier), the selection criteria pertaining to this factor analysis precluded the selection of most of these instruments. The selection criteria were primarily based on whether the instrument measured “an aspect of spirituality that not only was assumed to be a part of spirituality but that also was unique relative to other available measures” (p. 158). Additional selection criteria pertained to the author’s assessment of the validity and reliability of the instruments as well as the degree to which the structure of the instrument was considered to be “parsimonious” (p. 158). The factor model resulting from this study can therefore not be seen as a true reflection of the wide diversity of instruments measuring spirituality.

behaviour. These selection criteria (which are discussed in further detail in the methods section of this report) are oriented toward a view of spirituality as a subjective experience and therefore do not include external measures of religion (such as religious practices, religious affiliation and religious involvement).

On the whole, my literature review suggests that current operational definitions, based on the conceptualization of spiritual experience as defined above, may be usefully and meaningfully combined into a common metric that provides an overall measure of spirituality. Spirituality involves spiritual experiences that result from a relationship with the sacred and are associated with an existential search regardless of whether or not these experiences are mediated by religion. Thus, spirituality is considered to encompass spiritual experiences associated with religious practices and beliefs although it can also exist in isolation from religion. This has several implications for performing a meta-analysis based on combining findings derived from diverse operational definitions of spirituality. First, measures of religious experience should not be excluded from such an analysis. If we were to exclude these measures, then we would exclude the many studies that include the measurement of spiritual experience occurring within the context of religion. Therefore, literature searches should include keywords associated with religion or religiosity in order to avoid the exclusion of an important body of empirical literature. Second, it means that only those instruments that allow for the separate measurement of spirituality, distinct from religious practice, can be used. Although the goal here is not to view spirituality in isolation of religious influence (for many people spirituality is at least partly defined by their religious affiliation), the goal is to separate a person's experience of his/her spirituality from external measures such as religious commitment, church attendance and measures of religious or

spiritual practice and behaviour. The distinction therefore does not lay so much in separating spirituality from religion or religiousness, but rather in separating a person's experience of his or her spirituality from what have been deemed "objective" indicators of religiosity.

Emerging Theoretical Relationships

So far my discussion has involved formulating a conceptualization of spirituality that is meaningful to continued theoretical development and consistent with a wide variety of operational definitions. Based on this conceptualization of spirituality, I identified operational criteria that allow for the selection of relevant measurements of spirituality. The next step is to examine the literature for potential correlates that would be of interest to theoretical development and practical application. A preliminary overview of the existing literature on spirituality reveals a rapidly evolving accumulation of empirical findings related to spirituality and a large variety of outcome measures, yet few studies clearly refer to a theoretical framework that involves the spiritual dimension. Nevertheless, trends in spirituality research reveal a number of hypothetical relationships between spirituality and other phenomena.

An overarching trend in spirituality research using quantitative methods relates to the examination of relationships between spirituality and outcomes associated with health. Even though the theoretical bases of hypothesized relationships are rarely mentioned, the underlying hypothesis is that spirituality, however measured or defined, has a general positive association with various measures of health and well-being (Thoresen, 1999). Clearly, this research is driven by a desire to substantiate the significance of spirituality in theory and practice related to health or well-being. Several extensive literature reviews have been conducted to provide direction for this kind of research (Abeles et al., 1999; Aldridge,

1993; Larson et al., 1998; Levin, 1994; Thoresen, 1999). The NIHR Panel for Scientific Research on Spirituality and Health (Larson et al., 1998) for example, reviewed a large collection of studies that examined the relationships between measures of religion/spirituality and physical health, mental health, alcohol and drug problems and neuroscientific observations. Research panels consisting of scholars from a wide variety of disciplinary backgrounds were established to examine each of these relationships and to provide direction for further research. An additional panel examined the many definitions of spirituality. Another literature review initiative was undertaken by the Fetzer Institute and the National Institute on Aging Working Group (Abeles et al., 1999). This team of reviewers examined the dimensions of spirituality mentioned earlier in relation to a large variety of behavioural, social, psychological and physiological health outcomes.

Together, these research reviews provide very useful and concrete direction to continued theoretical development on spirituality in the context of health. However, the relationship between spirituality and well-being or perceived quality of life is not explicitly revealed. This is surprising in consideration of the relationships between spirituality and well being or quality of life that are inherent to various operational definitions of spirituality such as the Spiritual Well-Being Scale (Paloutzian & Ellison, 1982), the Functional Assessment of Chronic Illness Therapy – Spiritual Well-Being (Brady, Peterman, Fitchett, Mo, & Cella, 1999) and the Mental, Physical, and Spiritual Well-Being Scale (Vella-Brodrick & Allen, 1995). Furthermore, an extensive literature review of the relationships between spirituality and well being or quality of life has not been reported despite the recent development of several theoretical frameworks that include these relationships (Mytko & Knight, 1999). Finally, it is surprising that current reviews are primarily oriented toward examining

spirituality in relation to fairly objective constructs associated with health outcomes whereas the subjective experience of spirituality might more logically relate to subjective outcomes such as perceived quality of life, well being or life satisfaction.

Theoretical development of notions of spirituality remains very much in an infancy stage; the concept of spirituality still needs to be clearly defined and related concepts need to be identified (Thoresen, 1999). Despite this limited theoretical foundation, the drive underlying most spirituality research seems to be primarily motivated by a desire to defend the significance of spirituality by examining potential relationships between spirituality and a wide variety of health-related outcomes. However, a current systematic review that statistically combines findings derived from primary studies examining relationships between spirituality and other subjective phenomena such as perceived quality of life, well being and life satisfaction has not been reported. Substantial empirical support of this relationship would contribute to current theoretical developments involving the relationship between spirituality and quality of life.

Defining quality of life.

An extensive discussion on the conceptualization and operationalization of quality of life goes beyond the scope of this thesis, however several guiding principles need to be clarified to ensure the selection of a homogeneous collection of operational definitions of quality of life that can be meaningfully combined and examined in relation to spirituality. Based on a literature review examining the relationship between spirituality and quality of life, Mytko and Knight (1999) generated the following quality of life definition: “The term ‘quality of life’ usually refers to a multidimensional construct that includes the patient’s perspective of their overall quality of life and their assessment of specific components of

quality of life (i.e., physical, psychological and social well-being)” (p. 445). This definition combines two distinct characteristics of quality of life that are pertinent to our understanding of the potential relationship between spirituality and quality of life: (a) quality of life is a subjective phenomenon that encompasses other concepts (such as physical, psychological, and social dimensions of life) and (b) it is seen as a construct that encompasses multiple dimensions. The underlying assumption is that these two characteristics of quality of life constitute a unified phenomenon. Mytko and Knight (1999) continue by describing two models that specifically relate to the multidimensional aspect of this definition. “Models of multidimensional quality of life have included a religiosity and spirituality domain. The religiosity and spirituality domain, however, has been conceptualized either (1) as an independent component of quality of life or (2) as an overarching personal life perspective or world view which influences all quality of life domains (e.g., physical, functional, emotional)” (Mytko & Knight, 1999, p. 445). However, both models are based on the assumption that the multiple dimensions of quality of life relate to one another in a predictable and consistent manner and should therefore be seen as part of one overarching quality of life construct. This assumption, however, is questionable considering the diverse nature of each of these dimensions.

Since we cannot assume the existence of consistent relationships among the different dimensions that are commonly associated with quality of life, it may be more appropriate to view each of these dimensions as “causal sources” of quality of life (Beckie & Hayduk, 1997). In this sense, quality of life can be conceptualized as “a global personal assessment of a single dimension which may be causally responsive to a variety of other distinct dimensions” (p. 22). Quality of life is not seen as a construct that consists of multiple

dimensions, but rather as a “unidimensional entity” (p. 25) that, while having several predictor variables, remains a unique concept in its own right. In addition, quality of life is conceptualized as a subjective phenomenon that includes other subjective concepts such as life satisfaction and well-being. Within this conceptual approach, spirituality can be seen as a concept that is predictive of quality of life, but that remains distinct from other related concepts such as physical, social and psychological well-being. Similarly, physical, social and psychological well-being are also predictive of quality of life but remain distinct from one another.

Measuring quality of life.

In consideration of these issues it will be useful to categorize operational definitions of quality of life in terms of the two distinct conceptualizations discussed above. Commonly identified multidimensional measures of quality of life may be included in a meta-analysis of the relationship between spirituality and quality of life, since exclusion of such measures would unnecessarily constrain the sample of selected studies. Nevertheless, it must remain clear that the conceptual model of quality of life underlying this meta-analysis is one in which quality of life is seen as a subjective concept in its own right as opposed to a construct consisting of multiple dimensions or concepts. The focus is on a person’s perception of his/her quality of life rather than the multiple dimensions of a quality of life construct. In consideration of the conceptualization of quality of life as a subjective experience, measures of life satisfaction and subjective well-being should be included as operational definitions of quality of life. The important criterion becomes whether the instrument measures a person’s subjective experience. Objective measures based on medical outcomes or external measures of a person’s physical, psychological or social functioning do not meet this criterion.

Based on the previous discussion, quality of life as a phenomenon, in and of itself, is defined as a person's subjective experience of his/her quality of life. The question remains as how to measure quality of life as a unique and subjective phenomenon. One operational definition that reflects this conceptualization of quality of life is the single item measurement used by Brady et al. (1999): "I am content with the quality of my life right now" (p. 420). This item is part of the FACT-G (Cella et al., 1993) quality of life measurement tool. Although the FACT-G operationalized quality of life as a multi-dimensional construct, Brady et al. argue that the single item of contentment with quality of life can be seen as an overall measure of the person's perception of his/her quality of life. Other operational definitions of quality of life include measures of satisfaction with life such as LIFESAT (Poloma & Pendleton, 1990), the Life Satisfaction Questionnaire (Fugl-Meyer, Branholm, & Fugl-Meyer, 1991) and the Satisfaction with Life Scale (Diener, Emmons, Larson, & Griffin, 1985), as well as measures of subjective well-being (Diener, Suh, Lucas, & Smith, 1999).

Naturally, reliability issues associated with single-item operational definitions of quality of life as well as the small number of instruments that are entirely subjective in nature necessitate the inclusion of other operational definitions of quality of life. The difficulty therefore is to assess the degree to which a particular instrument measures a subjective experience as opposed to objective external indicators of multiple quality of life dimensions since various quality of life instruments include both subjective and objective components. Again, the decision here should be based on the intent to include as many as possible compatible instruments. Consequently, multi-dimensional operationalizations should be included based on the assumption that the combination of multiple dimensions provides a valid indicator of subjective quality of life. As Beckie and Hayduk (1997) suggest, "It is

entirely reasonable to postulate that a unidimensional QOL variable can result from a global assessment spanning diverse and complex domains” (p. 25).

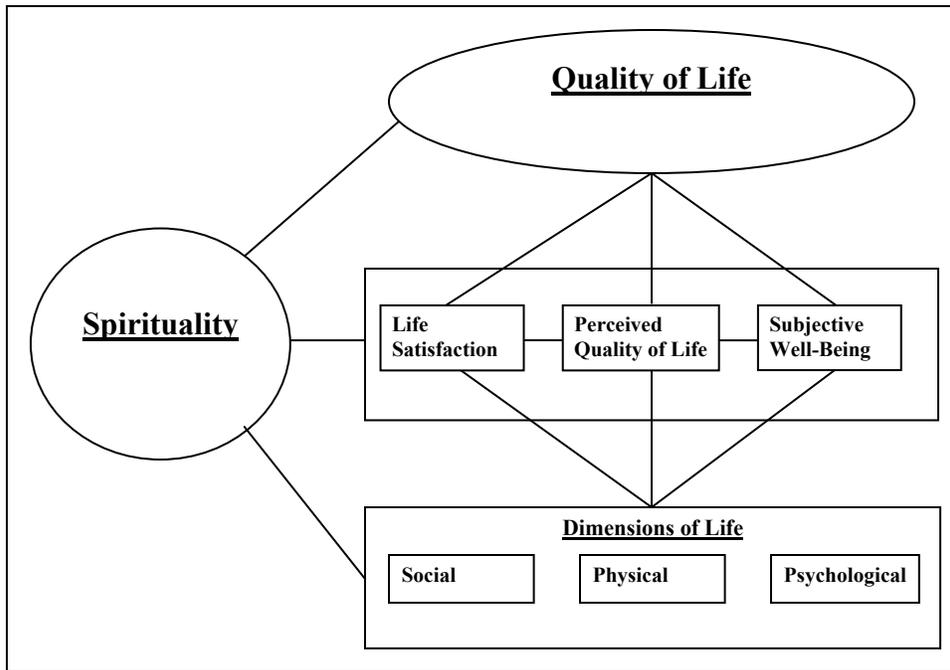
Based on these conceptual considerations, the criteria for including a quality of life instrument are based on whether an instrument contains subjective measures that either measure quality of life directly in terms of perceived quality of life, subjective well-being or life-satisfaction, or indirectly by combining measures of multiple causal predictors, such as multiple dimensions of life. Any instrument that meets either of these criteria should be included even if the instrument might include some items that explicitly measure objective indicators. Once all the data are collected, subsequent regression analysis techniques can be used to identify whether instruments that include external or multidimensional measures have a significantly different outcome in comparison with measures that are purely subjective or unidimensional in nature.

Theoretical Framework

The above conceptualizations and operational definitions of spirituality and quality of life are integrated into the theoretical framework depicted in Figure 1. Within this theoretical framework quality of life is conceptualized as a person’s perception of his/her quality of life however it is defined by the individual. Operational definitions that approximate this conceptualization are the subjective instruments measuring life satisfaction, perceived quality of life and subjective well-being. A less direct approach to measuring quality of life includes operationalizations pertaining to different dimensions of life. Although each dimension in isolation may not consistently predict quality of life, it is hypothesized that these dimensions taken as a whole affect quality of life. The hypothesized relationship between spirituality and quality of life underlying this meta-analysis is thus operationalized by examining the

correlations between diverse operational definitions of spirituality and measures of perceived quality of life or combined measures of multiple dimensions of life.

Figure 1. Relationship between Spirituality and Quality of Life.



Research Question and Objectives

The following research question emerges from my discussion so far: Is there a significant relationship between measures of spirituality and quality of life? This question can be examined by employing meta-analytical methods to aggregate findings derived from a growing body of primary quantitative research initiatives surrounding the relationship between spirituality and quality of life. I therefore identified the following objectives to guide such an analysis:

1. To synthesize results from completed studies to determine whether there is empirical support for a relationship between spirituality and quality of life.

2. To provide an estimate of the strength of this relationship.
3. To hypothesize and examine the existence of any potential related variables affecting this relationship.
4. To contribute to the conceptualization of spirituality in relation to quality of life.

Research Design

One of the primary advantages of a meta-analysis is the potential to address biases that may influence a conventional literature review (Lipsey & Wilson, 2001; Moody, 1990; Wolf, 1986). Such biases include a variety of researcher biases affecting the selection of studies and the synthesis of seemingly incongruent study findings. Other biases specifically pertaining to meta-analyses include publication bias and statistical biases related to the combination of potentially heterogeneous effect sizes derived from primary studies with diverse samples and methodological approaches (Wolf, 1986). The research design presented below was formulated to address such biases.

The research design of this study followed the steps for quantitative research synthesis outlined by Moody (1990): (a) delineate the domain of study, (b) define admissible studies, (c) locate studies, (d) code and classify study variables, (e) determine a common scale or metric, (f) analyze across studies, (g) interpret and report results, (h) explicate theory and research outcomes, and (i) project future research trends. The first step involved defining the key concepts of spirituality and quality of life as discussed earlier. The conclusions outlined in the discussion provided the basis for the next step of selecting studies that addressed the identified research question and objectives in a manner congruent with the conceptual definitions.

Criteria for Selection of Studies

The selection of relevant studies can be a primary source of researcher bias. One way of overcoming this bias is to formulate explicit inclusion and exclusion criteria that provide unambiguous direction to the selection process. The selection criteria should accurately represent the substantive domain of inquiry as well as consider the degree to which studies with different methodological characteristics can be meaningfully combined into a common metric (Hall, Tickle-Degnen, Rosenthal, & Mosteller, 1994; Lipsey & Wilson, 2001; White, 1994)³. I therefore outlined the selection criteria according to substantive and methodological inclusion criteria and general exclusion criteria while considering Lipsey and Wilson's (2001) recommendation to specify "(a) the distinguishing features of a qualifying study, (b) the research respondents, (c) key variables, (d) research design, (e) cultural and linguistic range, (f) time frame, and (g) publication type" (p. 17). This recommendation formed the basis for the selection criteria used in this meta-analysis as described below.

Substantive inclusion criteria.

The substantive inclusion criteria for the determination of relevant studies were as follows:

1. Descriptive studies that quantitatively estimated the relationship between a person's experience of spirituality and quality of life.
2. Studies that operationalized spirituality and quality of life in a manner that was consistent with the conceptualization and subjective nature of these concepts (i.e., the instruments were not based on external ratings of particular practices or behaviour).

³ Even though effect sizes can be extracted from a wide variety of research designs, Lipsey and Wilson (2001) specifically warn against combining effect sizes derived from vastly diverse research designs such as combining findings from experimental and descriptive studies into a common metric. Nevertheless, a meta-analysis can include primary studies with diverse research designs when different mean effect sizes are calculated to reflect the differences in research design.

In consideration of these substantive criteria, it should be noted that the selection of studies was not constrained by demographic or other sample characteristics because the research question pertaining to spirituality applied to individuals of all backgrounds. Demographic data were however recorded to allow for the examination of potential cultural or religious covariates.

Methodological inclusion criteria.

In addition to these substantive inclusion criteria, the selection of studies was also guided by methodological inclusion criteria. Selected studies were:

1. Quantitative correlational or descriptive studies that provided sufficient statistical data to calculate an estimated effect size of the correlation between spirituality and quality of life (i.e., statistics pertaining to the relationship between spirituality and quality of life were provided).
2. Studies that measured spirituality by using subjective ratings (i.e., self-reported ratings on a Likert-type scale) of spiritual experiences, beliefs or behavior associated with:
 - a. an existential search for meaning *and*
 - b. a relationship with a transcendent reality (e.g., God or divine being, Ultimate Reality, or Ultimate Truth).
3. Studies that measured quality of life by:
 - a. using subjective ratings of a person's self-reported quality of life, satisfaction with life or subjective well-being, *or*
 - b. using instruments that measured quality of life by statistically combining subjective ratings of multiple dimensions of life (e.g., physical, social,

psychological and emotional dimensions). This refers primarily to broad multi-dimensional measures of quality of life or wellbeing.

Methodological inclusion criteria accounting for the quality of studies were not considered in the selection process because the limitations of available research reports did not necessarily enable the coder to accurately evaluate the rigor of each study (Lipsey & Wilson, 2001). The determination of study quality was generally constrained by the extent of the research report. In addition, in consideration of the domain of inquiry and its stage of development, it was found that an insufficient number of highly rigorous studies would be identified. As stated by Lipsey and Wilson (2001), “Many areas of research, especially those that deal with applied topics, provide virtually no perfect studies and the ones closest to textbook standards may be conducted in circumstances that are unrepresentative of those in which the meta-analyst is most interested” (p. 9). Methodological characteristics, however, were coded and thereby accounted for in the statistical analysis.

Exclusion criteria.

The following types of studies were not included in this meta-analysis:

1. Case studies and studies that did not report primary research findings involving human participants.
2. Studies of an exclusively qualitative design.
3. Studies that did not operationalize quality of life, wellbeing or life-satisfaction, or
 - a. Studies that exclusively used diagnostic indicators to measure quality of life (e.g., based on variables such as social economic status or illness-outcomes),

or

- b. Studies that exclusively measured only one dimension of quality of life or wellbeing (e.g., physical, emotional, psychological, social or functional).
- 4. Studies that did not operationalize spirituality, or
 - a. Studies that exclusively measured spirituality in terms of the frequency of religious or spiritual practices or behavior as opposed to an individual's self-rated experience of spirituality.
- 5. Studies that were not reported in English.
- 6. Studies that were published before 1991.

The selection criteria were originally tested and refined by applying them to five pre-selected studies on spirituality and quality of life (Sawatzky, 2002). The final selection criteria were integrated into a selection criteria coding form (see Appendix A) allowing all studies that were reviewed to be labeled according to the criteria. This was done to facilitate subsequent analysis of types of studies that were systematically excluded.

Overview of the selection process.

The selection process started with identifying potentially relevant studies by screening the titles, keywords and abstracts of the citations derived from the comprehensive database search described in the next section. This process was completed by the primary researcher and one other researcher. The exclusion and inclusion criteria were used to distinguish those studies that clearly were not representative of the population of studies of interest to this meta-analysis from studies that were. Studies that were identified as being of potential interest were marked for retrieval.

Several challenges pertaining to the reliable identification of potentially relevant studies became apparent throughout this process. These challenges mostly involved the

difficulty of identifying studies that may have operationalized spirituality and quality of life in accordance with the conceptual foundations of this meta-analysis. Though the manner in which the concepts of spirituality and quality of life were operationalized was usually sufficiently discussed in the research design section of the articles that were reviewed, the original citation information (i.e., the title, the keywords and the abstracts) often did not provide much information about how these concepts were actually operationalized in the study. For example, though I distinguished the conceptualizations of spirituality and religion in my theoretical framework, this distinction was often not that clearly identified in the operational definitions used to measure spiritual and/or religious phenomena. In addition, the terms “religion” or “religiosity” were used in many studies as demographic variables, whereas other studies operationalized “religion” or “religiosity” as an experiential phenomenon associated with religious orientation, religious commitment, religious coping, religious consciousness and religious beliefs. Religion as a demographic variable was clearly not of interest to this meta-analysis. However, studies involving the measurement of religion as a degree of religious orientation or religious consciousness may be of interest depending on the manner in which these concepts were operationalized. The difficulty was therefore to ascertain from the citation information which operational definitions were used to represent the concepts associated with spirituality and/or religion.

To address this difficulty, several guidelines were constructed to expedite the identification of potentially relevant operationalizations of spirituality during the screening process. First, studies using the terms “religion” or “religiosity” were excluded when the terms were exclusively mentioned in a list of demographic variables and when there was no indication in the citation information that the concepts were more broadly operationalized. In

addition, studies that used quality of life or wellbeing instruments that explicitly included a spiritual subscale were only included if the subscale was mentioned as a separate measure in the citation information.⁴ An example is the spiritual subscale of the “Functional Assessment of Cancer Therapy Scale” (Cella et al., 1993) that measures “important aspects of spirituality, such as a sense of meaning in one’s life, harmony, peacefulness, and a sense of strength and comfort from one’s faith” (Brady et al., 1999, p. 420).⁵

Another challenge pertaining to the operationalization of spirituality was that the vagueness of many questionnaires precluded the ability to determine how the questionnaire might have been interpreted by the study participants. Examples include single item instruments based on a question such as, “How spiritual or religious would you say you are?” With these types of questions it was unclear exactly what was measured because we do not know how the term spirituality or religiosity was interpreted by the participants. However, these instruments were of interest based on the assumption that participants may intuitively or naturally have considered the existential and the relational dimensions of spirituality when responding to these types of questions. These instruments were therefore included and coded as ambiguous measures of spirituality, thereby accounting for potential differences between these types of instruments and the more explicitly existential and relational instruments in the analysis.

Similar guidelines were formulated to expedite the process of identifying possibly relevant operational definitions of quality of life. The conceptualization of quality of life was

⁴ It was inferred that bivariate correlational statistics pertaining to spirituality and quality of life were unlikely to be presented when the spiritual subscale of a quality of life instrument was not exclusively mentioned in the abstract.

⁵ Studies including only one quality of life instrument that contained a spiritual subscale were included if the correlation between the spiritual subscale and an independent measure of overall quality of life could be estimated from the data provided in the report.

further clarified to distinguish the measurement of instrumental health outcomes, such as symptoms and physical functioning from “manifestation[s] of ultimate values” that represent the more global concept of quality of life (Green & Kreuter, 1999). Therefore, instruments that exclusively measured biomedical outcomes were not considered to be representative of the concept of perceived quality of life underlying the theoretical framework of this meta-analysis. Similarly, instruments that exclusively measured one domain of quality of life, such as psychological wellbeing, were also excluded.

Again, the challenge was to identify the operational definitions that were used based on the information that was available in the citations. Studies that explicitly mentioned the exclusive use of biomedical or instrumental outcome-measures pertaining to physiological functioning or mental health (e.g., measures of adaptation, coping or depression) as indicators of wellbeing or quality of life were excluded from this analysis. However, some studies operationalized psychological wellbeing by using an instrument measuring satisfaction with life or general wellbeing. Mention of psychological wellbeing in the abstract could therefore not be seen as a sufficient indicator of the type of operational definition that was used in the study. The following rule was designed to address this difficulty: Measures of psychological wellbeing were excluded from this analysis unless there was explicit mention in the citation (title, keywords and abstract) that psychological wellbeing referred to a broader construct of quality of life, general wellbeing or life satisfaction.⁶

⁶ A related concern pertains to those studies that derived a general conclusion about quality of life or wellbeing based on findings from diverse operational definitions, each representing a particular aspect of quality of life or well being. We only included the combination of multiple indicators of quality of life when the indicators were meaningfully and statistically combined by the researcher into a total score representing overall quality of life, wellbeing or life satisfaction (i.e., the instruments were developed with the purpose of combining scores from multiple dimensions into a total score of quality of life or wellbeing).

On the whole, the initial screening of potentially relevant studies was an iterative process throughout which selection criteria and guidelines for selection were clarified and refined. Various ambiguous citations were discussed throughout this process and consensus was reached based on the selection guidelines. Consequently, since this initial screening process relied entirely on the information that was available in the citation provided by the electronic databases, some studies may have been excluded when insufficient or misleading information was presented. Nevertheless, when in doubt, the general rule at this stage was to err toward over-inclusion rather than accidental exclusion of potentially relevant studies.

Search Strategy

The sampling procedures for this meta-analysis were guided by the selection criteria as well as by an extensive search strategy which was designed to construct a sample of studies that was congruent with the research questions and conceptual definitions underlying this meta-analysis. According to Hedges (1994), “The sampling procedure must be designed so as to yield studies that are representative of the intended universe of studies” (p. 35). Hedges explains that, although the notion of “exhaustive sampling” is used to achieve a sufficiently representative sample of studies, one should not assume that the aggregated sample of studies involves the complete extent of variability that may be present in the universe of potential studies pertaining to the domain of inquiry. This relates particularly to the issue of publication bias or the “file drawer problem” as discussed by Glass, McGaw, and Smith (1981). White’s (1994) modes of searching were used to obtain the sample of representative studies. These modes included: (a) footnote chasing: reviewing bibliographies of selected articles, (b) consultation, (c) searches in subject indexes: electronic database searches, (d) browsing, and (e) citation searches of electronic databases.

Computerized Databases Searches.

The computerized database searches were based on White's (1994) recommendation to use "natural language terms" as well as "controlled-vocabulary terms" to compose a comprehensive search statement (p. 49). Natural language terms are those terms directly associated with the domain of inquiry such as may be revealed in abstracts or the full text of potential studies of interest. Controlled vocabulary terms involve subject headings and descriptors that are used for indexing. In addition, Clarke and Oxman (2000) recommended reviewing keywords or subject indexes associated with a pre-selected collection of studies to ensure that the list of search terms is comprehensive as well as specific to the domain of interest. Reviewers also need to consider that different indexes may use different keyword (Clarke & Oxman, 2000). Reed and Baxter (1994) pointed out that the selection of relevant terms should be based on a "clear definition of the topic, which

1. precisely reflects the scope as well as the limits of the search,
2. includes all important concepts,
3. indicates relationships among concepts, and
4. provides criteria for inclusion/exclusion of materials" (p. 60).

Based on these recommendations, a list of terms associated with spirituality and quality of life was constructed to direct the search process. Conceptual terms related to spirituality included the terms spiritual(-well being), religion/religiosity/religiousness, sacred(-ness), transcendent/transcendence, existential, and transpersonal. Conceptual terms that reflected this study's conceptualization of quality of life included quality of life, (subjective) well(-)being, and satisfaction with life or life-satisfaction. These terms were

validated by reviewing the subject headings and descriptors in each of the databases and by reviewing the citation information of several pre-selected studies in relation to each database.

Computerized citation searches also provide a good basis for retrospective and prospective retrieval of potentially relevant published study reports (White, 1994). For example, citation indexes were used to search references for subsequent articles citing a particular source of interest. According to White, “Forward citation searching, still underutilized, has much to recommend it, since it tends to produce hits different from those produced by retrievals with natural language or controlled vocabulary” (p. 52). Citation searches were initiated after the other database searches were completed and a sample of articles for inclusion were identified. Selected articles were then subjected to manual footnote chasing by reviewing the reference lists and, if appropriate, to prospective citation searches using the Social Science Citation Index and the Science Citation Index Expanded (Reed & Baxter, 1994).

The selection of pertinent databases is important in consideration of potential biases introduced by the characteristics of each database. It is generally highly recommended to search multiple databases to retrieve an as complete as possible sample of studies relevant to the research question (Clarke & Oxman, 2000; Reed & Baxter, 1994). Utilizing multiple databases was particularly important in consideration of the multi-disciplinary orientation of my research question (Reed & Baxter, 1994). Both concepts, spirituality and quality of life, have been examined within the contexts of multiple disciplines such as sociology, psychology, medicine and nursing. I included the following databases in my search to reflect the multi-disciplinary nature of this meta-analysis: CINAHL; PubMed; EMBASE; HEALTHSTAR; PsychINFO; SOCIAL SCIENCES CITATION INDEX; SCIENCE

CITATION INDEX EXPANDED and the ARTS AND HUMANITIES INDEX. A brief overview of characteristics pertaining to each of these databases is provided in Appendix B.

Another concern, particularly for meta-analysis, is the importance of limiting the potential of publication bias or “the file drawer problem” by attempting to retrieve unpublished studies (Rosenthal, 1984; White, 1994). The original research proposal was developed to include unpublished studies by searching for dissertations and theses through computerized searches of the Dissertation Abstracts International (Reed & Baxter, 1994). I considered this index, in combination with other standard indexes, such as CINAHL and PsychINFO that also index dissertation materials, as sufficiently broad in scope for the retrieval of dissertation studies. Nevertheless, a comprehensive search of unpublished literature was not very realistic in consideration of Rosenthal’s assertion that “journals are filled with the 5% of the studies that show Type 1 errors, while the file drawers back at the lab are filled with the 95% of the studies that show non-significant (e.g., $p > .05$) results” (p. 108). Although efforts were made to retrieve unpublished studies by searching the dissertation abstracts index, it must be recognized that only a small portion of the unpublished literature was retrieved. Therefore, another method of addressing the file drawer problem was accomplished by computing a “fail safe N” to identify the “number of additional studies in a meta-analysis that would be necessary to reverse the overall probability obtained from our combined test to a value higher than our critical value for statistical significance” (Wolf, 1986).⁷

⁷ Though the original intent was to include dissertations to represent unpublished studies, the dissertations were not included in the final analyses due to limited time and resources. The retrieval and coding of dissertations is however being continued with the intent to include the dissertations in a subsequent analysis thereby addressing the potential concern of publication bias. At the time of this writing, 37 dissertations published in 2000 and 2001 had been selected for retrieval.

Overview and results of the search process.

The above suggestions for effective literature searches were combined to formulate a comprehensive sequential search strategy. My search first began with pre-selecting several studies that clearly reflected the selection criteria. Each database was then searched for these studies to identify keywords and subject headings used to classify potential studies of interest. This process was used to identify substantive as well as methodological terms that were representative of the concepts of spirituality and quality of life as indexed in each database. In addition, a comprehensive review of definitions of medical subject headings (National Library of Medicine, 2002) associated with spirituality and quality of life was conducted to identify subject headings of potential interest. A search string consisting of the following terms was compiled based on the results of these analyses: Spiritual(ity), Religion / Religiosity / Religiousness, Transcendent / Transcendence, Existential, Transpersonal and Sacred(ness), Quality of Life, Well(-)Being, Life-Satisfaction.

Several remarkable findings surfaced when these terms were cross-referenced with the subject headings of each of the pre-selected studies. First, the study by Riley et al. (1998), though containing the word “spiritual” in its title, was not classified under any subject heading associated with the above terms pertaining to spirituality or religiosity in the PubMed database. This finding confirmed the need to search beyond subject headings by including titles and abstracts in the search fields. Second, though EMBASE and PubMed used the same medical subject headings, these subject headings were not consistently applied as determined by the finding that different subject headings were used in EMBASE and PubMed to categorize the same pre-selected studies. In addition, one study, by Kim, Heinemann, Bode, Sliwa, and King (2000), was not indexed in PubMed and

HEALTHSTAR though it was indexed in EMBASE. These findings confirm the need to use multiple databases to identify potential studies of interest. Third, though CINAHL subject headings include the same medical subject headings as MEDLINE, CINAHL also includes many minor subject headings that are especially applicable to spirituality. In addition, the subject headings applied to these pre-selected articles in CINAHL were different from those used in PubMed. The main difference was the inclusion of the subject heading “spirituality” whereas PubMed did not use spirituality as a subject heading for any of the pre-selected articles.⁸

In the analysis of terms and subject headings it was concluded that each subject heading was associated with at least one of the key terms included in the initial search string. No additional terms pertaining to spirituality or quality of life were derived from the review of database subject headings. The initial search string was therefore used to search the titles, abstracts, subject headings and keywords of all databases. The results of this search as well as pertinent characteristics of each of the databases are displayed in Appendix B.

The next step in the search strategy was to compile all the retrieved citations into a single database using the End Note software (Version 4). Each citation was then screened by two coders by manually reviewing the titles, keywords and abstracts in accordance with the exclusion criteria. In total 2,969 citations were screened after the dissertations, publications prior to 1991, and duplicate retrievals were removed from the cumulative database.

In addition to the above search strategies, reference lists of 82 studies of potential interest, including several dissertations, were reviewed to identify additional studies. Several

⁸ An interesting finding resulting from the analysis of medical subject headings was that prior to 2002 the MLN MeSH did not contain any subject heading specific to spirituality. Consequently, most spirituality-related studies before 2001 were classified under other subject headings such as Humanity/Religion and Psychological Phenomena and Processes/Religion and Psychology.

review sources pertaining to topics of religion and spirituality were also consulted. As a result of these reference searches, 26 studies were added to the overall database. An ancestry search was completed by entering the citation information of 44 published studies into the cited references search of the Social Sciences Citation Index and the Science Citation Index Expanded as suggested by White (1994). Based on this search 264 citations were identified. A search of these citations using the key terms identified earlier resulted in a collection of 71 citations. However, 52 of these were already part of the database and of the remaining 19 studies only one study of interest was identified based on the selection criteria.⁹ Finally, several experts in the field were consulted to identify additional studies. However, this process did not result in any additional studies.

The comprehensive database of studies consisted of 3040 citations that were screened in accordance with the selection process described earlier.¹⁰ A total of 371 published studies were selected for retrieval after the screening process was completed. Three of these studies could not be retrieved. A selection criteria form was completed by one of the coders for each of the 368 studies that were retrieved. Of these studies, 59 were excluded because of the exclusion criteria: 1, 2, or 5 (i.e., the report was not a study, the study was entirely qualitative in nature, or the study was not reported in English). An additional 192 studies were excluded because one of the criteria pertaining to the measurement of spirituality and quality of life was not met. A total of 251 studies were excluded from the analysis at this stage.¹¹

⁹ This result can be explained by the fact that a search of the Social Science Index Expanded and the Science Citation Index had already been conducted using the previously identified keywords. The results of this search therefore provided support for the comprehensive nature of the original search. Based on these findings it was decided not to complete the ancestry search for the remainder of the selected studies.

¹⁰ This number is based on 2969 studies from the database searches plus the 71 studies from the ancestry search.

¹¹ An analysis of the exact reasons for which studies were excluded was difficult because not all items of the coding forms were completed when the coder identified a reason for exclusion. In addition, several items of the coding forms were overlapping in meaning (e.g., the methodological categories pertaining to the measurement of spirituality and quality of life overlapped with the exclusion criteria pertaining to spirituality and quality of

The remaining 120 studies were examined in terms of the availability of sufficient statistical data to calculate an effect size. The criterion pertaining to the calculation of an effect size was re-specified so that the reasons for not being able to calculate an effect size were identified. The results of this analysis indicated that 43 studies did not provide any explicit quantitative information about the relationship between spirituality and quality of life. These studies were originally selected because they contained the appropriate operational definitions. However, they failed to report the correlation between the variables of interest. Another 12 studies only reported multiple regression coefficients without providing any measure of the bivariate relationship between spirituality and quality of life. Three more studies reported findings that were statistically insignificant without providing any indication of the magnitude of the relationship. These non-significant results were initially included in the analysis by entering a Pearson correlation of zero for the estimated effect size as suggested by Lipsey and Wilson (2001). One more group of excluded studies consisted of five studies that only reported confounded correlations between combinations of several scales or subscales without reporting the correlation specific to spirituality or quality of life. Based on this selection process, 60 studies (including the three studies with non-significant findings) were identified for inclusion in the final analysis.

Coding and Classifying Study Variables

The next step in the meta-analysis process was to identify those variables that were of potential interest and to code these variables in a manner that allowed for statistical analysis. According to Stock (1994), items for coding should be selected on the basis of substantive as

life). Several studies were discussed throughout the selection process to reach consensus on the disposition of studies for which one of the coders was uncertain. The coding process can therefore not be considered to be entirely independent. Because of this process it was not possible to estimate a truly independent inter-rater reliability coefficient between the two coders.

well as methodological considerations. However, many of these considerations only become apparent as a study progresses and as the researcher becomes increasingly familiar with the domain of inquiry and the statistical challenges and biases that need to be addressed (Stock, 1994; Woodworth, 1994). Developing a code book is therefore seen as an iterative process that develops throughout the data collection phase of the study design (Stock, 1994).

Nevertheless, a priori coding of items can be planned in anticipation of substantive and methodological study characteristics that may systematically affect the homogeneity of effect sizes (Lipsey & Wilson, 2001; Stock, 1994). An awareness of potential sources of variation was needed to identify those types of variables that may have covaried with spirituality and/or quality of life. For example, in light of the discussion distinguishing spirituality from religion it seemed appropriate to code as many as possible religious variables to examine the effect of these variables on the relationship between spirituality and quality of life. A similar argument can be made for potential covariates such as other demographic sample characteristics and methodological differences. In consideration of these potential sources of variance, Stock (1994) suggested using the following categories in constructing a coding form: “report identification, setting, subjects, methodology, treatment, process and effect size” (p. 127).

The code book used to collect data for this meta-analysis (see Appendix C) was based on Stock’s (1994) general recommendations for classifying study variables. Several adaptations were necessary, however, to reflect substantive and methodological considerations associated with spirituality and quality of life. In particular, the treatment

category,¹² as presented by Stock, was adapted to include items pertaining to different operationalizations of spirituality and quality of life. The resulting category was therefore termed “operational definitions.” In addition, the process category was not included because this category did not represent substantive or methodological items of importance to the current meta-analysis.¹³

The general format of the code book used for this meta-analysis consisted of the following categories: report identification, setting, subjects, methodology, operational definitions and effect size. Every category defined in terms of the items it contained and the different levels or options associated with each item were described in the code book (see Appendix C). However, although pre-determined and clearly defined items were relatively easy to code, some items were not as readily classified (Lipsey & Wilson, 2001). For example, I was not able to pre-determine all the different measures of spirituality and quality of life that were used in the sample of selected studies. The diversity of information associated with demographic variables such as ethnicity and religious affiliation also constrained the ability to code these variables in a manner that was consistent across all studies. Some items on the coding form were therefore left open-ended to allow for qualitative descriptions rather than quantitative classifications. These items were analyzed after the coding process was completed to identify potential commonalities and differences that could be coded in a manner suitable for statistical analysis (Lipsey & Wilson, 2001).

¹² The treatment category suggested by Stock (1994) pertains mostly to studies of experimental design. In these types of designs the nature and time of the treatment as well as the characteristics of the control group(s) need to be coded. This obviously does not apply to meta-analyses that exclusively involve descriptive studies.

¹³ The process category as described by Stock (1994) includes items pertaining to the “disposition of coders”, the amount of time it takes to complete the coding form as well as the coder’s confidence in the accuracy of the study report.

Besides identifying and defining coding items, consideration also needed to be given to the organization of items in the coded database. For example, individual studies might have examined several sub-sample characteristics in relation to selected operationalizations of the research variables. Each of these sub-sample-specific relationships between variables might have provided a unique effect size of importance to the meta-analysis and therefore needed to be “coded separately for each effect size” (Lipsey & Wilson, 2001; Stock, 1994, p. 128). This required a flexible coding form allowing for the coding of one or more effect sizes in relation to each primary research study (Lipsey & Wilson, 2001; Stock, 1994). Lipsey and Wilson (2001) therefore recommended separating the coding protocol into two modules: “one module to code information that applies to the whole study and another to code effect size information” (p. 74). In following these recommendations, the final coding form that was used for this meta-analysis consisted of separate report coding forms and effect size coding forms (see Appendix C). The effect size forms were completed for each sub-sample and operational relationship for which an effect size was calculated. These forms were combined into a comprehensive spreadsheet associating each study coding form with its respective effect size coding form(s).

Primary and secondary effect sizes.

Another important aspect of coding effect sizes was to distinguish primary from secondary effect sizes in the final analysis. This was necessary because only effect sizes that were independent from other effect sizes could be used for the calculation of an overall mean effect size (Lipsey & Wilson, 2001). Effect sizes that were derived from the same sample or sub-sample can not be combined in calculating a mean effect size. Therefore, although the coding form allowed for the coding of numerous effect sizes that may have been of interest

for differential analysis, the coders needed to identify the primary effect size to be used to calculate the mean effect size. The following criteria were formulated to guide the selection of the primary effect size:

1. If different operationalizations of spirituality and/or quality of life were used in relation to the same sample, then the primary effect size was selected based on those operational definitions that most closely reflected the operational criteria of spirituality and quality of life underlying this meta-analysis.
2. The primary effect size related to a unique sample (i.e., no other primary effect sizes had been coded in relation to the same sample).
3. If the above criteria did not clearly distinguish the primary ES then:
 - a. I selected the ES that related to the primary instrument as opposed to the subscale, or
 - b. I selected the most conservative approach if there was more than one possible primary instrument (i.e., the smallest ES was selected).

I defined secondary effect sizes as those effect sizes of interest that were not independent from a primary effect size because:

1. they were derived from the same sample or a sub-sample from which the primary effect size was calculated, or
2. they were based on a subscale of an instrument used in the computation of the primary effect size.

Based on these criteria, I was able to select effect sizes that were independent and that reflected the conceptual and operational criteria of spirituality and quality of life discussed earlier.

Coding characteristics of operational definitions.

Another consideration in coding effect size information pertains to the wide variety of instruments used to measure spirituality and quality of life. The specified conceptual and operational criteria were fairly inclusive in nature to avoid systematically excluding findings associated with particular operational definitions. In an attempt to account for the diversity of instruments included in the analysis, it was decided to code each of the instruments in relation to pre-determined categories derived from the conceptual framework underlying this analysis. As described in the coding forms (see Appendix C), quality of life instruments were categorized according to: (a) general measures of multiple dimensions of quality of life, (b) disease specific multidimensional measures of quality of life, (c) global or uni-dimensional (often single item) measures of overall quality of life, (d) measures of well-being, (e) measures of life-satisfaction, and (f) other.

Operational definitions of spirituality were coded in accordance with the existential and relational or transcendent attributes specified by the conceptual definition of spirituality. Due to the prevalence of diverse and often ambiguous operationalizations of spirituality, it became important to distinguish those instruments that clearly reflected the conceptualization of spirituality used in the meta-analysis from instruments that measured the existential and relational or transcendent dimensions of spirituality in a less explicit manner. Instruments measuring spirituality were classified based on their degree of fit between the operational definition and the conceptualization of spirituality used in this meta-analysis. The resulting categories of spirituality instruments were as follows:

1. Explicit measures of existential and relational / transcendent dimensions included those instruments that contained items that clearly measured the existential as well as the relational or transcendent dimensions of spirituality.
2. The category of primarily existential measures of spirituality referred to broad instruments containing items that primarily measured the existential dimension but which may also be interpreted as having relational or transcendent connotations.¹⁴
3. Primarily relational or transcendent measures of spirituality referred to broad instruments containing items that primarily measured the relational dimension of spirituality while also containing items that may be interpreted as having existential connotations. Examples include measures of closeness to God, prayer experiences, meditation, and faith that also have connotations with meaning, purpose or direction in life.
4. Ambiguous measures of spirituality were those instruments that asked people to provide a general rating of their degree of spirituality or religiosity or religious beliefs. It was unclear exactly what was measured here because we do not know how the term spirituality or religiosity was interpreted by the participants. These instruments were of interest based on the assumption that participants may intuitively consider the existential and the relational dimensions of spirituality when responding to these types of instruments.

¹⁴ Instruments that exclusively measured existentialism (purpose and meaning in life) without any reference to spirituality, religiosity, spiritual or religious beliefs, or any other association with the transcendent aspects of spirituality were excluded.

Issues of reliability throughout the coding process.

One last consideration in the coding process involved the avoidance of errors and biases pertaining to the coding of data. Formulating unambiguous descriptions of each coded item was the primary means of avoiding “coding errors” associated with judgment (Orwin, 1994). In addition, mundane coding errors associated with data entry were minimized by avoiding the necessity of reentering or copying data from one database to another. For this reason most of the data were directly entered into an electronic coding form.¹⁵

Despite these preventive measures, statistical analysis of coding error is recommended since coding errors and biases can never be completely avoided. Stock (1994) recommended calculating measures of inter-rater reliability to evaluate the degree of consistency between two coders. However, the only items in the coding form that required some degree of judgment were those two items associated with the descriptive operational categories used to classify instruments measuring spirituality and quality of life and the one item distinguishing primary from secondary effect sizes. These items were rated in accordance with elaborate descriptions of the different codes associated with each variable. These descriptions were refined and discussed throughout the coding process. The same items were re-analyzed after the coding process to identify any inconsistencies. The few inconsistencies that did occur were discussed and a decision about which code to assign was reached by consensus among the two coders and, at times, other members of the research team. The coding of instruments and the identification of primary effect sizes therefore followed a consensus approach which made the calculation of an inter-rater reliability

¹⁵ The second coder completed most coding forms on paper copies after which the data were transferred into the electronic database.

coefficient rather meaningless. The remaining items on the coding forms were descriptive in nature and did not require any ratings based on the coders' own judgments.

Statistical Methods of Analysis

Once the coding process was completed, the next step in the meta-analysis was the selection of an appropriate effect size statistic and the use of statistical methods to combine those effect sizes across studies (Moody, 1990). Various experts in the field of meta-analysis methodology have formulated extensive statistical procedures for calculating effect sizes, weighting mean effect sizes and estimating the effect of other potential covariates. These statistical procedures have been organized in a practical manner by Lipsey and Wilson (2001) in their book, "Practical Meta-Analysis," which provided the primary source of reference for the statistical methods discussed below.

Methods for Calculating Independent Effect Sizes

The first step in applying any statistical procedure to the aggregated findings from primary studies is to determine a standardized scale that is appropriate for the types of research designs included in the sample of studies. The Pearson product moment correlation provides a standardized measure of the direction and degree of relationship between two variables. The Pearson product moment correlation was a convenient effect size parameter for the current meta-analysis, which consisted entirely of descriptive (correlational) studies. With the exception of statistics derived from multivariate designs that only reported multivariate results derived from regression analyses or structural equation modeling, most descriptive findings were readily transformed into a Pearson correlation (Lipsey & Wilson, 2001). The Pearson correlation was also easily transformed into a Fisher's z correlation which in turn allowed us to calculate the standardized variance associated with each effect

size.¹⁶ This standardized variance was then used to weight each effect size in calculating a mean effect size across the selected studies, which could then be transformed back to a Pearson product moment correlation.¹⁷

Before combining the correlations into a mean effect size, Lipsey and Wilson (2001) recommended that the effect of outliers on the distribution of effect sizes be assessed, and that individual effect sizes be adjusted in consideration of common sources of error.¹⁸ The adjustment accounting for differences in reliability was likely most applicable to the current meta-analysis in consideration of the use of many recently developed instruments of spirituality and quality of life. The methods of adjusting for instrument reliability are extensively discussed by Hunter and Schmidt (1990).¹⁹

¹⁶ The Pearson's r to Fisher's z(r) transformation is as follows: $ES_{z_r} = .5 \log_e \left[\frac{1+y}{1-y} \right]$. The standard error is derived from the associated sample size by applying the following formula: $SE_{z_r} = \frac{1}{\sqrt{n-3}}$ (Lipsey & Wilson, 2001).

¹⁷ Lipsey and Wilson (2001) explained that the standard error of the effect size provides a more accurate source of weighting than simply weighting each effect size according to the sample size from which it was derived. Because the standard error is inversely related to the square root of the sample size, the parameter for weighting each effect size is the inverse of the error variance which is calculated as follows: $\omega_{z_r} = \frac{1}{SE_{z_r}^2} = n - 3$

(Lipsey & Wilson, 2001). The mean effect size is then calculated by multiplying each independent effect size by its respective inverse variance weight and dividing the sum of these multiplied values by the sum of all inverse variance weights as described in the following formula provided by Lipsey and Wilson (2001):

$\overline{ES} = \frac{\sum \omega_i * ES_i}{\sum \omega_i}$. The standard error of the mean effect size is simply calculated by taking the square root

of the summed inverse variance weights.

¹⁸ Common sources of error “include adjustment for unreliability of the variables(s) involved in the effect size, for restrictions in the range of those variables, for dichotomization of continuous variables, and for imperfect construct validity” (Lipsey & Wilson, 2001, p. 108). Adjustments such as these need to be made before any further calculations because they adjust each individual effect size as well as the error variances associated with those effect sizes.

¹⁹ Mean reliability coefficients are entered into the following formulas to adjust each effect size as well as the associated standard error: $ES'_r = \frac{ES_r}{\sqrt{r_{xx}} \sqrt{r_{yy}}}$ and $SE'_{z_r} = \frac{SE_{z_r}}{\sqrt{r_{xx}} \sqrt{r_{yy}}}$ where ES_r is the Pearson product

Unfortunately, many researchers did not report reliability coefficients in relation to their specific study sample and simply relied on reliability coefficients reported by other authors. Nevertheless, adjustment for instrument reliability was still desirable even if some of the reliability data were not provided (Lipsey & Wilson, 2001). The following strategies were applied sequentially to provide an estimate of the reliability coefficients pertaining to each effect size. Priority was giving to test-retest coefficients that were estimated in the study from which the effect size was derived. A second choice was to use the Cronbach's alpha coefficients that were estimated in the study from which the effect size was derived. If no reliability information was presented in relation to the sample in the study, then test-retest or Cronbach's alpha coefficients that were reported by the original instrument developers, other studies, or instrument review resources that provided psychometric information pertaining to the instrument were used to provide an estimate of reliability. If no reliability information could be found in the literature then instruments consisting of multiple items received the average reliability coefficient of all instruments; instruments consisting of a single item received the lowest reliability coefficient of all instruments. The minimum value for the reliability coefficients of the spirituality instruments was 0.62 and the mean reliability coefficient of these instruments was based on 53 reported reliability coefficients ($r = 0.84$). The minimum reliability coefficient entered for the quality of life instruments was 0.56 and the average of the 53 available reliability coefficients was 0.81. The reliability coefficients used for calculating each of the effect sizes are outlined in Appendix D.

moment correlation prior to the Fisher's $z(r)$ transformation, and r_{xx} and r_{yy} refer to the reliability coefficient of each measure. ES'_r is transformed into a Fisher's $z(r)$ which is then entered into the meta-analysis equations (Lipsey & Wilson, 2001). SE'_{Z_r} refers to the adjusted standard error that is calculated after the transformation to Fisher's $z(r)$. These methods are discussed in detail by Hunter and Schmidt (1990).

Accounting for Variance in the Distribution of Effect Sizes

The next step in the meta-analysis was to account for variance in the distribution of the sample of effect sizes. Clearly, a mean effect size only has meaning when there is indication that the assumptions underlying parametric statistical tests have been met.

Outlier analysis.

Conventional outlier analyses were performed in accordance with recommendations offered by Tabachnick and Fidell (1996) to account for multiple sources of methodological error or poor validity of operational definitions that did not reflect the conceptualizations of spirituality and quality of life underlying this meta-analysis (Lipsey & Wilson, 2001). The degree of dispersion of outliers as well as their effect on the resulting mean effect size was examined to determine the appropriate approach toward constraining their effect.

Analysis of homogeneity of variances pertaining to the distribution of effect sizes.

The next assumption involved the degree of homogeneity of variance within the distribution of primary effect sizes. This homogeneity of variance assumption was examined by determining the significance of the Q statistic, which provides an estimate of the degree of heterogeneity in the distribution. The Q statistic was calculated by comparing the error variances associated with each effect size and the significance was determined using a Chi-Square distribution.²⁰

Naturally, the effect sizes derived from fairly diverse sample characteristics and operational differences associated with the measurement of spirituality and quality of life were unlikely to be homogeneously distributed. One way of addressing this issue was to use a

²⁰ The homogeneity analysis was calculated as follows: $Q = \sum \omega_i (ES_i - \overline{ES})^2$. The significance of Q was derived from a chi-square table where a significant Q was indicative of a heterogeneous distribution (Lipsey & Wilson, 2001).

“fixed effects model” of analysis and to employ regression analysis methods to identify all variables that had systematic effects on the effect size distribution (Lipsey & Wilson, 2001). However, identifying such variables was dependent on the ability to determine a priori all moderating variables so that they could be coded and entered into the equation thereby accounting for all systematic sources of variance and obtaining a homogeneous distribution of residual variance. Another assumption underlying a fixed effects model is whether the universe of all conceivable studies that address the study question is sufficiently similar to the sample of studies used for the analysis (Hedges, 1994). In other words, “The studies gathered for a meta-analysis are a representative sample of a known universe with known characteristics” (Hunt, 1997, p. 107).

The assumptions underlying the fixed effects model were difficult to support considering the conceivably diverse sample characteristics and operationalizations associated with spirituality and quality of life. An alternative was the use of a random effects model where, in consideration of the wide diversity of studies, it could be assumed that sources of variance associated with the distribution of the effect sizes were likely to be randomly distributed. Hedges (1994) explained that in this model “the study sample is presumed to be literally a sample from a hypothetical collection (or population) of studies” (p. 31). The variability in the random effects model is therefore much larger and will result in a more conservative combined effect size.²¹ An additional benefit of the random effects model is that generalizations based on the findings of a random effects model can be applied to a large

²¹ The random effects model adds an estimate of random variances to the equation of total variance. The random variance is computed based on the Q statistic and the inverse variance weight as follows:

$$v_o = \frac{Q - (k - 1)}{\sum \omega_i - (\sum \omega_i^2 / \sum \omega_i)}$$

The total variance is calculated by adding the sampling error and the random variance. The total variance is then entered into the equation for calculating the inverse variance weight, the mean Fisher’s z(r), and the Q statistic for homogeneity as described earlier.

variety of situations that do not need to reflect the particular characteristics of the sample of studies used in the meta-analysis.²²

Assuming the existence of random sources of variance does not exclude the existence of systematic sources of variance (i.e., variables that co-vary with spirituality or quality of life). Certainly one of the strengths of meta-analysis is to examine potential covariates or moderator variables. For this reason, much attention was devoted to designing a coding tool that allowed for the quantification of such variables. Regression analyses could have been used to identify the degree to which differences among studies systematically contributed to the overall variance (Lipsey & Wilson, 2001). However, regression analysis strategies do not account for random sources of variance and are therefore based on the fixed effects model. An alternative to a regression analysis based on a fixed effects model is the mixed effects model, which is based on the assumption that “the effects of between-study variables, ... [such as different operational definitions], are systematic but that there is a remaining unmeasured (and possibly unmeasurable) random effect in the effect size distribution in addition to sampling error” (Lipsey & Wilson, 2001, p. 124). The mixed effects model is based on entering systematic sources of variance, random sources of variance, and variance associated with expected sampling error into the overall equation. This approach allowed for the identification of potential moderator variables associated with the relationship between spirituality and quality of life while also allowing for sources of random variance. This model made the most sense in consideration of the desire to identify potential moderator variables while at the same time recognizing the existence of diverse sources of variance associated with broad conceptual relationships between spirituality and quality of life. The

²² These formulas were integrated into an Excel spread sheet to facilitate the computation of random and fixed effects model analyses.

mixed effects model was therefore used in the final analysis of factors affecting the relationship between spirituality and quality of life.²³

Preliminary Analyses

The selection process described earlier resulted in a sample of 60 studies. This sample was reviewed and analyzed by examining potential discrepancies in the selection process. The first step in this analysis involved a thorough review of all the instruments used to represent the two primary concepts of interest. Initially, the inclusion of studies was determined by the coder who based his or her decision primarily on the information available in the research reports. The coders took the approach of including all instruments of potential interest. Subsequent analysis of the extent to which each instrument represented existential as well as relational or transcendent aspects of spirituality as specified in the selection criteria was based on a thorough review of the instrument items, and a search for original publications and subsequent psychometric testing of the instruments. Several studies were excluded as a result of this analysis.

Analysis of Selection Process

The first group of excluded studies included six studies in which the relational or transcendent dimensions of spirituality were not represented in any of the instruments. Five of these studies examined the Existential Subscale of the McGill Quality of Life Questionnaire (MQOL) (Cohen, Mount, Strobel, & Bui, 1995) in relation to quality of life (Cohen et al., 1997; Cohen, Hassan, Lapointe, & Mount, 1996; Cohen et al., 1995; Cohen, Mount, Tomas, & Mount, 1996; Pratheepawanit, Salek, & Finlay, 1999). The inclusion of the

²³ Explicating the statistical methods associated with the mixed effects model goes beyond the scope of this report. The mixed effects analysis was performed by using an SPSS macro provided by Lipsey and Wilson (2001). The macro uses matrix algebra formulas to estimate the random variance component.

MQOL was questioned by both coders throughout the coding process. The MQOL was not included in the final analysis because its items are exclusively existential in nature and do not represent the relational or transcendent defining characteristics of spirituality. The existential subscale consists of six items pertaining to the meaningfulness of life, achievement of goals, life worth, control over life, feeling good about oneself, and the degree to which every day is perceived as a burden or gift. The grouping of these items into a single scale was based on factor analyses by Cohen and Hassan et al. (1996) and Cohen and Mount et al. (1996). The scale was fittingly conceptualized by the authors as a measurement of existential wellbeing.

One other excluded study (van Wegberg et al., 1998) also used another quality of life instrument that contained an exclusively existential subscale. The eight-item spiritual subscale of the SELT-M (van Wegberg et al., 1998) was originally developed in the German language. Sample translated items of this scale include, “It is difficult for me to see positive meaning in my illness”, and “I have found new goals in my life” (p. 1095). The relational or transcendent aspects of spirituality are not represented in any of the items.²⁴

Another other study was excluded because of inconsistencies in the operational definitions of quality of life, well-being, or life-satisfaction. The study by Ayele, Mulligan, Gheorghiu and Reyes-Ortiz (1999) used the Life Satisfaction Index-B (Neugarten, Havighurst, & Tobin, 1961). This study was excluded because the scale underlying this instrument is based on the researcher’s rating of participants’ responses to open-ended questions pertaining to life satisfaction (Bowling, 1997). The scale is therefore not a specifically self-rated scale as specified in the inclusion criteria.

²⁴ A similar argument for exclusion may be made for the spiritual subscale of the WHOQOL-100 (WHOQOL-Group, 1998), which also consists of items measuring meaning and purpose in life. However, the prompt of this scale specifically refers to spiritual and / or religious beliefs which may be interpreted as having transcendent connotations. The three studies using the spiritual subscale of the WHOQOL-100 therefore were included (Gioiella, Berkman, & Robinson, 1998; Pippalla & Char, 2001; WHOQOL-Group, 1998).

Finally, a study by Bowman and Sanders (1998) was not included in the analysis of primary effect sizes because this study reported results from a sub-sample of the same general sample used by Levin and Taylor (1998). These effect sizes could therefore not be considered to be independent. The study with the largest sample size was coded for the primary effect size. In addition, two reports by Peterman, Fitchett, Brady, Hernandez and Cella (2002) and Brady, Peterman, Fitchett, MO and Cella (1999) reported findings from the same study. The findings from these two reports are therefore not independent. The second report by Peterman et al. (2002) was therefore coded as a secondary effect size and not included in the analysis.

All these decisions were consistent with decisions made throughout the selection process and served the purpose of creating a sample of studies that represented the conceptualizations of spirituality and quality of life discussed earlier. On the whole 62 primary effect sizes from 51 studies were included in the subsequent analysis.

Analysis of distribution of effect sizes

After compiling all the information from the final selection of studies in a comprehensive database, the distribution of primary effect sizes was analyzed for skewness and outliers in accordance with standard statistical recommendations (Tabachnick & Fidell, 1996). As demonstrated in Figure 2, the original distribution of unadjusted correlations was positively skewed (Skewness statistic = 1.122, SE = 0.304; Kurtosis statistic = 1.521, SE = 0.599) thereby threatening the assumptions underlying parametric testing. As demonstrated in the bar-graph below, the positive skew appears to be caused by an accumulation of studies around the zero correlation value as well as several studies with extremely high correlations (see Figure 2).

Outliers.

Further analysis of the distribution of the 62 primary effect sizes revealed three potential outliers that contributed to the positively skewed distribution (see Figure 3).

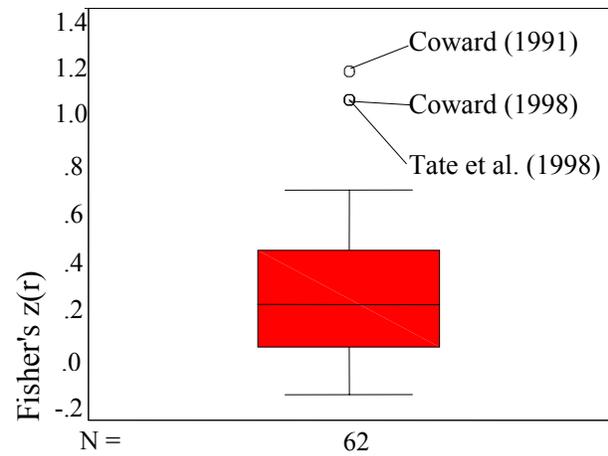
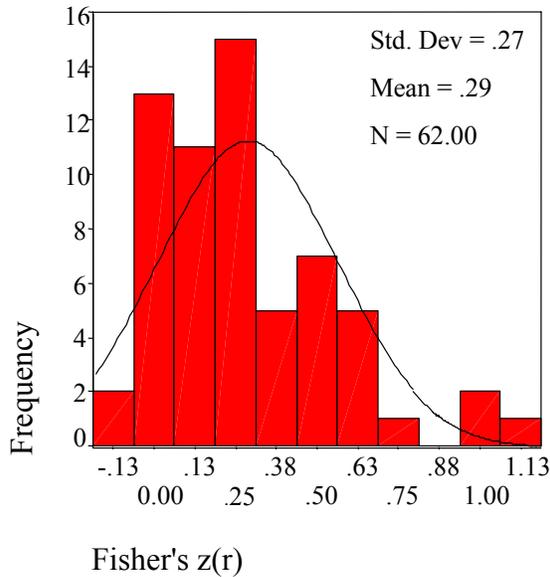


Figure 2. Histogram of the Distribution of Correlations

Figure 3. Box Plot with Outliers

Remarkably, two of the outliers were derived from studies by the same author (Coward, 1991, 1998). In both of these studies spirituality, represented by the Self-Transcendence Scale (Reed, 1987), was examined in relation to life satisfaction measured with the Cognitive Well-Being Scale (CWBS) (Coward, 1990) in one study (Coward, 1998) and a compilation of the CWBS and Bradburn's Affect Balance Scale (ABS) (as cited in Bowling, 1997) in the other study (Coward, 1991). The life satisfaction measure in the study from 1991 was originally included because the combined scales included two items from the CWBS that directly measure life satisfaction. However, the combination of the CWBS and the 10-item

ABS clearly resulted in a confounded measurement of life satisfaction and affect. The combined scale is therefore less congruent with other more explicit instruments of life satisfaction included in the sample of selected studies. The Affect Balance Scale used on its own was systematically excluded throughout the selection process. This study is therefore not congruent with the population represented by all the other studies. Outliers for which there is reason to believe that they do not represent the intended population can be safely removed from the analysis (Lipsey & Wilson, 2001; Tabachnick & Fidell, 1996).

Exclusion of the other two outliers is, however, not as readily justified because the instruments used in these studies are clearly representative of the relevant concepts. Removing these outliers may result in an introduction of bias and less generalizable results (Tabachnick & Fidell, 1996). One of these outliers was associated with the second study by Coward (1998) that resulted in a high Pearson correlation of 0.82 in a sample of 16 women diagnosed with breast cancer who participated in a breast-cancer support group. Not surprisingly, the small sample size used in this study caused this effect size to be very imprecise (95% CI: 0.53 - 0.94). However, other than the exceptionally small and narrowly defined sample characteristics, no differences between this study and other studies using comparable instruments were identified to explain the high correlation. The same combination of instruments was used in a third study by Coward (1996) involving a 143 individuals who attended a health-related event for the general public. This study resulted in a correlation of 0.60 between the Self-Transcendence Scale (STS) (Reed, 1991) and the Cognitive Well-Being Scale. The STS was also used by Mellors, Riley, and Erlen (1997). The correlation between the STS and Ferrans and Powers' (1992) Quality of Life Index used

in this more recent study was much smaller ($r = 0.46, p < 0.01$), and thus consistent with the range of correlations found in the 62 selected estimates.

The approach used for this outlier was therefore to constrain its extraneous effect by reducing its value to the next group of effect sizes in the distribution as suggested by Lipsey and Wilson (2001). Examination of the distribution as depicted in Figure 4 reveals that the two remaining outliers clearly represented a break in the distribution. The two effect sizes on the other side of this break had values (after Fisher's transformation) of 0.69 and 0.65. The most conservative approach was to reduce the impact of this outlier by reducing its effect size to the value of the last study before the break in the distribution. Remarkably, the study in closest proximity to the group of outliers was also a study by Coward (1996) resulting in a correlation of 0.60 between the Self Transcendence Scale and the two life-satisfaction items of the Cognitive Wellbeing Scale discussed earlier.

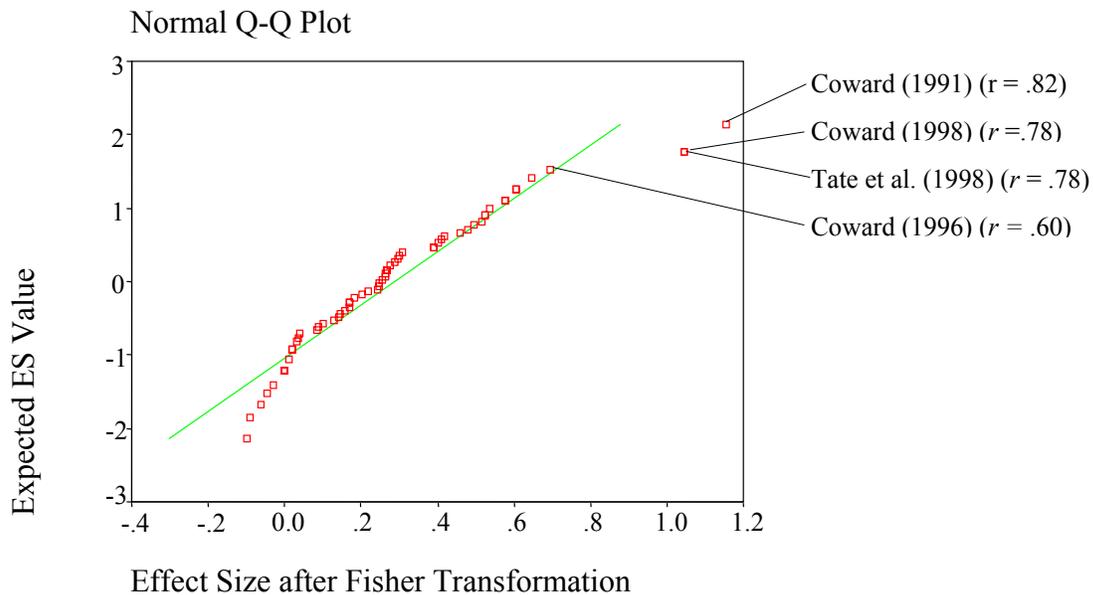


Figure 4. Comparison of Outliers to Regression Line

The same approach was used for the third outlier which was associated with a study by Tate, Riley, Perna, and Roller (1997) who examined the FACT-SP (Cella, 1997) in relation to various scales measuring quality of life and life satisfaction in a sample of 219 people with traumatic spinal cord injury, traumatic amputation of lower limbs, postpolio syndrome, recurrent breast cancer or recurrent prostate cancer. Most of the reported data were derived from regression analyses for which no zero-order correlation was reported. The only bivariate correlation presented in this study was between the FACT-SP and the FACT-G (Cella et al., 1993).²⁵ No explanations for the high correlation of 0.72 in a sample of 66 men and women with various chronic conditions could be explained by the sample or design characteristics. The FACIT-SP (or FACT-SP) was correlated with the FACT-G in one other study reported by Peterman et al. (2002) and Brady et al. (1999) resulting in bivariate correlation coefficients of 0.58 ($p < 0.0001$) for a sample of 1,610 people diagnosed with cancer and/or HIV/AIDS. The instruments and other characteristics of these studies are clearly representative of the intended population of studies for this meta-analysis. This outlier was therefore also not excluded but its impact was reduced using the same approach as discussed earlier.

Non-significant unreported correlations.

The positive skew of the distribution can also be explained by the inclusion of three studies that resulted in a statistically non-significant and unreported correlation. These studies were initially coded as having a correlation of zero. However, an examination of the confidence intervals of these studies, assuming a correlation of zero, reveals that two of the three studies had a power so low that the 95% confidence interval was -0.47 to $+0.47$ for the

²⁵ Based on the presented findings and the use of this scale in other studies, it can be assumed that the FACT-G total score excluded the items from the FACT-SP. The FACT-SP is therefore seen as a separate measure.

first study (Gioiella et al., 1998) and -0.35 to $+0.35$ for the second study (Post-White et al., 1996). Both of these studies did not have sufficient power to identify a statistically significant correlation below 0.35 at $p < 0.05$. These two studies were therefore excluded from further analysis because it was impossible to reliably estimate a correlation due to the large amount of variance and associated small sample sizes ($N = 18$ and $N=32$, respectively). The third study with a statistically non-significant result (Harvey, Bond, & Greenwood, 1991) was based on a sample of 137 participants resulting in a 95% confidence interval of ± 0.168 . Entering a zero value for this correlation is much more accurate than for the other two studies. This study was therefore left in the analysis in unadjusted form.

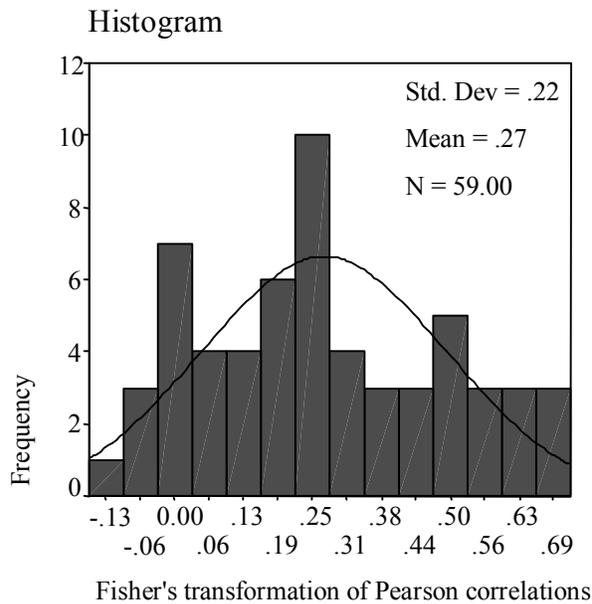


Figure 5. Distribution of Correlations after Adjustment For Outliers and Non-Significant Correlations that were not Reported.

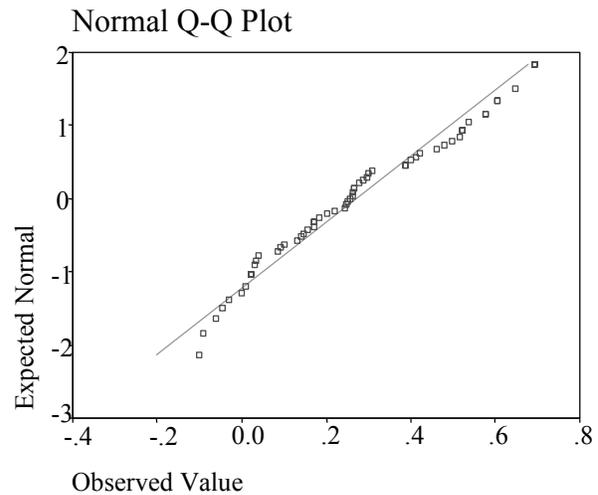


Figure 6. Comparison of Adjusted Values to Regression Line.

After these adjustments for outliers were made the distribution of correlations became roughly normal with no further outliers identified (see Figures 5 and 6). The remaining

skewness and kurtosis statistics were favorable at values of 0.278 (SE = 0.311) and -0.896 (SE = 0.613), respectively.²⁶

Confidence intervals of original effect sizes.

The preliminary analysis also involved an examination of the distribution of the 95% confidence intervals associated with each of the primary effect sizes.²⁷ A graph of each effect size with its accompanying confidence interval provides a useful overview of the generally large amount of variability within each of the studies (see Figure 7). Twenty-five studies with statistically insignificant results are easily identified by those confidence intervals that cross the zero value. This exemplifies one of the primary advantages of a meta-analysis; though no statistically significant conclusions could be drawn from each of these studies in isolation, the aggregation of their findings results in a substantial reduction in error variance and a statistically significant mean effect size as discussed in the following section. Finally, because the y-axis of this graph represents equal increments between each of the studies, the graph also functions as a Q – Q plot clearly revealing the three potential outliers as deviating from the imaginary regression line that may be drawn through each of the effect size values.²⁸

²⁶ It must be noted that the final analyses were performed on correlations that were adjusted for measurement error as described earlier. The distribution of these correlations was again more positively skewed which may be explained by the fact that, on average, extreme correlations are more strongly adjusted than smaller correlations resulting in a wider overall distribution. However, the distribution of correlations adjusted for measurement does not present an accurate reflection of the actual distribution because the variance of these correlations was also adjusted in the analyses. This means that the correlations that received the largest adjustment for instrument reliability (causing the distribution to be positively skewed) also received less weight in the analysis (due to a decrease in their inverse variance weights). Only the distributions of the observed values (after Fisher's transformation) are presented in the plots.

²⁷ The confidence intervals were calculated by adding or subtracting a z-value of 1.96 multiplied by the standard error to the Fisher correlation and then transforming the Fisher correlations back to the Pearson correlations to facilitate interpretation using the formulas discussed in the methods section.

²⁸ For ease of interpretation, this error bar graph presents the values based on the actual Pearson correlations (before Fisher's transformation).

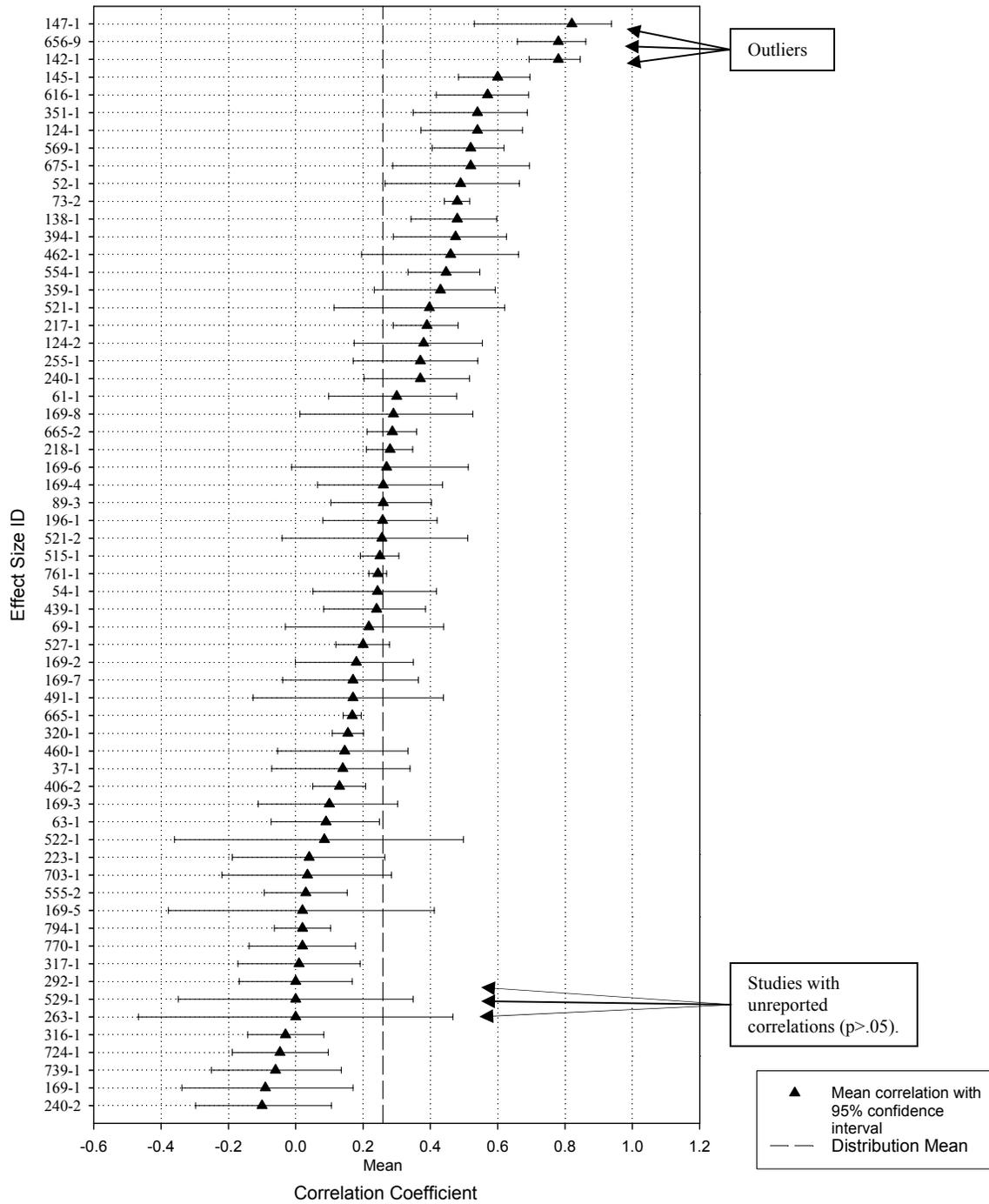


Figure 7. 95% Confidence Interval for Unadjusted Correlations

Results

Sample Description

The overall sample consists of 59 primary effect sizes derived from 48 studies with a total sample of 22,554 individuals. Several studies were oriented toward individuals with various health-related conditions. Of the total sample of participants, 801 individuals were taken from a population of people with a chronic medical condition, 159 individuals were defined as having a mental illness, 98 individuals were diagnosed with HIV/AIDS, 1,700 individuals were diagnosed with a form of cancer and 1 study with 58 people specifically focused on individuals receiving palliative care. The remaining samples were not associated with a particular health condition. However, several studies specifically drew their samples from specific ethnic groups as discussed later.

Aggregating the demographic characteristics of all the studies resulted in a diverse representation of people from various cultural and religious backgrounds. However, the demographic sample characteristics of many primary samples could not be determined from the study reports. Information about ethnic background was available for only 12,413 individuals (55% of the total sample). Of these, 8,396 (67.6%) individuals identified themselves as “white” or Caucasian, 3,238 (23.7%) were classified as black or “African American”, 716 (5.8%) as South American, 96 (0.8%) as Asian, 46 (0.4%) as Hispanic and 120 (1.8%) individuals were classified as “other”. The meaning of this type of demographic information in a meta-analysis is questionable because the method by which demographic information is collected and the manner in which ethnic and religious classifications are made are unique to each of the primary studies. Nevertheless, the presented overview of demographics does suggest that individuals labeled “Caucasian” or “White” were probably

overrepresented in the combined sample, which brings to question the generalizability of the findings to different cultural or ethnic groups.

Not surprisingly, the distribution of religious demographics was similar to the distribution of ethnic groups. Information about religious affiliation of almost one half of the studies ($n = 25$) could not be determined because religious information was either entirely omitted from the study reports, or because the effect size was associated with a subsample while demographic information was only available for the overall sample of participants. As a result religious information was available for only 19.7% ($n = 4,453$) of the 22,554 participants represented in this meta-analysis. Of the 4,453 participants for which religious data were available, 1,547 individuals in the total sample described themselves as being affiliated with a Protestant religion (34.7%), 1,569 individuals were affiliated with a Catholic religious background (35.2%) and the Jewish, Hindu and Islam religions were represented by 180, 4, and 3 individuals, respectively (4.1%, 0.1%, and 0.1%, respectively). In addition, 399 individuals (9.0%) described themselves as having no religious affiliation, 665 individuals were classified as “other” (14.9%), and the religious backgrounds of 86 individuals were unknown (1.9%).

Description of Operational Definitions

One of the most important findings of this meta-analysis pertains to the instruments that were used to measure spirituality, quality of life, well-being and life-satisfaction. A description of each instrument representing spirituality is provided in appendix E. On the whole, 13 instruments associated with 20 primary effect sizes consisted of items that explicitly represented existential and relational dimensions of spirituality as discussed earlier. Twelve primary effect sizes were based on five spirituality instruments that were mostly

existential in nature while also containing items with relational or transcendent connotations. Six spirituality instruments associated with 14 primary effect sizes were primarily relational or transcendent in nature, and the category of ambiguous measures of spirituality or religiosity was represented by ten instruments used to determine 13 effect sizes pertaining to the relationship between spirituality and quality of life.

Measurements of quality of life included five instruments that were coded as general multidimensional instruments of quality of life, five instruments representing the disease specific multidimensional quality of life category, three instruments measuring overall (not multidimensional) quality of life, two instruments measuring wellbeing, and 13 instruments reflecting various approaches to measuring life satisfaction. Two instruments used in studies by Beery, Baas, Fowler and Allen (2002) and Landis (1996) were classified as “other” because they did not correspond explicitly with any of the five categories. The first instrument, the Index of Wellbeing by Campbell, Converse, and Rodgers (1976) consisted of nine items including one item measuring life satisfaction and eight other items such as, “My present life is enjoyable ... miserable” and “My present life is useless ... worthwhile” (p. 13). The other instrument, the Psychosocial Adjustment to Illness Scale by Derogatis (as cited in Derogatis Psychological Tests, n.d.), which was used in a study by Landis (1996), measured quality of life in terms of psychological adjustment. However, though this instrument was considered to be distinct from other quality of life instruments in that it measures psychological adjustment, the 46 items represent seven domains that are commonly associated with quality of life. The domains of the PAIS are as follows: (a) “Health Care Orientation,” (b) “Vocational Environment,” (c) “Domestic Environment,” (d) “Sexual

Relationships,” (e) “Extended Family Relationships,” (f) “Social Environment,” and (g) “Psychological Distress” (Derogatis Psychological Tests, n.d.).

Mean Effect Size

The mean effect size was calculated to provide an indicator of the presence and strength of the relationship between spirituality and quality of life. The analysis of primary effect sizes adjusted for instrument error²⁹ was indicative of a moderate relationship between spirituality and quality of life represented by a Pearson correlation of 0.341 and a 95% confidence interval between 0.282 and 0.398 using a random effects model of analysis (see Appendix F).³⁰ As predicted in my earlier discussion on statistical methods, the homogeneity analysis resulted in a significant Q-value ($Q = 635.29$, $df = 58$, $p < .0000$), which indicates that the variance in the sample of effect sizes are heterogeneously distributed and could not be accounted for by sampling error alone. The assumptions underlying the fixed effects model are therefore not met and the random effects model was used for the remaining analyses (Hunter & Schmidt, 1990; Lipsey & Wilson, 2001).³¹

The mean effect size was also examined based on an analysis of primary effect sizes prior to adjustment for instrument reliability to examine the potential effect of these adjustments on the overall findings. The resulting mean effect size of 0.256 (95% CI: 0.214 - 0.298) was substantially lower than the mean effect size based on the adjusted correlations. Again, the null hypothesis of homogeneity was not met as represented by a significant Q-

²⁹ See Appendix D for an overview of the different reliability coefficients used to calculate the adjustments as described in the statistical analysis section.

³⁰ The effect size values presented in this discussion are based on the Fisher’s transformed z-values which have been reversed back to Pearson’s correlations for ease of interpretation. Both the Fisher’s Z(r) values and the Pearson’s correlations are presented in Appendix F.

³¹ The possibility exists that this additional variance can be accounted for by moderating factors incorporated into a fixed effects model. However, data collected on potential moderating variables only accounted for part of the variance beyond sampling error as demonstrated in subsequent regression analyses. The significant residual Q-value indicates that the residual variance remains heterogeneously distributed ($Q = 448.23$, $df = 49$, $p < .0001$), which provides support for the use of a random or mixed effects model.

value in the fixed effects model ($Q = 476.06$, $df = 58$, $p < .0001$). The reduction in the mean effect size indicates that the reliability coefficients had a positive effect on the overall analysis. This is not surprising when one considers that the adjustments for reliability have the average effect of widening the distribution and increasing the error variances.³²

Nevertheless, the mean effect size adjusted for instrument reliability would be considered to more accurately reflect the magnitude of the relationship between spirituality and quality of life because part of the systematic variance associated with instrument reliability is removed from the overall analysis (Hunter & Schmidt, 1990; Lipsey & Wilson, 2001).³³

A final mean effect size analysis was performed on the overall sample of 62 primary effect sizes before the removal and adjustment of the three outliers and two non-significant unreported correlations to determine the impact these adjustments had on the overall calculations.³⁴ In this analysis, the mean effect size prior to adjustment for instrument reliability was estimated to be 0.262 with a 95% confidence interval between 0.218 and 0.304 and a significant Q -value 536.45 ($df = 61$, $p < .0001$) using a random effects model. Comparison with the mean effect size that was not adjusted for instrument reliability

³² As discussed earlier, the mean effect sizes were adjusted for measurement error by dividing each correlation by the square roots of the reliability coefficients of both variables resulting in an increase of the primary effect sizes and widening the distribution (extreme values receive a proportionally larger increase than small values). The standard errors associated with each effect size were adjusted using the same approach resulting in an increased error variance or a decreased inverse variance weight (Hunter & Schmidt, 1990).

³³ Hunter and Schmidt (1990) provide an additional method of attenuating the overall effect size for instrument reliability based on the average reliability adjustments across all studies. However, due to the large amount of single-item instruments in the present meta-analysis, this approach was considered to be less accurate because the single item measures, which are on average less reliable, would receive the same treatment in the analysis as the multi-item measures.

³⁴ Because of the large coefficients and the accompanying small reliability coefficients associated with the outliers, this analysis could only be performed based on effect sizes that were not adjusted for instrument reliability because subsequent adjustments would cause the estimated correlations to be above one (which is statistically meaningless and which does not allow for the necessary Fisher's transformation).

indicates that the effect of the outliers was minimal in the overall analysis.³⁵ In addition, the comparison of the three effect sizes indicates that the correlation between spirituality and quality of life is fairly stable across the three different analyses, which provides increased confidence in the resulting moderate strength of the relationship as defined by Cohen (1988).³⁶

Our confidence in the overall precision of the effect size is also supported by calculating a “failsafe N” as originally suggested by Rosenthal (as cited in Lipsey & Wilson, 2001). The formula provided by Orwin (as cited in Lipsey & Wilson, 2001; Hunter & Schmidt, 1990) for calculating a failsafe N was used to estimate the number of studies with non-significant findings needed to reduce the effect size to a criterion value.³⁷ The results indicate that an additional 46 studies with an effect size value of zero would be needed to reduce the effect size toward the value of 0.20 which would still be considered a moderate effect size. To reduce the effect size to a small value of 0.10 we would need to find 151 studies with effect sizes that approximate zero. This brief analysis provides additional confidence in the finding of a moderately strong relationship between spirituality and quality of life. The probability of finding 151 published studies that resulted in a close to zero correlation between spirituality and quality of life seems small considering the exhaustive

³⁵ Comparison is made between the mean effect sizes that were not adjusted for measurement error because adjustment for instrument reliability caused the Pearson correlations to be adjusted to values larger than thereby making a Fisher’s $z(r)$ transformation impossible (and statistically meaningless).

³⁶ In his work on power analysis, Cohen (1989) provided the following guidelines for interpreting the strength of relationships between two variables based on Pearson’s correlation which have been commonly applied to many research findings: small effect size ($r \leq 0.10$), medium effect size ($r = 0.25$), large effect size ($r \geq 0.40$).

³⁷ The fail-safe N was calculated in accordance with Orwin’s recommendation by using the following formula:

$$Fail - safeN = k \left[\frac{\overline{ES} Z_r}{\overline{ES}_c} - 1 \right]$$

where k indicates the number of studies used in the analysis and \overline{ES}_c is the value of the criterion mean effect size (Orwin as cited in Lipsey and Wilson, 2001). It must be recognized that this approach only provides a crude indicator of the reliability of the study findings because the formula does not take into account the variances within each of the primary studies.

search process used to finding the current collection of studies. Nevertheless, it is of course possible that such studies do exist in the non-published literature.

Analysis of Moderator Variables

Besides merely providing an indicator of the strength of the relationship between spirituality and quality of life, meta-analysis techniques can also be used to examine the potential moderating impact of other variables. This is particularly important considering the fairly large amount of unexplained variance which is assumed to be attributed to random sources in the random effects model. The mixed effects model allows us to reduce this random variance as much as possible by examining factors that may have a systematic effect. One of these potential moderating variables was already discussed by comparing the mean effect size adjusted for instrument reliability with the unadjusted mean effect size. The significance of the effects of moderators such as instrument reliability, other design characteristics and sample characteristics was examined through single factor ANOVA and multivariate linear regression analyses assuming a mixed effect model as discussed in the methods section.³⁸

Potential indicators of study quality.

Data on several study quality indicators were collected to examine whether some of the variance between studies could be explained by differences associated with the peer review status of the journal in which the study was published, the sampling method, the research design, and the percentage of attrition. In terms of peer review status, all the studies

³⁸ Multivariate analyses assuming a fixed effects model were also performed to examine whether the random variance component could be completely eliminated from the model. However, combining all the factors discussed in the following section (except gender) into a multivariate regression analysis based on a fixed effects model resulted in a comparatively large and significant Q-value for the residual variance which indicates that there are other sources of variance not accounted for by the model ($Q_{Residual} = 448.74$, $df = 49$, $p < .0001$) (Lipsey & Wilson, 2001).

that were selected for primary effect size calculations were published in peer reviewed journals as determined in Ulrich's International Periodicals Directory. For sampling method, the collected data were dichotomized and dummy coded according to probability sampling methods and non-probability sampling methods. The studies with probability sampling ($n = 7$) had a slightly lower mean effect size ($r = 0.31, p = 0.0055$) than those with non-probability sampling methods ($r = .35, n=52, p < 0.0001$). The effect of probability versus non-probability sampling methods was shown to be not statistically significant when the dummy coded variable was entered as a single variable in a regression analysis ($p = 0.6813$). However, the insignificant results of the regression analysis may be partly attributed to a lack of power³⁹ associated with the small number of studies using probability sampling methods. The results of the present analysis are therefore inconclusive.

In terms of research design, all the effect sizes were based on descriptive results that did not involve any experimental manipulation. One study by Coward (1998) was based on a quasi-experimental design where data were collected before and after a therapeutic intervention for women diagnosed with breast cancer. However, the effect size was based on the baseline data collected prior to the intervention because the data after the intervention were obviously confounded by the intervention itself. Similarly, though four other studies were based on longitudinal designs (Fry, 2001; Kim et al., 2000; Maton & Zimmerman, 1992) the bivariate correlations used for the primary effect sizes were derived from data collected at one point in time. Two other studies used comparative descriptive designs to examine various variables across different groups (Riley et al., 1998; WHOQOL-Group, 1998). Clearly a statistical analysis of the research designs in this sample of studies would be

³⁹ Lack of power was indicated by a large standard error for the probability sampling groups ($n = 7$, $ES_{Zr} = .3154, SE_{Zr} = .1136$).

meaningless because nearly all studies used a form of descriptive design. The classification of research designs in this sample could therefore not be used as an indicator of study quality.

Finally, attrition is commonly cited as an important indicator of study quality. However, comparing attrition and response rates across descriptive designs seems rather meaningless because most of the samples were simply convenience samples. For example, someone approaching 16 participants using a convenience sampling method (e.g., Gioiella et al., 1998) might report a 100% response rate with zero attrition. Comparison with the response rate of 68% reported in a study by Levin, Chatters, and Taylor (1995), which was based on a nation-wide stratified random survey, would not provide any indicator of the quality of either study. In addition, many correlation matrices are based on list-wise deletion of missing values which are often not explicitly reported in relation to each correlation in the matrix.

On the whole, though study characteristics that are commonly seen as indicative of study quality were collected for the purposes of this meta-analysis, the available data did not allow for a reliable statistical examination of indicators of study quality.

Moderating effects of sample characteristics.

Other variables that may systematically affect the variance among effect sizes include common demographic characteristics of the individual samples associated with each effect size. To determine a potential moderating effect of gender on the relationship between spirituality and quality of life a regression analysis was performed on the percentage of male participants included in each of the primary effect size samples. However, data pertaining to gender was missing for 7 of the 57 effect sizes. These studies were omitted from this part of the analysis. The remaining sample of 52 revealed no significant effect of gender on the total

effect size across studies when gender was entered as a single independent predictor of the mean effect size ($p = 0.9894$).⁴⁰

The analysis of a potential moderating effect of age was constrained by the difficulty of identifying a meaningful age variable. Entering average sample age as a single independent variable in the regression equation did not significantly explain any of the variance in the relationship between spirituality and quality of life (Beta = 0.1287, $p = 0.3520$). However, data pertaining to the average age of the sample was missing for nine studies. In addition, average age is a very crude indicator of the age distribution of each of the samples. To further explore any potential effect of age, the sample of studies was divided into those studies that exclusively used samples of older adults (ages 55 and older) and those that were based on more general or younger samples. A dummy variable distinguishing samples explicitly consisting of older adults (55 years and older) from samples of general age groups was created. Nine effect sizes associated with samples of older adults were identified. Contrary to what was expected, an ANOVA analysis revealed a slightly higher mean effect size for the general age samples compared to the samples consisting exclusively of older adults ($r = 0.3460$, $p < .0001$; $r = 0.3341$, $p = 0.0012$). However, subsequent regression analysis using the variable “age group” as a single predictor of the mean effect size indicated that this difference was not statistically significant ($p = 0.9060$). In addition, it must be recognized that the resulting statistic was very confounded by the fact that the two dummy codes were not mutually exclusive because the general samples also included older adults. A

⁴⁰ Regression equations are used in the analysis of each independent variable to estimate the bivariate relationship with the mean effect size while using the estimated variance derived from the random effect model of analysis. Since only one independent variable is entered into the equation, the resulting Beta and p-value provide a direct indicator of the bivariate correlation for continuous variables.

reliable analysis of age on the relationship between spirituality and quality of life could therefore not be performed.

The problem of confounding categories also pertained to analyses of the moderating effects of ethnicity and religion. Though demographic data of ethnicity and religion were collected, few mutually exclusive ethnic or religious groups could be identified. As suggested by Lipsey and Wilson (2001), comparisons of moderating variables are only meaningful when an independent effect size associated with each mutually exclusive categorical variable could be identified. The current data set only contained one primary and one secondary effect size associated with a particular religious group (Bienenfeld, Koenig, Larson, & Sherrill, 1997; Chamberlain & Zika, 1992). Any potential differences in ethnic background were represented by four independent effect sizes pertaining to samples that were coded as exclusively Caucasian or “White” (Coward, 1998; Fry, 2001; Neill & Kahn, 1999; Thomas & Holmes, 1992), and seven independent effect sizes (five primary) in six studies involving exclusively “Black” or African American samples (Carmel, 2001; Chumbler, 1996; Coke, 1992; Levin & Taylor, 1998; Maton & Zimmerman, 1992; Thomas & Holmes, 1992). Statistical analyses comparing these two groups of studies with results based on mixed or general samples resulted in a large amount of variance within each group, as depicted in Figure 8, thereby precluding any meaningful statistical conclusion. In addition, comparison of these variables with mixed or general samples seems rather meaningless because the mixed samples include varying percentages of participants labeled Caucasian or African-American.

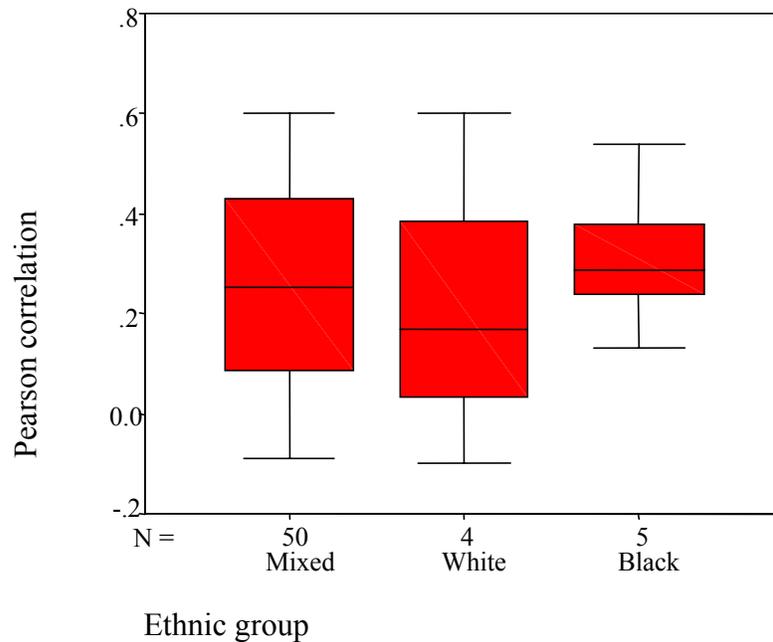


Figure 8. Comparison of Ethnic Groups

Operational definitions as moderating variables.

A final set of analyses was conducted examining various operational characteristics as potential moderating variables. It was originally hypothesized that part of the variance in the primary effect sizes might be explained by the nature of the instruments used to operationalize spirituality and quality of life. The diversity of instruments certainly warrants an examination of their potentially moderating effect. As discussed earlier, instruments were coded according to their substantive characteristics based on an analysis of the items of the instrument itself. Spirituality instruments were categorized according to: (a) those instruments that contained items that were existential as well as relational or transcendent in nature, (b) those instruments that were primarily existential in nature but still had transcendent connotations, (c) those instruments that focused primarily on the relational or

transcendent aspects of spirituality, and (d) ambiguous instruments where the items were so general in nature that it was impossible to determine how the items would be interpreted by the participants (e.g., How spiritual or religious are you?). These categories were dummy coded for the regression analyses. The results of the ANOVA analysis were indicative of a significant difference between the four groups of instruments ($Q_{between} = 8.1288$, $df = 3$, $p = .0434$). The mean effect sizes for the different groups were 0.30 ($n = 20$, $p < .0001$) for the combined existential and relational instruments, 0.50 ($n = 12$, $p < .0001$) for the existential instruments, 0.23 ($n = 15$, $p = .0032$) for the relational instruments, and 0.39 ($n = 12$, $p < .0001$) for the ambiguous instruments. These differences are conveniently displayed in an error bar plot of the 95%CI pertaining to each group (see Figure 9). These findings were confirmed when the dummy coded spirituality categories are entered into a linear regression equation. The resulting model variance was heterogeneously distributed between groups ($Q_{Model} = 8.6075$, $df = 3$, $p = .0350$) and the residual variance was no greater than expected from sampling error ($Q_{Residual} = 61.1802$, $df = 55$, $p = .2638$). On the whole, the categorization of spirituality instruments significantly accounted for some of the variance between effect sizes (R-square = .12). Consistent with the ANOVA analysis as displayed in Figure 9, the category that accounted for the larger estimates of correlation was the existential instrument.

ANOVA Spirituality Categories

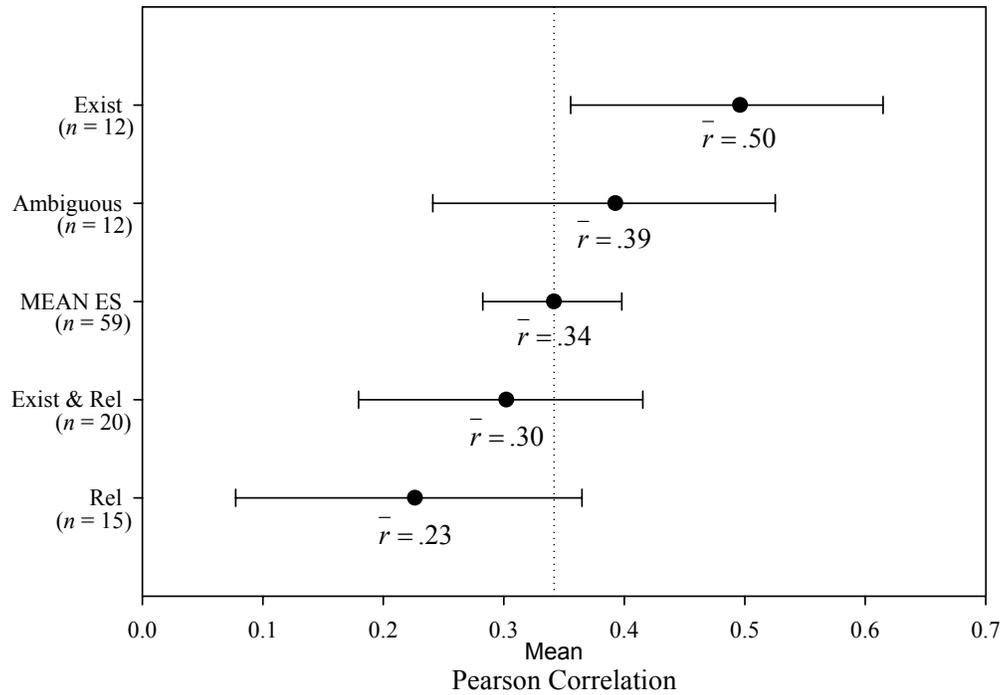


Figure 9. 95% CI Associated with Each Categorical Group of Spirituality Instruments.

Further consideration of the different operational categories of the spirituality instruments led to the hypothesis that these findings may be explained by the confounding effect of adjusting all effect sizes for instrument reliability. This hypothesis was tested with a linear regression analysis of effect sizes that were not adjusted for measurement reliability. The resulting findings revealed a similar pattern as described in the previous paragraph ($R\text{-Square} = 0.1407$, $Q_{Model} = 9.712$, $df = 3$, $p = .0212$). Again, the existential category seemed to be associated with the largest estimates of the effect size ($B = 0.1594$, $p = 0.0271$). In addition, entering instrument reliability into the model as a continuous independent variable did not result in any statistically significant effect ($Beta = -0.113$, $p = 0.5405$). It does

therefore not appear that the variance between the operational categories can be explained by differences in instrument reliability.

Similar analyses were performed to examine differences across operationalization categories of the quality of life instruments. The quality of life instruments were categorized as: (a) general multidimensional QOL instruments, (b) disease specific multidimensional QOL instruments, (c) global QOL instruments, (d) measures of subjective well-being, (e) measures of life satisfaction, and (f) a category of other or unclassifiable measures as described earlier. All these categories were again dummy coded and entered into a linear regression equation. The results revealed that the categorization of quality of life instruments explained 27% of the variances in a mixed effects model. However, the small number of primary effect sizes in several categories resulted in a large amount of variance within each of the categories which made the interpretation of the model questionable.

Based on the above findings, I decided to try and increase the accuracy of the model by collapsing the quality of life variables. I used two approaches to collapse the categories. In the first approach, I combined the general and disease specific multidimensional QOL instruments and compared them to all other instruments arguing that the multidimensional instruments were operationally distinct from the other group of instruments, which were mostly unidimensional in nature. However, the resulting regression model was non-significant with an R-square of 0.0302. An explanation can be found in an ANOVA analysis of the original QOL categories which resulted in a large difference in mean effect sizes between the general and disease specific multidimensional QOL instruments. The mean effect size for the general multidimensional QOL instruments was small and statistically insignificant ($ES_{Zr} = 0.1031, p = 0.2378$) whereas studies using the disease specific QOL

instruments resulted in statistically significant mean effect sizes of 0.4824 ($p < 0.0001$).

Collapsing these two groups into one dummy variable would obviously reduce the explanatory power of the resulting model.

I therefore decided to keep the original categories for the general and disease specific multidimensional QOL instruments and to combine all the other instruments into a single group representing overall quality of life. The resulting model significantly accounted for between studies variance ($Q = 8.3583$, $df = 2$, $p = 0.01523$) while residual variance was no larger than expected by sampling error as indicated by a non-significant p-value for residual heterogeneity ($Q = 62.2623$, $df = 56$, $p = 0.2632$). However, this model explained much less of the variance in the effect sizes ($R\text{-square} = .12$). The difference between the three groups is displayed in Figure 10, which is based on the original Pearson correlations represented in each of the groups.

The above analyses only examined the categorization of instruments in relation to spirituality or quality of life. The potential interaction between these different categories was not accounted for in these analyses. However, the sample size was too small to examine the 12 different combinations of instruments in a model based on the four spirituality categories and the 3 collapsed quality of life categories in a meaningful manner. However, the potential for interaction between the different operational categories is revealed in the analysis of a regression model involving all of the potential moderator variables.

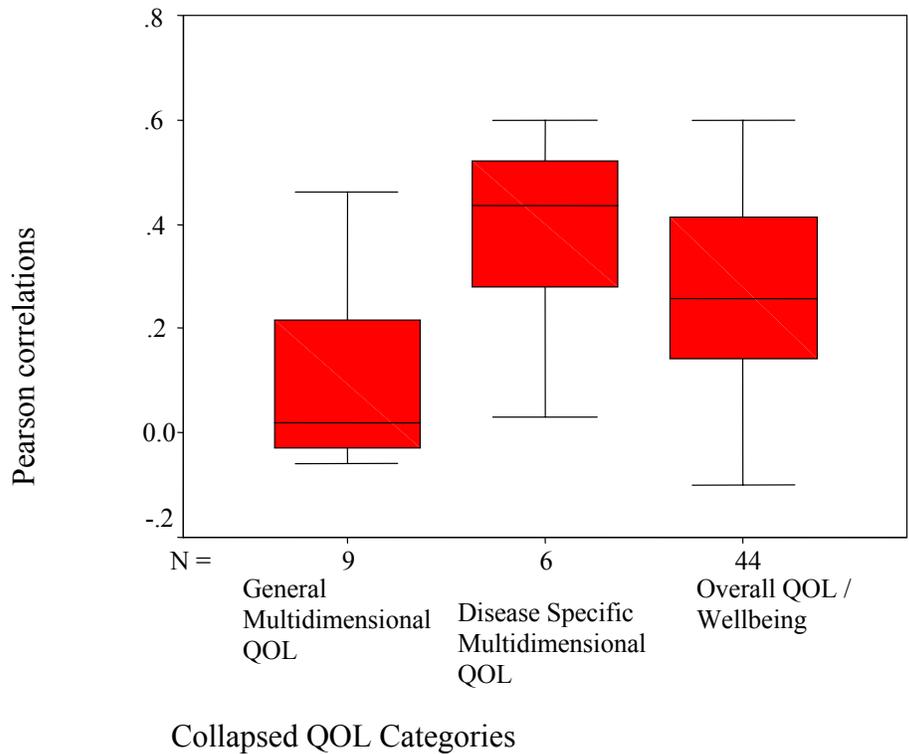


Figure 10. Box Plots of Collapsed QOL Categories.

Multivariate regression analyses

Based on the analyses thus far, several multivariate regression models were constructed involving the four categories of the spirituality instruments, the three collapsed categorical variables for the quality of life instruments, the sampling method variable comparing non-probability to probability samples, and the variable distinguishing African American samples and Caucasian samples from other ethnically diverse samples. The variable for gender was not included because of the large amount of missing data for this variable. In addition, preliminary analyses indicated that gender did not explain any of the variance pertaining to the relationship between spirituality and quality. The resulting model significantly explained 32% of the variance in the effect sizes ($Q_{Model} = 30.4568, df = 9, p =$

.0004) while the residual variance was no larger than expected by sample error as demonstrated by an insignificant residual Q-value $Q_{Residual} = 63.4813$, $df = 49$, $p = .0800$) (see Appendix G). However, the stability and reliability of this model were threatened by large variances associated with the dummy coded categorical variables for sampling method, African-American ethnicity, Caucasian ethnicity, and age group. None of the coefficients associated with these variables were significant. These findings are consistent with the preceding bivariate analyses of each of these variables discussed earlier.

A trimmed model was therefore constructed by removing the variables pertaining to sample characteristics and sampling methods. The resulting model only included the four categories of spirituality instruments and the three collapsed categories of quality of life instruments (see Appendix H). The resulting model significantly explained 27% of the variance in the primary effect sizes ($R^2 = .27$, $Q_{Model} = 23.60$, $df = 5$, $p = .0003$), and the remaining variance was not greater than expected from sampling error assuming a mixed effects model ($Q_{Residual} = 63.18$, $df = 53$, $p = .1598$). Both the spirituality instrument categories and the three collapsed quality of life instrument categories significantly contributed to the model. In particular, the general QOL, well-being and life-satisfaction instruments as well as two instruments categorized as “other” all collapsed into one category had a significant positive moderating effect in comparison with the general multidimensional quality of life instruments ($p = .0003$). In addition, the spirituality instruments that were primarily relational or transcendent in nature had a constraining effect on the mean effect size

$p = .0224$). The mean effect size for the overall model was .3562 which is comparable to the previously reported mean effect size of the random effects model (see Appendix F).⁴¹

A final regression analysis was performed to examine whether the model remained significant when applied to effect sizes that were not adjusted for instrument reliability (see Appendix I). The reliability coefficients of quality of life instruments and spirituality instruments were added to the equation as continuous independent variables. In this analysis the model remained significant and the general conclusions were similar. However, the overall mean effect size was reduced to 0.261 (comparable to the random effects analysis performed earlier), and the explanatory power of the model increased as indicated by a significant R^2 of 0.35. Despite these overall differences, the variables representing the reliability coefficients of the spirituality and quality of life instruments did not significantly contribute to the equation ($Beta_{QLREL} = .1796, p = .1360, Beta_{SPREL} = -.0363, p = .8250$). Consistent with the previous findings, the model also suggested that the spirituality instruments that were classified as being primarily relational in nature had a constraining effect on the overall effect size ($p = .0144$).

On the whole, the results from these analyses are supportive of a moderate positive relationship between spirituality and quality of life that is moderated primarily by factors pertaining to the operational definitions of spirituality and quality life. The meaning and implications of these findings are explicated in the following discussion.

⁴¹ The total variance in the mixed effects model is different from the total variance in the random effects model. This difference is explained by the different mathematical approach for calculating the random variance of the total variance. In the random effects model the random variance was estimated using a formula based the “method of moments” procedure where as the random variance in the mixed model was estimated using an iterative method of “maximum likelihood” (Lipsey & Wilson, 2001, p.140).

Discussion

The primary research question pertaining to the magnitude of the relationship between spirituality and quality of life was addressed by using various statistical approaches to estimate the mean effect size across a diverse collection of descriptive studies. The results indicate that we can be fairly confident that the bivariate mean effect size is of moderate magnitude, as defined by Cohen (1988), regardless of the statistical approach used in the original studies. The meaning of the mean effect size, however, is obscured by large amounts of variability in the results. The following discussion is oriented toward evaluating the global meaning of the effect size in relation to various factors associated with spirituality and quality of life derived from the research findings. The implications are primarily theoretical in nature and provide potentially important direction to continued theoretical development.

Magnitude of the Relationship between Spirituality and Quality of Life

A global evaluation of the magnitude of the relationship between spirituality and quality of life revealed several important findings despite the fact that the relationship between these two concepts may have been systematically affected by methodological factors that remain largely unaccounted for. First, the presence of a moderate relationship between spirituality and quality of life is fairly consistently supported. Even if the mean effect size is affected by unidentified factors, which it undoubtedly is, 44 of the 59 included studies reported a correlation larger than .10. In addition, except for the statistically insignificant mean effect size pertaining to the category of general, multi-dimensional quality of life instruments, mean effect sizes associated with each of the other operational categories pertaining to spirituality and quality of life were larger than .24 ($p < .05$) (as determined by the ANOVA analyses and illustrated in Figures 8 and 10). Furthermore, the calculation of a

fail-safe N revealed that approximately 151 studies resulting in no correlation between spirituality and quality of life ($r = 0.00$) would be needed to reduce the mean effect size to .10. Based on these findings, it is fairly reasonable to conclude that the relationship between spirituality and quality of life is, on average, of moderate magnitude despite the variance within the distribution of effect sizes.

The second important conclusion derived from the mean effect size is that the moderate magnitude of the relationship between spirituality and quality of life supports the notion that these two concepts are related yet conceptually distinct. In other words, the presence of spirituality as a unique concept is supported by the finding that its relationship with quality of life is not of such a large magnitude that would lead us to conclude that spirituality and quality of life are largely overlapping or analogous concepts. The mean effect size of 0.34 (based on the random and mixed effects models discussed earlier) suggests that the conceptual overlap between spirituality and quality of life is less than 12% ($.34^2$) as depicted in Figure 11. Even if we use the maximum value of 0.40 derived from the upper 95% confidence interval in the random effects model, we would still conclude that spirituality and quality of life remain conceptually distinct ($R^2 = 16\%$). Conceptual overlap of 50% or greater would require a minimum effect size of .70 which is highly unlikely considering the finding that only three studies (the outliers) resulted in a Pearson correlation above .70 (Coward, 1991, 1998; Tate et al., 1997). This finding supports the theoretical framework underlying this study wherein spirituality is depicted as a unique concept that stands in relationship to quality of life. The proposed conceptual framework contradicts the commonly assumed quality of life framework where spirituality is represented as an overlapping concept that falls under the multi-dimensional quality of life umbrella.

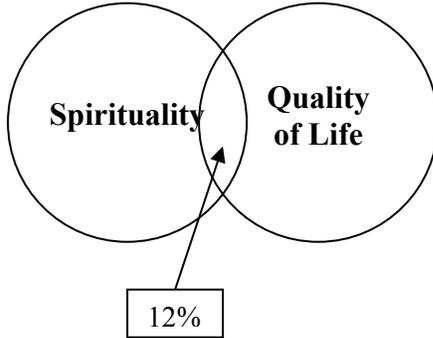


Figure 11. Venn-Diagram of Potential Conceptual Overlap

My conclusions have important implications for the postulated conceptual relationships between spirituality and quality of life as represented by multidimensional quality of life instruments. Many quality of life instruments provide an overall measure of the magnitude of one's quality of life by averaging scores across multiple dimensions that are considered to be indicators of one's quality of life. Several instruments, such as the WHOQOL-100 (WHOQOL-Group, 1998), the McGill QOL Scale (Cohen et al., 1995), and the FACT-G (Cella et al., 1993), were explicitly designed with the intent to incorporate items pertaining to spirituality under a large quality of life umbrella. The theoretical model underlying these instruments is based on the assumption that the various dimensions of quality of life co-vary in a predictable and relatively strong manner in relation to overall quality of life.

The findings of this meta-analysis do not support this theoretical model for several reasons. The first reason pertains to the moderate magnitude of the effect size as discussed in the previous paragraph. The second reason pertains to the large variances in the distribution of primary effect sizes, which indicate that the relationship between spirituality and quality of

life is moderated by other factors (methodological or substantive) which remain largely unknown. The variance can only be explained by assuming that spirituality and quality of life are to some degree independently affected by other factors. In other words, the effect of a moderating factor on one's spirituality may be different from the effect of this same factor on one's quality of life. In this case, spirituality can not be seen as a reliable indicator of quality of life unless we account for each of the moderating factors or assume that all the moderating effects are a result of random factors.

The above argument, based on the findings of this meta-analysis, is consistent with findings from other studies that refute the theoretical assumption that an overall measure of quality of life can be derived by averaging the scores or ratings of multiple dimensions putatively associated with quality of life. In accordance with Beckie and Hayduk's (1997) findings, "It seems entirely reasonable to claim that QOL remains a unidimensional concept despite the multiplicity, diversity, and complexity of its causes" (p. 35). Therefore, in terms of the present meta-analysis, the mean effect size of .37 ($ES_{Zr} = .3921, 95\% \text{ CI: } .3037 - .4805, n = 44$) associated with the global QOL, wellbeing and life-satisfaction measures may be seen as the more conceptually accurate effect size representative of the relationship between spirituality and overall quality of life.⁴² On the whole, the conceptual uniqueness of the spirituality dimension provides support for the notion that quality of life is best measured as a unique phenomenon that remains conceptually distinct from other dimensions commonly associated with, but not components or elements of, quality of life.

⁴² The global, wellbeing and life-satisfaction measures are mostly unidimensional in nature. However, this classification is somewhat confounded because a few studies included measures that may be considered multidimensional measures of well-being or life-satisfaction.

Factors Affecting the Relationship between Spirituality and Quality of Life

Once we have determined that spirituality and quality are conceptually unique phenomena, we can discuss the various factors affecting the relationship between these phenomena in a more meaningful manner. The analysis was structured according to three groups of moderating factors. The first group pertained to those methodological factors associated with overall research design and study quality. It was hypothesized that these factors would affect the findings based on the assumption that more rigorously developed research designs would result in a more accurate or precise estimate. The analyses however led to inconclusive findings. Sampling method was the only design characteristic statistically analyzed that resulted in different mean effects; studies with probability samples had slightly lower effect sizes than non-probability samples ($\bar{r}_{prob.} = .31, p < .0055, \bar{r}_{non-prob.} = .35, p < .0001$). However, the homogeneity analysis revealed that the between group homogeneity assumption was not met ($Q_{Between} = .1686, df = 1, p = .6813$), which indicates that there were other factors that explained the difference in effect sizes between these two groups. Not surprisingly, sampling method did not explain any of the variance found in the observed associations between spirituality and quality of life ($R^2 = .0028$) when entered as a single independent binary variable in a linear regression equation.

Similar conclusions were drawn for the sample characteristics pertaining to gender, age and ethnicity. The mean effect size remained remarkably consistent across sample differences in gender, age and ethnicity. However, the findings were inclusive because of the inability to create independent (mutually exclusive) samples of each categorical variable and because a lack of statistical power resulted in large standard errors associated with the relevant coefficients. The only statistically significant explanation for the variance among the

primary effect sizes pertained to differences in how the concepts of spirituality and quality of life were operationalized which accounted for 27% of the variance (see Appendix H). This finding has several implications for ongoing empirical research and theoretical development pertaining to the conceptualization of, and measurement of, spirituality in relation to quality of life.

Conceptualization and measurement of spirituality.

As discussed earlier, the debate on how to conceptualize and measure spirituality has been ongoing for several decades and no conclusive empirical answer to this question has been found to date (Larson et al., 1998; Sloan, Bagiella, & Powell, 1999). For the purposes of this meta-analysis, spirituality was conceptualized in accordance with the definition posed by Larson et al. Based on this definition, I concluded that explicit measures of spirituality must reflect existential as well as relational or transcendent attributes of spirituality. These existential and relational attributes of spirituality pertain to several instruments that contain explicit items pertaining to each of these characteristics (see Appendix E). A factor analysis of one of these instruments, the Spiritual Well-Being Scale (Paloutzian & Ellison, 1982), resulted in a classification of items according to two factors representing existential and relational or transcendent dimensions.⁴³ Based on this background, each spiritual instrument was classified according to four categories that reflect the degree to which the instrument represents the proposed conceptualization of spirituality. The multivariate meta-analysis involving these four categories resulted in several pertinent conclusions pertaining to the conceptualization and measurement of spirituality.

⁴³ Though these factors were termed Existential and Religious by the original authors of the scale, the religious factor may be more appropriately termed Relational or Transcendent as all the items reflect characteristics of a relationship with a divine being or God.

One conclusion involves the overall amount of variance that may be attributed to differences in the operational definition of spirituality. The regression and ANOVA analyses involving only the four categories of the spirituality instruments revealed that the variance in the effect sizes is partly explained by the nature of the spirituality instruments used ($R^2 = .1407$). These differences are depicted in Figure 9, which represents the mean effect sizes and confidence intervals of each category as derived from the ANOVA analysis. From this figure it becomes clear that the mean effect size can vary anywhere between .23 and .50 depending on whether the instruments were primarily relational or existential in nature. These findings indicate that the manner in which spirituality is operationalized may be an important consideration in the interpretation of primary studies pertaining to the relationship between spirituality and quality of life.

Several explanations for the variance of effect sizes across different operational characteristics of spirituality can be identified. First, the variance may be explained by the notion that the instruments that are primarily relational or existential in nature actually measure two distinct phenomena. This notion is supported by the finding that the mean effect sizes of the primarily existential instruments and the primarily relational instruments are fairly widely dispersed and fall outside of the confidence interval of the overall mean effect size (see Figure 9). However, this conclusion is premature without first considering other potential differences between the two groups of studies. In addition, it must be remembered that both of these categories are conceptually overlapping because the inclusion criteria allowed only for the inclusion of instruments that had existential as well as relational or transcendent connotations or instruments that were considered to be ambiguous in nature. It

seems therefore unlikely that such a wide spread would be entirely explained by conceptual differences when the two operational categories are to some extent overlapping.

The other explanation for the spread of mean effect sizes across the operational categories of spirituality relates to the psychometric properties of the instruments themselves. As discussed earlier, the analyses were based on effect sizes that were adjusted for instrument reliability. It may therefore be that the relational instruments were less reliable than the existential instruments which would cause these two categories to disperse in the analysis. However, a comparison of the average reliability correction factor associated with the existential and relational instruments revealed that studies in both categories received approximately the same average adjustment for instrument reliability. The average reliability adjustment factor for effect sizes resulting from the use of primarily existential instruments was 0.923 and the reliability adjustment factor for the relational instruments was 0.922.⁴⁴ The reliability coefficients associated with the two groups of instruments, therefore, do not explain any of the variance in the primary effects.⁴⁵

A more plausible explanation of the differences in effect sizes across different operationalizations of spirituality relates to the possibility that moderating effects of other factors were in action. One of these factors may be the potential differences in the type of quality of life instruments used in each of the categories. In other words, there may be an interaction between the operational categories of spirituality and the operational categories of quality of life. Unfortunately, to my knowledge, meta-analytic approaches involving post-hoc

⁴⁴ The average adjustment factor was calculated by averaging the square roots of the reliability of the spirituality instrument associated with each effect size in the group (see Appendix D for reliability coefficients associated with each effect size).

⁴⁵ In addition, adding the reliability coefficients of spirituality instruments to the regression model involving effect sizes that were not adjusted for instrument reliability did not result in a large increase in the R-square value, which provides additional support to the notion that differences in the primary effects could not be explained by differences in the reliability of the two groups of spirituality instruments.

analyses of this nature have not yet been constructed, and there was insufficient statistical power, given the number of primary effect sizes, to introduce conventional interaction terms into the linear regression model. However, a basic analysis comparing the type of quality of life instruments used in the existential group as compared to the transcendent group revealed that the quality of life instruments in the transcendent group exclusively measured life satisfaction whereas the quality of life instruments in the existential group represented most of the quality of life categories as depicted in Figure 12. Not surprisingly, an ANOVA analysis of all the quality of life categories revealed that the mean effect size for the satisfaction with life instruments was slightly lower than the overall mean effect size ($\bar{r} = .31, n = 36, p < .0001$). The differences between effect sizes associated with different spirituality categories may therefore be partly explained by an interaction associated with the particular combination of instruments used. However, the nature and magnitude of the potential interaction between quality of life and spirituality instruments is unknown and worthy of further consideration.

Obviously, there are many other factors that may explain the differences in effect sizes among the types of operationalization of spirituality. In their critique of the current state of empirical literature on spirituality and health, Sloan et al. (1999) emphatically discuss the need for controlling demographic factors, social factors and behavioural factors in examining any effect of spirituality on health-related constructs; they argue that “failure to control for these factors can lead to a biased estimation of this association” (p. 665). Unfortunately, the relatively small number of studies that were included in this meta-analysis preclude the ability to analyze the potential interactions between spirituality categories and other variables associated with sample characteristics and study characteristics. Ongoing primary research

initiatives examining other potential moderating variables pertaining to the relationship between spirituality and quality of life are necessary to further explore the moderating effects of such factors.

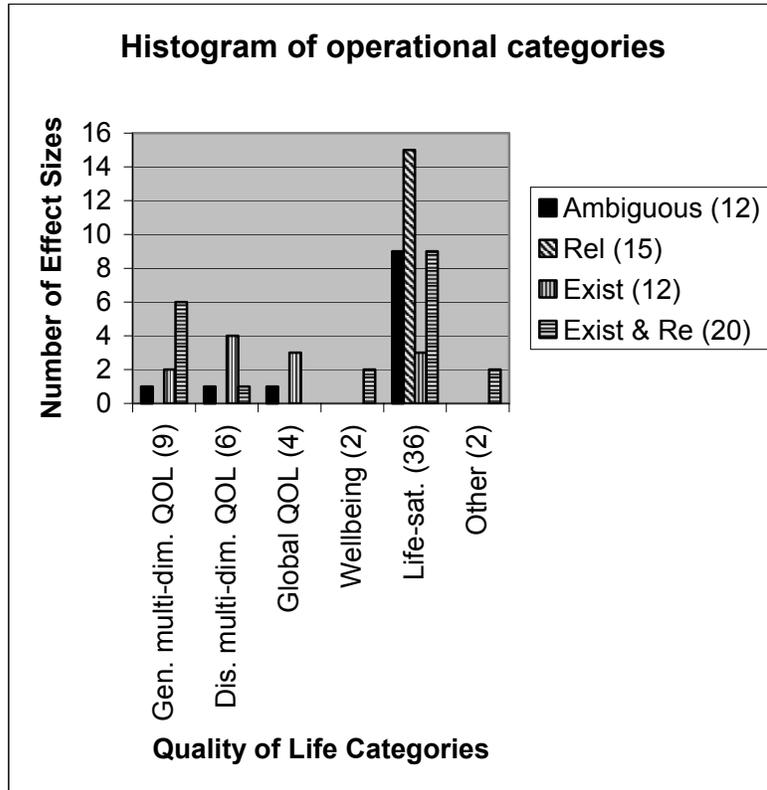


Figure 12. Combination of Operational Categories Represented by each Effect Size.

We can conclude that though part of the variance in effect sizes between each of the operational categories of spirituality might be explained by conceptual differences inherent in each of these categories, there is reason to believe that much of the variance may be attributed to interactions between the spirituality categories and other moderating variables such as the operational categories of quality of life. Whether the current meta-analysis supports the proposed conceptualization of spirituality can not be determined with any degree of certainty. Ongoing research examining the potential effects of different types of

operationalizations of spirituality in relation to quality of life and other outcomes is needed to test the definition of spirituality provided by Larson et al. (1998). The covariance between the relational and existential characteristics of spirituality is of particular interest.

Another recommendation pertaining to the measurement of spirituality remains to be made. The finding that differences in operational definitions significantly explain part of the variance in effect sizes is concerning considering the wide diversity of instruments in use. Most spirituality instruments were only used in one or two studies. The few instruments that were used in three or more studies included the FACIT-SP (Cella et al., 1997), the spiritual subscale of the WHOQOL-100 (WHOQOL-Group, 1998), the Self-Transcendence Scale (Reed, 1991) and the Spiritual Well-Being Scale by Ellison and Paloutzian (1983). Further evidence of the multiplicity of instruments measuring spiritually-related phenomena is evident in the extensive list of 120 instruments reviewed by Hill and Hood (1999). A large number of additional instruments were cited and reviewed by MacDonald et al. (MacDonald, Kuentzel, et al., 1999; MacDonald, Friedman, et al., 1999; MacDonald et al., 1995) and Larson et al. (1998). This diversity and inconsistent use of instruments constrain the synthesis of study results as was demonstrated by the findings of this meta-analysis. It may be that other factors pertaining to the relationship between spirituality and quality of life would have significantly explained part of the variance in effect sizes if the variances across different operational definitions were more rigorously controlled. The importance of selecting instruments with a history of psychometric testing and empirical application cannot be underestimated if we want to draw conclusions by synthesizing findings from different studies.

Operational characteristics of quality of life measurement.

The same concerns pertaining to the measurement of spirituality also apply to the moderating effect of different operational definitions of quality of life. One important implication pertaining to the conceptualization and measurement of quality of life has already been identified. However, a few additional remarkable findings need to be discussed in further detail. One of these findings involves a comparison of the general, multi-dimensional quality of life instruments with the disease-specific quality of life instruments. Contrary to what might be expected, the effect sizes associated with each of these categories are very different (see Figure 13). However, a comparison of sample and research design characteristics between these two categories did not reveal any significant differences. Three of the four operationalizations of spirituality were represented in both groups of studies, the cumulative sample sizes were very comparable and there was also no large difference between the average reliability adjustments related to the effect sizes in each group. Again, several explanations may account for the different findings pertaining to general, multi-dimensional instruments and disease-specific multidimensional instruments. One explanation is that the difference may be attributed to a third moderating factor for which no data were collected. A second explanation is that the instruments pertaining to each of these categories actually measure different phenomenon (i.e., the instruments are not conceptually congruent). A third explanation pertains to the variability within each group. The ANOVA results, reported in Figure 13, reveal that the confidence intervals associated with each group are not mutually exclusive. The possibility exists that the distribution of effect sizes for each group is actually the same and that the mean difference found in the meta-analysis is largely due to sampling error associated with small sample sizes.

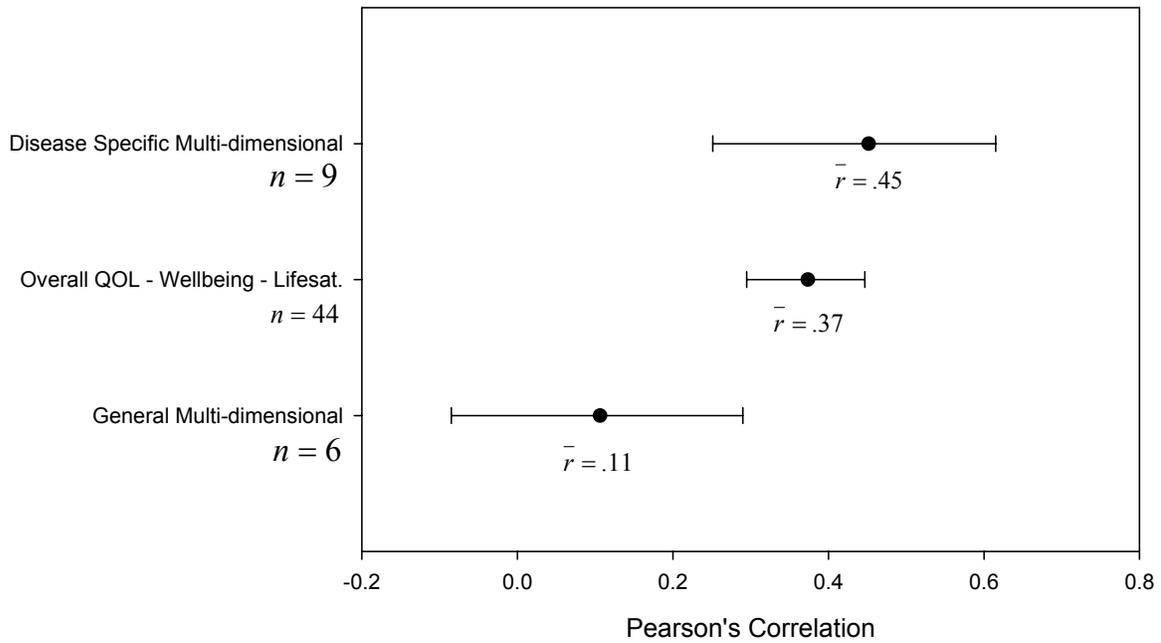


Figure 13. 95% CI Associated with Each Categorical Group of QOL Instruments as Revealed by the ANOVA Analysis.

Unfortunately, the relatively small number of studies in each group precludes the ability to conclude which explanation might best explain the findings. However, the results do present an important consideration for the interpretation of primary research findings. The large amount of variability across the different operationalizations of quality of life suggest that primary research findings involving these instruments in examining a relationship between spirituality and quality of life should be cautiously interpreted and compared to other studies before any reliable conclusions can be drawn.

Limitations

The results of this meta-analysis can only be interpreted correctly when potential limitations associated with the meta-analysis process have been taken into consideration.

Many critiques of meta-analyses have been identified since the publication of foundational works by Glass, by Schmidt and Hunter, and by Rosenthal and Rubin in the 1970's (as cited in Lipsey & Wilson, 2001). Common critiques associated with meta-analytic methods include biases in the selection of studies, publication bias, comparing apples and oranges, and not accounting for the influence of extraneous variables (Hunter & Schmidt, 1990; Lipsey & Wilson, 2001; Wolf, 1986). However, before discussing the limitations of this particular meta-analysis, it must be said that the limitations of meta-analyses in general mostly reflect the limitations associated with the primary research studies and the overall body of empirical knowledge associated with the topic of interest. The limitations themselves therefore have important implications for ongoing research on spirituality and quality of life.

Bias in the selection process.

One criticism relates to the issue of sampling bias. Clearly, the results of the meta-analysis are largely determined by the selection of studies included in the analysis. The primary method employed to address this potential bias was the establishment of explicit inclusion and exclusion criteria to guide the selection process. It is nevertheless conceivable that people may interpret the inclusion criteria in different ways. Though the rigor of the selection process may have been strengthened by calculating an inter-rater reliability estimate for all the studies reviewed and selected by the two independent researchers, the resources available for this meta-analysis did not allow for this type of analysis. In addition, the selection process was fairly iterative in nature and decisions about dubious studies were reached through consensus discussions between the two raters. An inter-rater reliability assessment would therefore be rather meaningless. The selection process has been described

in great detail in the methods section of this report to allow the reader to draw his or her own conclusions regarding the representativeness of the final sample of selected studies.

Another bias pertaining to the selection process results from the decision to only include published reports of primary studies. It has commonly been argued that studies with favourable results are more readily published than studies with statistically insignificant results (Lipsey & Wilson, 2001; Wolf, 1986). This limitation was addressed by calculating a failsafe N as described in the results section. The results indicated that 151 studies with an effect size of zero would be needed to reduce the mean effect size to 0.10. Based on the current sample of studies, the likelihood of finding this number of additional studies seems rather remote.⁴⁶

The apples and oranges phenomenon.

The criticism of comparing apples with oranges pertains to the argument that “logical conclusions cannot be drawn by comparing and aggregating studies that include different measuring techniques, definitions of variables (e.g., treatments, outcomes), and subjects because they are too dissimilar” (Wolf, 1986, p. 14). This criticism is certainly of concern in the present meta-analysis. In response to this concern, attempts were made to account for some of the variance by statistically controlling for differences in research design. The results of these analyses were however inconclusive.

A more pertinent concern is the wide range of instruments that have been used to measure spirituality and quality of life. However, as discussed by Hunter and Schmidt (1990), the nature of this concern depends on the nature of the research question guiding the

⁴⁶ Continuation of this research initiative involves the coding of approximately 30 dissertations pertaining to spirituality and quality of life that were completed in 2000 and 2001. It is hypothesized that the inclusion of these unpublished reports will not dramatically alter the mean effect size, though it may add additional power to the regression analyses of moderating factors thereby strengthening the overall research design of this meta-analysis.

analysis. The research question underlying this meta-analysis was so general in nature that systematic exclusion of certain operational definitions would likely lead to biased results. The approach of assessing the moderating effects of differences in operational definitions was therefore used to avoid excluding potentially important material. Nevertheless, the results did indicate that differences in the effect sizes may be partly explained by differences in the operational definitions. However, the overall differences were not large enough to dramatically alter the mean effect size.

Lack of causal evidence.

Another important limitation relates to the inability to use the results to draw conclusions about the causal nature of the relationship between spirituality and quality of life. This limitation clearly reflects the current state of empirical knowledge and has already been discussed. The concerns are two-fold: (a) lack of control for confounding or moderating factors in primary research designs and (b) the nature of correlational analyses that do not allow for causal inferences or inferences pertaining to the direction of the relationship. The first point has already been discussed in great detail and has been succinctly summarized by Thoresen (1999) in his review of research on spirituality and health; he wrote, "Prospective studies have not accounted for many psychosocial and sociocultural factors that could explain health outcomes" (p. 294). This limitation is therefore primarily a reflection of the available empirical literature.

The second point is one that clearly represents the methodological nature of the current body of empirical knowledge. Though the original selection criteria stated that only descriptive studies would be included in this analysis, it must also be noted that the extensive search methods failed to identify any studies that were experimental in nature. This is largely

explained by the nature of the variables involved; variables such as quality of life and spirituality are simply very difficult, if not impossible, to manipulate and control.

Nevertheless, the importance of statistical and methodological control of other factors which can be incorporated into primary research designs can not be underestimated. The implications of this limitation for the interpretation of the findings derived from this meta-analysis are therefore similar to the limitations associated with the descriptive studies that were included. The need for more rigorously designed research initiatives has been clearly exposed.

The mechanism by which spirituality may affect quality of life.

Another limitation involves the inability to describe the mechanism by which spirituality relates to quality of life. Though attempts have been made to collect data pertaining to potential mediating factors, the sample of 59 studies was too small to examine the variables that might explain how spirituality affects quality of life. Answering this question was further constrained by the inconsistent reporting of variables of interest in the primary studies. For example, Thoresen (1999) suggests that the relationship between spirituality and health-related outcomes may be mediated by differences in “health habits, social support, psychodynamic or other cognitive behavioral effects (e.g., psycho-neuroimmunology), or supernatural or ‘superempirical’ effects (e.g., distant healing)” (p. 296). However, few studies provided sufficient statistical information about the variables associated with any of these mechanisms. Without this type of information, aggregation of research findings will always remain at the descriptive level.

Generalizability of the results.

A final concern pertains to the generalizability of the results. As a rule, results derived from a meta-analysis are more generalizable than any one study because the resulting effect size represents a greater variety of primary sample characteristics than can be achieved through a single primary study. However, as identified in the presentation of the pooled demographic data, the current meta-analysis remains constrained by the under-representation of multiple cultural and religious groups and people of different nationalities. The results are therefore primarily generalizable to Caucasian or African-American people living in North-America or Western Europe.

On the whole, though they need to be interpreted with caution, the research findings of this meta-analysis provide important direction for further research and theoretical development. Recommendations for further research include the need for more rigorously controlled research designs, the need to incorporate instruments that measure potentially important moderating variables, as well as the need to achieve consensus pertaining to the conceptualization of spirituality and quality of life. A final recommendation pertains to the need to replicate current findings. Many common critiques of the meta-analysis process could be addressed by replicating meta-analyses and comparing potential differences in results.

Conclusion

In conclusion, results of this meta-analysis have revealed various findings pertaining to the conceptualization of spirituality as a distinctive concept that relates to people's quality of life. The implications are mostly theoretical in nature and raise questions about the commonly assumed multidimensional conceptualization of quality of life. The findings

demonstrate the value of meta-analytical methods in revealing general characteristics of conceptual relationships that are easily obscured by isolated primary research findings. On the other hand, other questions pertaining to the nature and direction of the relationship between spirituality and quality remain largely unanswered and indicate the need for continued primary research. Ongoing qualitative meta-analyses are needed to further explicate the meaning of the findings of this meta-analysis pertaining to the relationship between spirituality and quality.

Various concerns pertaining to the overall body of research on spirituality have been identified. Of primary concern is the wide diversity of instruments used to measure spirituality. However, constraining the measurement of spirituality to narrowly defined theoretical or philosophical notions contradicts the subjective nature of this concept as understood by people from different cultural or religious backgrounds. A conceptualization of spirituality that addresses all the peculiarities associated with every culture and religion is not likely to be developed. A more appropriate approach is to identify the common attributes of diverse conceptualizations of spirituality. For example, the broadly defined existential and transcendent or relational attributes of spirituality provided useful direction to the selection and classification of instruments for the purposes of this meta-analysis. Further research is needed to validate whether these attributes can indeed be combined to provide a conceptualization of spirituality that is broadly defined yet remains conceptually distinct from other concepts such as the dimensions commonly associated with quality of life.

References

- Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Duez, N. J., Filbierty, A., et al. (1993). European Organization for Research and Treatment of Cancer QLQ-C30: A quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute*, 85, 365-376.
- Abeles, R., Ellison, C., George, L. K., Idler, E., Krause, N., Levin, J., et al. D. (1999). *Multidimensional measurement of religiousness/spirituality for use in health research*. Kalamazoo, MI: Fetzer Institute & National Institute on Aging Working Group.
- Aldridge, D. (1993). Is there evidence for spiritual healing? *Advances, The Journal of Mind-Body Health*, 9(4), 4-21.
- Allport, G. W., & Ross, J. M. (1967). Personal religious orientation and prejudice. *Journal of Personality and Social Psychology*, 5, 432-443.
- Aycock, T. J. (1999). Attitude Toward Christianity Scale. In P. C. Hill & R. W. Hood (Eds.), *Measures of religiosity* (pp. 80-82). Birmingham, AL: Religious Education Press.
- Ayele, H., Mulligan, T., Gheorghiu, S., & Reyes-Ortiz, C. (1999). Religious activity improves life satisfaction for some physicians and older patients. *Journal of the American Geriatrics Society*, 47, 453-455.
- Baetz, M., Larson, D. B., Marcoux, G., Bowen, R., & Griffin, R. (2002). Canadian psychiatric inpatient religious commitment: An association with mental health. *Canadian Journal of Psychiatry-Revue Canadienne De Psychiatrie*, 47, 159-166.
- Bassett, R. L. (1999). Intrinsic Religious Motivation Scale. In P. C. Hill & R. W. Hood (Eds.), *Measures of religiosity* (pp. 135-137). Birmingham, AL: Religious Education Press.
- Beckie, T. M., & Hayduk, L. A. (1997). Measuring quality of life. *Social Indicators Research*, 42, 21-37.
- Beery, T. A., Baas, L. S., Fowler, C., & Allen, G. (2002). Spirituality in persons with heart failure. *Journal of Holistic Nursing*, 20(1), 5-25.
- Benight, C. C., Flores, J., & Tashiro, T. Y. (2001). Bereavement coping self-efficacy in cancer widows. *Death Studies*, 25, 97-125.
- Bienenfeld, D., Koenig, H. G., Larson, D. B., & Sherrill, K. A. (1997). Psychosocial predictors of mental health in a population of elderly women: Test of an explanatory model. *American Journal of Geriatric Psychiatry*, 5(1), 43-53.

- Boivin, M. J., Kirby, A. L., Underwood, L. K., & Silva, H. (1999). Spiritual Well-Being Scale. In P. C. Hill & R. W. Hood (Eds.), *Measures of religiosity* (pp. 382-385). Birmingham, AL: Religious Education Press.
- Blaine, B., & Crocker, J. (1995). Religiousness, race, and psychological well-being: Exploring social psychological mediators. *Personality and Social Psychology Bulletin, 21*, 1031-1041.
- Bonomi, A. E., Patrick, D. L., Bushnell, D. M., & Martin, M. (2000). Validation of the United States' version of the World Health Organization Quality of Life (WHOQOL) instrument. *Journal of Clinical Epidemiology, 53*(1), 1-12.
- Bowling, A. (1997). *Measuring health: A review of quality of life measurement scales*. Buckingham, UK: Open University Press.
- Bowman, P. J., & Sanders, R. (1998). Unmarried African American fathers: A comparative life span analysis. *Journal of Comparative Family Studies, 29*(1), 39-56.
- Bradburn, N. M. (1969). *The structure of psychological well-being*. Chicago, IL: Aldine.
- Brady, M. J., Peterman, A. H., Fitchett, G., Mo, M., & Cella, D. (1999). A case for including spirituality in quality of life measurement in oncology. *Psycho-Oncology, 8*, 417-428.
- Bufford, R. K., Paloutzian, R. F., & Ellison, C. W. (1991). Norms for the Spiritual Well-Being Scale. *Journal of Psychology and Theology, 19*(1), 56-70.
- Burkhardt, M. A. (1989). Spirituality: An analysis of the concept. *Holistic Nursing Practice, 3*(3), 69-77.
- Burkhardt, M. A. (1993). Characteristics of spirituality in the lives of women in a rural Appalachian community. *Journal of Transcultural Nursing, 4*(2), 12-18.
- Burriss, C. T. (1999). Religious Orientation Scale. In P. C. Hill & R. W. Hood (Eds.), *Measures of religiosity* (pp. 144-153). Birmingham, AL: Religious Education Press.
- Byrd, K. R., Lear, D., & Schwenka, S. (2000). Mysticism as a predictor of subjective well-being. *International Journal for the Psychology of Religion, 10*, 259-269.
- Campbell, A., Converse, P., & Rodgers, W. L. (1976). *The quality of American life*. New York: Sage.
- Carmel, S. (2001). The will to live: Gender differences among elderly persons. *Social Science & Medicine, 52*, 949-958.
- Cella, D., Tulskey, D. S., Gray, G., Sarafian, B., Linn, E., & Bonami, P. (1993). The Functional Assessment of Cancer Therapy Scale: Development and validation of the general measure. *Journal of Clinical Oncology, 11*, 570-579.

- Cella, D. F. (1997). *Manual of the Functional Assessment of Chronic Illness Therapy (FACIT Scales) - Version 4*. Evanston, IL: Center on Outcomes, Research & Education (CORE), Evanston Northwestern Healthcare and Northwestern University.
- Chamberlain, K., & Zika, S. (1992). Religiosity, meaning in life, and psychological well-being. In J. F. Schumaker (Ed.), *Religion and mental health* (pp. 138-148). New York, NY: Oxford University Press.
- Chiu, L. (2000). Lived experience of spirituality in Taiwanese women with breast cancer... including commentary by Clark MB and Daroszewski EB with author response. *Western Journal of Nursing Research*, 22(1), 29-53.
- Chumblor, N. R. (1996). An empirical test of a theory of factors affecting life satisfaction: Understanding the role of religious experience. *Journal of Psychology and Theology*, 24, 220-232.
- Clarke, M., & Oxman, A. D. (2000). Cochrane reviewers' handbook [updated June 2000]. In: Review Manger [Computer program] (Version 4.1). Oxford, England: The Cochrane Collaboration.
- Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, S. R., Bultz, B. D., Clarke, J., Kuhl, D. R., Poulson, M. J., Baldwin, M. K. et al. (1997). Well-being at the end of life: Part 1. A research agenda for psychosocial and spiritual aspects of care from the patient's perspective. *Cancer Prevention & Control*, 1, 334-342.
- Cohen, S. R., Hassan, S. A., Lapointe, B. J., & Mount, B. M. (1996). Quality of life in HIV disease as measured by the McGill Quality of Life Questionnaire. *AIDS*, 10, 1421-1427.
- Cohen, S. R., Mount, B. M., Strobel, M. G., & Bui, F. (1995). The McGill Quality of Life Questionnaire: A measure of quality of life appropriate for people with advanced disease. A preliminary study of validity and acceptability. *Palliative Medicine*, 9, 207-219.
- Cohen, S. R., Mount, B. M., Tomas, J. J. N., & Mount, L. F. (1996). Existential well-being is an important determinant of quality of life - Evidence from the McGill Quality of Life Questionnaire. *Cancer*, 77, 576-586.
- Coke, M. M. (1992). Correlates of life satisfaction among elderly African Americans. *Journals of Gerontology*, 47, 316-320.
- Conner-Spady, B., Cumming, C., Nabholtz, J. M., Jacobs, P., & Stewart, D. (2001). Responsiveness of the EuroQol in breast cancer patients undergoing high dose chemotherapy. *Quality of Life Research*, 10, 479-486.

- Cotton, S. P., Levine, E. G., Fitzpatrick, C. M., Dold, K. H., & Targ, E. (1999). Exploring the relationships among spiritual well-being, quality of life, and psychological adjustment in women with breast cancer. *Psycho-Oncology*, 8, 429-438.
- Coward, D. D. (1990). Correlates of self-transcendence in women with advanced breast cancer (Doctoral dissertation, University of Arizona, 1990). *Dissertation Abstracts International*, 52, 158.
- Coward, D. D. (1991). Self-transcendence and emotional well-being in women with advanced breast cancer. *Oncology Nursing Forum*, 18, 857-863.
- Coward, D. D. (1996). Self-transcendence and correlates in a healthy population. *Nursing Research*, 45, 116-121.
- Coward, D. D. (1998). Facilitation of self-transcendence in a breast cancer support group. *Oncology Nursing Forum*, 25, 75-84.
- Derogatis, L. R. (1986). The psychosocial adjustment to illness scale (PAIS). *Journal of Psychosomatic Research*, 30, 77-91.
- Derogatis psychological tests. (n.d.). Retrieved July 18, 2002, from <http://www.derogatis-tests.com/info.html>.
- Diener, E., Emmons, R., Larson, R., & Griffin, S. (1985). The Satisfaction with Life Scale. *Journal of Personality Assessment*, 49, 71-75.
- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, 125, 276-302.
- Dorahy, M. J., Lewis, C. A., Schumaker, J. F., Akuamoah Boateng, R., Duze, M. C., & Sibiya, T. E. (1998). A cross-cultural analysis of religion and life satisfaction. *Mental Health, Religion and Culture*, 1(1), 37-43.
- Dyson, J., Cobb, M., & Forman, D. (1997). The meaning of spirituality: A literature review. *Journal of Advanced Nursing*, 26, 1183-1188.
- Elkins, D. N., Hedstrom, L. J., Hugher, L. L., Leaf, J. A., & Saunders, C. (1988). Toward a humanistic-phenomenological spirituality. *Journal of Humanistic Psychology*, 28(4), 5-18.
- Ellison, C. W. (1983). Spiritual well-being: Conceptualization and measurement. *Journal of Psychology and Theology*, 11, 330-340.
- Emblen, J. D. (1992). Religion and spirituality defined according to current use in nursing literature. *Journal of Professional Nursing*, 8(1), 41-47.

- Fabricatore, A. N., Handal, P. J., & Fenzel, L. M. (2000). Personal spirituality as a moderator of the relationship between stressors and subjective well-being. *Journal of Psychology and Theology, 28*, 221-228.
- Fenzel, L. M. (1996). *Spiritual Involvement Scale*. Unpublished manuscript, Loyola College in Maryland at Baltimore, MD.
- Ferrans, C. E., & Powers, M. J. (1992). Psychometric assessment of the Quality-of-Life Index. *Research in Nursing & Health, 15*, 29-38.
- Ferrell, B. R., Grant, M. M., Funk, B. M., Otis-Green, S. A., & Garcia, N. J. (1998). Quality of life in breast cancer survivors: Implications for developing support services. *Oncology Nursing Forum, 25*, 887-895.
- Ferrell, B. R., Hassey Dow, K., & Grant, M. (1995). Measurement of the quality of life in cancer survivors. *Quality of Life Research, 4*, 523-531.
- Fitchett, G., Rybarczyk, B. D., DeMarco, G. A., & Nicholas, J. J. (1999). The role of religion in medical rehabilitation outcomes: A longitudinal study. *Rehabilitation Psychology, 44*, 333-353.
- Francis, L. J., & Stubbs, M. T. (1987). Measuring attitudes toward Christianity: From childhood to adulthood. *Personality and Individual Differences, 8*, 741-743.
- Fry, P. S. (2001). Predictors of health-related quality of life perspectives, self-esteem, and life satisfactions of older adults following spousal loss: An 18-month follow-up study of widows and widowers. *Gerontologist, 41*, 787-798.
- Fugl-Meyer, A. R., Branholm, L. B., & Fugl-Meyer, K. S. (1991). Happiness and domain-specific life satisfaction in adult northern Swedes. *Clinical Rehabilitation, 5*, 25-33.
- Genia, V. (1991). The Spiritual Experience Index: A measure of spiritual maturity. *Journal of Religion and Health, 30*, 337-347.
- Genia, V., & Cooke, B. A. (1998). Women at midlife: Spiritual maturity and life satisfaction. *Journal of Religion & Health, 37*, 115-123.
- George, L. K., Larson, D. B., Koenig, H. G., & McCullough, M. E. (2000). Spirituality and health: What we know, what we need to know. *Journal of Social and Clinical Psychology, 19*(1), 102-116.
- Gioiella, M. E., Berkman, B., & Robinson, M. (1998). Spirituality and quality of life in gynecologic oncology patients. *Cancer Practice, 6*, 333-338.
- Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills, CA: Sage.

- Green, L. W., & Kreuter, M. W. (1999). *Health promotion planning: An educational and ecological approach*. Mountainview, CA: Mayfield.
- Hall, J. A., Tickle-Degnen, L., Rosenthal, R., & Mosteller, F. (1994). Hypotheses and problems in research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 17-28). New York: Russell Sage Foundation.
- Harvey, C. D., Bond, J. B., & Greenwood, L. J. (1991). Satisfaction, happiness, and self-esteem of older rural parents. *Canadian Journal of Community Mental Health, 10*(2), 31-46.
- Hedges, L. V. (1994). Statistical considerations. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 29-38). New York: Russell Sage Foundation.
- Highfield, M. (1992). Spiritual health of oncology patients. *Cancer Nursing, 15*, 1-8.
- Hill, P. C., & Hood, R. W. (1999). *Measures of religiosity*. Birmingham, AL: Religious Education Press.
- Hill, P. C., Pargament, K. I., Hood, R. W., McCullough, M. E., Swyers, J. P., Larson, D. B. et al. (2000). Conceptualizing religion and spirituality: Points of commonality, points of departure. *Journal for the Theory of Social Behaviour, 30*(1), 51-77.
- Hoge, D. R. (1972). A validated Intrinsic Religious Motivation Scale. *Journal for the Scientific Study of Religion, 11*, 369-376.
- Holland, J. C., Kash, K. M., Passik, S., Gronert, M. K., Sison, A., Lederberg, M., et al. (1998). A brief spiritual beliefs inventory for use in quality of life research in life-threatening illness. *Psycho-Oncology, 7*, 460-469.
- Holland, J. C., Passik, S., Kash, K. M., Russak, S. M., Gronert, M. K., Sison, A., et al. (1999). The role of religious and spiritual beliefs in coping with malignant melanoma. *Psycho-Oncology, 8*, 14-26.
- Hong, S. M., & Giannakopoulos, E. (1994). The relationship of satisfaction with life to personality-characteristics. *Journal of Psychology, 128*, 547-558.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York: Russell Sage Foundation.
- Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
- Idler, E. L. (1987). Religious involvement and the health of the elderly: Some hypotheses and an initial test. *Social Forces, 66*, 226-238.
- Jackson, J. S. (1991). Methodological approach. In J. S. Jackson (Ed.), *Life in Black America*. Thousand Oaks, CA: Sage.

- Kash, K. M., Holland, J. C., Passik, S. D., Lederberg, M. S., Sison, A. C., & Gronert, M. K. (1995). The Systems of Belief Inventory (SBI) - A scale to measure spiritual and religious beliefs in quality-of-life and coping research. *Psychosomatic Medicine*, *57*, 62-62.
- Kauffman, J. H. (1979). Social correlates of spiritual maturity among North American Mennonites. In D. O. Moberg (Ed.), *Spiritual well-being: Sociological perspectives* (pp. 237-254). Washington, DC: University Press of America.
- Kennedy, J. E., Davis, R. C., & Taylor, B. G. (1998). Changes in spirituality and well-being among victims of sexual assault. *Journal for the Scientific Study of Religion*, *37*, 322-328.
- Kim, J., Heinemann, A. W., Bode, R. K., Sliwa, J., & King, R. B. (2000). Spirituality, quality of life, and functional recovery after medical rehabilitation. *Rehabilitation Psychology*, *45*, 365-385.
- King, M. B., & Hunt, R. A. (1975). Measuring the religious variable: National replication. *Journal for the Scientific Study of Religion*, *14*, 13-22.
- Koenig, H. G., Parkerson, G. R., & Meador, K. G. (1997). Religion index for psychiatric research: A 5-item measure for use in health outcome studies. *American Journal of Psychiatry*, *154*, 885-886.
- Landis, B. J. (1996). Uncertainty, spiritual well-being, and psychosocial adjustment to chronic illness. *Issues in Mental Health Nursing*, *17*, 217-231.
- LaPierre, L. L. (1994). A model for describing spirituality. *Journal of Religion and Health*, *33*, 153-161.
- Larson, D. B., Sawyers, J. P., & McCullough, M. E. (1998). *Scientific research on spirituality and health: A report based on the Scientific Progress in Spirituality Conferences*. New York: John M. Templeton Foundation.
- Levin, J. (1994). Religion and health: Is there an association, is it valid, and is it causal? *Social Science & Medicine*, *38*, 1475-1482.
- Levin, J. S., Chatters, L. M., & Taylor, R. J. (1995). Religious effects on health-status and life satisfaction among black-Americans. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences*, *50*, S154-S163.
- Levin, J. S., & Taylor, R. J. (1998). Panel analyses of religious involvement and well-being in African Americans: Contemporaneous vs. longitudinal effects. *Journal for the Scientific Study of Religion*, *37*, 695-709.
- Lewis, C. A., Lanigan, C., Joseph, S., & de Fockert, J. (1997). Religiosity and happiness: No evidence for an association among undergraduates. *Personality and Individual Differences*, *22*, 119-121.

- Lewis, C. A., & Maltby, J. (1995). Reliability and validity of the Francis Scale of Attitude Towards Christianity among U.S. adults. *Psychological Reports, 76*, 1243-1247.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- MacDonald, D. A. (2000). Spirituality: Description, measurement, and relation to the five factor model of personality. *Journal of Personality, 68*, 153-197.
- MacDonald, D. A., Friedman, H. L., & Kuentzel, F. G. (1999). A survey of measures of spiritual and transpersonal constructs: Part one - research update. *Journal of Transpersonal Psychology, 31*, 137-154.
- MacDonald, D. A., Kuentzel, F. G., & Friedman, H. L. (1999). A survey of measures of spiritual and transpersonal constructs: Part two - additional instruments. *Journal of Transpersonal Psychology, 31*, 155-177.
- MacDonald, D. A., LeClair, L., Holland, C. J., Alter, A., & Friedman, H. L. (1995). A survey of measures of transpersonal constructs. *Journal of Transpersonal Psychology, 27*, 171-231.
- Maranell, G. M. (1974). *Response to religion: Studies in the social psychology of religious belief*. Lawrence, KS: University Press of Kansas.
- Martinez, R. J. (1999). Close friends of God: An ethnographic study of health of older Hispanic adults. *Journal of Multicultural Nursing & Health, 5*(1), 40-45.
- Maton, K. I. (1989). The stress-buffering role of spiritual support: Cross-sectional and prospective investigations. *Journal for the Scientific Study of Religion, 28*, 310-323.
- Maton, K. I., & Zimmerman, M. A. (1992). Psychosocial predictors of substance use among urban Black male adolescents. *Drugs and Society, 6*(1-2), 79-113.
- McGrath, P. (1997). Putting spirituality on the agenda: Hospice research findings on the "ignored" dimension. *Hospice Journal Physical, Psychosocial, & Pastoral Care of the Dying, 12*(4), 1-14.
- McHorney, C. A., Ware, J. E., Jr., & Raczek, A. E. (1993). The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical Care, 31*, 247-263.
- McQuaide, S. (1998). Women at midlife. *Social Work, 43*, 21-33.
- National Library of Medicine. (2002, March 5). Medical subject headings. Retrieved April 1, 2002, from: <http://www.nlm.nih.gov/mesh/meshhome.html>
- Meller, S. (2001). A comparison of the well-being of family caregivers of elderly patients hospitalized with physical impairments versus the caregivers of patients hospitalized with dementia. *Journal of the American Medical Directors Association, 2*(2), 60-65.

- Mellors, M. P., Riley, T. A., & Erlen, J. A. (1997). HIV, self-transcendence, and quality of life. *Journal of the Association of Nurses in AIDS Care*, 8(2), 59-69.
- Mol, H. (1970). Religion and sex in Australia. *Australian Journal of Psychology*, 22, 105-114.
- Moody, L. (1990). Meta-analysis: Qualitative and quantitative methods. In L. Moody (Ed.), *Advancing nursing science through research* (Vol. 2, pp. 70-109). Newbury Park, CA: Sage.
- Mytko, J. J., & Knight, S. J. (1999). Body, mind and spirit: Towards the integration of religiosity and spirituality in cancer quality of life research. *Psycho-Oncology*, 8, 439-450.
- Neill, C. M., & Kahn, A. S. (1999). The role of personal spirituality and religious social activity on the life satisfaction of older widowed women. *Sex Roles*, 40, 319-329.
- Nelson, E. C., Landgraf, J. M., Hays, R. D., Wasson, J. H., & Kirk, J. W. (1990). The functional status of patients. How can it be measured in physicians' offices? *Medical Care*, 28, 1111-1126.
- Neugarten, B. L., Havigurst, R. J., & Tobin, S. S. (1961). The measurement of life satisfaction. *Journal of Gerontology*, 24, 465-469.
- Orwin, G. R. (1994). Evaluating coding decisions. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 139-162). New York: Russell Sage Foundation.
- Padilla, G. V., & Frank-Stromborg, M. (1997). Single instruments for measuring quality of life. In M. Frank-Stromborg & S. J. Olsen (Eds.), *Instruments for clinical health-care research* (2nd ed., pp. 114-134). London: Jones and Bartlett.
- Paloutzian, R., & Ellison, C. (1982). Loneliness, spiritual well-being and quality of life. In L. Peplau & D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research and therapy* (pp. 224-237). New York: John Wiley.
- Pargament, K. I. (1999). The psychology of religion and spirituality? Yes and no. *International Journal for the Psychology of Religion*, 9(1), 3-14.
- Pargament, K. I., Smith, B. W., Koenig, H. G., & Perez, L. (1998). Patterns of positive and negative religious coping with major life stressors. *Journal for the Scientific Study of Religion*, 37, 710-724.
- Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment*, 5, 164-172.
- Peacock, J. R., & Poloma, M. M. (1999). Religiosity and life satisfaction across the life course. *Social Indicators Research*, 48, 321-345.

- Peterman, A. H., Fitchett, G., Brady, M. J., Hernandez, L., & Cella, D. (2002). Measuring spiritual well-being in people with cancer: The Functional Assessment of Chronic Illness Therapy-Spiritual Well-Being Scale (FACIT-Sp). *Annals of Behavioral Medicine, 24*(1), 49-58.
- Pfeifer, S., & Waelty, U. (1995). Psychopathology and religious commitment--a controlled study. *Psychopathology, 28*(2), 70-77.
- Pippalla, R. S., & Chaar, M. D. C. (2001). An assessment of quality of life in ambulatory schizophrenics: World Health Organization Quality of Life assessment scale (WHOQOL-100) concepts quality of life of ambulatory schizophrenics. *Clinical Research & Regulatory Affairs, 18*(1-2), 53-65.
- Poloma, M. M., & Pendleton, B. F. (1990). Religious domains and general well-being. *Social Indicators Research, 22*, 255-276.
- Poloma, M. M., & Pendleton, B. F. (1991). The effects of prayer and prayer experiences on measures of general well-being. *Journal of Psychology and Theology, 19*(1), 71-83.
- Post-White, J., Ceronsky, C., Kreitzer, M. J., Nickelson, K., Drew, D., Mackey, K. W., et al. (1996). Hope, spirituality, sense of coherence, and quality of life in patients with cancer. *Oncology Nursing Forum, 23*, 1571-1579.
- Potts, R. G. (1996). Spirituality and the experience of cancer in an African-American community: Implications for psychosocial oncology. *Journal of Psychosocial Oncology, 14*(1), 1-19.
- Pratheepawanit, N., Salek, M. S., & Finlay, I. G. (1999). The applicability of quality-of-life assessment in palliative care: Comparing two quality-of-life measures. *Palliative Medicine, 13*, 325-334.
- Reed, P. G. (1983). Well-being and perspectives on life and death among death-involved and non-death-involved adults. (Doctoral dissertation, Wayne State University, 1982). *Dissertation Abstracts International, 43*, 3538
- Reed, J. (1991). Self-transcendence and mental health in oldest-old adults. *Nursing Research, 39*, 355-359.
- Reed, J. G., & Baxter, P. M. (1994). Using reference databases. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 51-70). New York: Russell Sage Foundation.
- Reed, P. G. (1986). Religiousness among terminally ill and healthy adults. *Research in Nursing & Health, 9*, 35-41.
- Reed, P. G. (1987). Spirituality and well-being in terminally ill hospitalized adults. *Research in Nursing & Health, 10*, 335-344.

- Reed, P. G. (1993). An emerging paradigm for the investigation of spirituality in nursing. *Research in Nursing & Health, 15*, 349-357.
- Rehm, R. S. (1999). Research. Religious faith in Mexican-American families dealing with chronic childhood illness. *Image: The Journal of Nursing Scholarship, 31*(1), 33-38.
- Riley, B. B., Perna, R., Tate, D. G., Forchheimer, M., Anderson, C., & Luera, G. (1998). Types of spiritual well-being among persons with chronic illness: Their relation to various forms of quality of life. *Archives of Physical Medicine & Rehabilitation, 79*, 258-264.
- Ringdal, G. I. (1996). Religiosity, quality of life, and survival in cancer patients. *Social Indicators Research, 38*, 193-211.
- Ringdal, G. I., & Ringdal, K. (1993). Testing the EORTC Quality of Life Questionnaire on cancer patients with heterogeneous diagnoses. *Quality of Life Research, 2*, 129-140.
- Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Newbury Park, CA: Sage.
- Ruffing Rahal, M. A. (1991). Initial psychometric evaluation of a qualitative well-being measure: The Integration Inventory. *Health Values: Health Behavior, Education and Promotion, 15*(2), 10-20.
- Sawatzky, R. (2002). *Spirituality and outcome measures: Examining the feasibility of a meta-analysis of existing literature*. Unpublished manuscript.
- Schag, C. A., & Heinrich, R. L. (1990). Development of a comprehensive quality of life measurement tool: CARES. *Oncology, 4*, 135-138.
- Schag, C.A., Heinrich, R.L., Aadland, R.L., & Ganz, P.A. (1990). Assessing problems of cancer patients: Psychometric properties of the cancer inventory of problem situations. *Health Psychology, 9*, 83-102.
- Schipper, H., Clinch, J., McMurray, A., & Levitt, M. (1984). Measuring the quality of life of cancer patients: The Functional Living Index-Cancer - development and validation. *Journal of Clinical Oncology, 2*, 472-483.
- Simmons, Z., Bremer, B. A., Robbins, R. A., Walsh, S. M., & Fischer, S. (2000). Quality of life in ALS depends on factors other than strength and physical function. *Neurology, 55*, 388-392.
- Slanter, W., Hall, T. W., & Edwards, K. J. (2001). Measuring religion and spirituality: Where are we and where are we going? *Journal of Psychology and Theology, 29*(1), 4-21.
- Sloan, R. P., Bagiella, E., & Powell, T. (1999). Religion, spirituality, and medicine. *Lancet, 353*, 664-667.

- Spitzer, W. O., Dobson, A. J., Hall, J., Chesterman, E., Levi, J., Shepherd, R., et al. (1981). Measuring the quality of life of cancer patients: A concise QL-index for use by physicians. *Journal of Chronic Diseases, 34*, 585-597.
- Stewart, A. L., Hays, R. D., & Ware, J. E. (1988). The MOS Short-form General Health Survey. *Medical Care, 26*, 724-735.
- Stewart, A. L., & Ware, J. E. (Eds.). (1992). *Measuring functioning and well-being: The medical outcomes study approach*. Durham, NC: Duke University Press.
- Stock, W. A. (1994). Systematic coding for research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 125-138). New York: Russell Sage Foundation.
- Sussman, S., Nezami, E., & Mishra, S. (1997). On operationalizing spiritual experience for health promotion research and practice. *Alternative Therapies in Clinical Practice, 4*, 120-125.
- Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: HarperCollins College.
- Tate, D. G., Riley, B. B., Perna, R., & Roller, S. (1997). Quality of life issues among women with physical disabilities or breast cancer. *Archives of Physical Medicine & Rehabilitation, 78*(12 Suppl 5), S18-25.
- Thomas, M. E., & Holmes, B. J. (1992). Determinants of satisfaction for Blacks and Whites. *Sociological Quarterly, 33*, 459-472.
- Thoresen, C. E. (1999). Spirituality and health: Is there a relationship? *Journal of Health Psychology, 4*, 291-300.
- Tongprateep, T. (2000). The essential elements of spirituality among rural Thai elders. *Journal of Advanced Nursing, 31*, 197-203.
- Tuck, I., McCain, N. L., & Elswick, R. K., Jr. (2001). Spirituality and psychosocial factors in persons living with HIV. *Journal of Advanced Nursing, 33*, 776-783.
- Van Wegberg, B., Bacchi, M., Heusser, P., Helwig, S., Schaad, R., von Rohr, E., et al. (1998). The cognitive-spiritual dimension - an important addition to the assessment of quality of life: Validation of a questionnaire (SELT-M) in patients with advanced cancer. *Annals of Oncology, 9*, 1091-1096.
- Vaughan, F., Wittine, B., & Walsh, R. (1996). Transpersonal psychology and the religious person. In E. D. Shafranske (Ed.), *Religion and the clinical practice of psychology* (pp. 483-510). Washington, DC: American Psychological Association.

- Vella-Brodrick, D. A., & Allen, F. C. (1995). Development and psychometric validation of the Mental, Physical, and Spiritual Well-Being Scale. *Psychological Reports, 77*, 659-674.
- Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual framework and item selection. *Medical Care, 30*, 473-483.
- Westlake, C., Dracup, K., Creaser, J., Livingston, N., Heywood, J. T., Huiskes, B. L., et al. (2002). Correlates of health-related quality of life in patients with heart failure. *Heart & Lung, 31*(2), 85-93.
- White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 41-53). New York: Russell Sage Foundation.
- WHOQOL-Group. (1998). The World Health Organization Quality of Life Assessment (WHOQOL): Development and general psychometric properties. *Social Science & Medicine, 46*, 1569-1585.
- Wolf, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. London: Sage.
- Wood, V., Wylie, M., & Sheafor, B. (1969). An analysis of a short self-report measure of life satisfaction: Correlation with rater judgments. *Journal of Gerontology, 24*, 465-469.
- Woodworth, G. (1994). Managing meta-analytic databases. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis*. (pp. 177-189). New York: Russell Sage Foundation.
- Wulff, D. M. (1999). Religious attitudes scales. In P. C. Hill & R. W. Hood (Eds.), *Measures of religiosity* (pp. 314-320). Birmingham, AL: Religious Education Press.
- Wyatt, G., Kurtz, M. E., Friedman, L. L., Given, B., & Given, C. W. (1996). Preliminary testing of the Long-Term Quality of Life (LTQL) instrument for female cancer survivors. *Journal of Nursing Measurement, 4*, 153-170.
- Zinnbauer, B. J., Pargament, K. I., Cole, B., Rye, M. S., Butter, E. M., Belavich, T. G., et al. (1997). Religion and spirituality: Unfuzzifying the fuzzy. *Journal for the Scientific Study of Religion, 36*, 549-564.
- Zinnbauer, B. J., Pargament, K. I., & Scott, A. B. (1999). The emerging meanings of religiousness and spirituality: Problems and prospects. *Journal of Personality, 67*, 889-919.

Appendix A:
Selection Criteria Form

<p>Instructions: Enter [1] if the criterion leads to the inclusion of a study and [0] for those criteria that cause a study to be excluded (i.e., only studies with no zero's will be included). First check all the exclusion criteria. Review all the inclusion criteria if all exclusion criteria are marked [1]. Do not complete the inclusion criteria sections if any of the exclusion criteria are marked [0].</p>		
	NUMBER	Enter study number.
	REVIEWER	Enter [1] for Rick and [2] for Glenna.
	AUTHOR	Enter the name of the first author
	YEAR	Enter the publication year
Enter [1] or [0]	CODE	<u>Substantive Inclusion Criteria</u>
	SUB1	1. Descriptive studies that examine the relationship between spirituality and quality of life.
	SUB2	2. Studies that operationalize spirituality and quality of life in a manner that is consistent with the conceptualization and subjective nature of these concepts as discussed in the research proposal (i.e., the instruments are not based on external ratings of particular practices or behaviors that are considered to be representative of these concepts).
		<u>Methodological Inclusion Criteria</u>
	MET1	1. Quantitative correlational or descriptive studies that provide sufficient statistical data to calculate an estimated effect size of the correlation between spirituality and quality of life (i.e., statistics pertaining to the relationship between spirituality and quality of life are provided).
		<u>Measurement of Spirituality</u>
	MET2	2. Studies that measure spirituality by using subjective ratings (i.e., self-reported ratings on a Likert-type scale) of spiritual experiences, beliefs or behaviours associated with:
		a. an existential search for meaning or purpose in life, <u>and</u>
		b. a relationship with a transcendent reality (e.g. God or divine being, Ultimate Reality, or Ultimate Truth).
		<u>Measurement of Quality of Life / Wellbeing / Life-satisfaction</u>
	MET3	3. Studies that measure quality of life by:
		a. using subjective ratings of a person's self-reported quality of life, satisfaction with life or subjective well-being, <u>or</u>
		b. using instruments that measure quality of life by statistically combining subjective ratings of multiple dimensions of life (e.g., physical, social, psychological and emotional dimensions). This refers primarily to broad multi-dimensional measures of quality of life or wellbeing.

		<u>Exclusion Criteria</u>
	EX1	1. Case studies and studies that do not report primary research findings involving human participants.
	EX2	2. Studies of an exclusively qualitative design.
	EX3	3. Studies that do not operationalize quality of life, wellbeing or life-satisfaction <u>or</u>
		b. Studies that exclusively use diagnostic indicators to measure quality of life (e.g., based on variables such as social economic status or illness-outcomes) <u>or</u>
		c. Studies that exclusively measure only one dimension of quality of life or wellbeing (e.g., physical, emotional, psychological, social or functional).
	EX4	4. Studies that do not operationalize spirituality <u>or</u>
		b. Studies that exclusively measure spirituality in terms of the frequency of religious or spiritual practices or behaviors as opposed to an individual's self-rated experience of spirituality.
	EX5	4. Studies that are not reported in English.
	EX6	5. Studies that were published during or after 1991 or dissertations that were completed during or after 1996.
	FINAL	Final selection: Enter "yes" only if all selection criteria have been met. Enter "no" any of the selection criteria has not been met.
Enter any additional comments pertaining to the selection or exclusion of this article here:		

Appendix B

Overview of Database Searches

Search Date	Database / source	Coverage	Updated	Disciplinary scope	Provider	Limits	Scope	Results
23-Apr-02	PubMed	1966 - present	Daily	Medicine / Allied health	National Library of Medicine	English	Mesh; all fields	1053
29-Apr-02	Dissertations Abstracts	1861 - present	Monthly	General	Proquest	1992-2002	Title; keywords; abstracts	1120
1-May-02	EMBASE	1988 - present	Monthly	Medical / biomedical	Elsevier Science	none	Title; Mesh; abstract	969
2-May-02	HEALTH-STAR	1975 - present	Monthly	Health services technology, administration & research	National Library of Medicine	none	Title; Mesh: abstract; text fields	1153
2-May-02	CINAHL		Monthly	Allied health	CINAHL Information Systems	Research and English	All fields	520
2-May-02	PsychINFO	1887 - present	Monthly	Psychology and psychological aspects of related disciplines, such as medicine, psychiatry, nursing, sociology, education, pharmacology, physiology, linguistics, anthropology, business, and law	American Psychological Association	English AND 1970-2002	All fields including: title; keywords and abstracts	1201
2-May-02	Science Citation Index Expanded Social Sciences Citation Index Arts and Humanities Citation Index	1989 - present	Weekly	Physical Sciences, Medical Sciences, Life Sciences, Applied Sciences, Agriculture, Humanities and Social Sciences, Law, Business	Institute for Scientific Information	English	Title; keywords; abstracts	1493

13-Jun-02	Science Citation Index Expanded Social Sciences Citation Index	1989 - present	Weekly	See above	See above	none	Citation search	71
Subtotal								7580
Dissertations ¹								- 1644
Publications prior to 1991 ²								- 973
Duplicates ³								- 1923
Total published reports								3040

¹ Dissertations were retrieved by the Dissertations Abstracts Database as well as by the PsychINFO database. Many of these dissertations are therefore duplicates.
² Number of studies that were published prior to 1991.
³ Duplicates were identified and removed from the database throughout the screening process.

Appendix C
Code Book

A Meta-Analysis of the Relationship Between Spirituality and Quality of Life			
Report Level Coding Form			
Rick Sawatzky - Revised June 15, 2002			
General comments: All fields must be filled in with either a quantitative statistic or code or a qualitative descriptor of an item that does not fit the pre-coded categories. Added categories will be coded once the data collection is completed. If information on a particular item is unknown then a "999" must be entered in that field. No field should be left completely blank. Remember to enter [0] for each zero value.			
CODE	VALUE		Explanation and comments
1. Report Identification			
STUDID		Study ID	Give the study a three digit ID that is unique for each study. It is important to code the study separately from the report since several reports might comment on the same study. Effect sizes are calculated in relation to the different studies.
AUTHORS		Author name(s)	
YEAR		Year of publication	
CALLNUM		Report call number	A three digit callnumber corresponding to the Endnote database is listed on the right top corner of each report.
NUMREP		Number of reports reviewed in relation to this study	
TYPEPUB		Type of publication 1. Peer reviewed journal article / book chapter 2. Non-peer reviewed journal / book chapter 3. Book 4. Thesis or dissertation 5. Conference paper Or enter a description if another category needs to be added.	The type of publication provides a way of evaluating the impact of publication characteristics on the cumulative effect size estimations.
SOURCE		Source derivation 1. Electronic database search 2. Ancestry search 3. Expert consultation 4. Browsing	
LANG		Language 1. Original study in English 2. Translated study	
CODERID		Coder Coder 1: Rick Sawatzky Coder 2: Glenna Stewart	

2. Setting		
SCOPE		Scope of sampling
		Select a descriptor of the general scope of the population from which the data was drawn.
		1. Local
		Data were drawn from a particular town, city, local community or institution.
		2. Regional
		Data were drawn from an area consisting of multiple communities, cities or towns.
		3. National
		Data were drawn from a whole nation geographically.
		4. International
		Data were drawn from a large variety of different nations.
SETTING		Setting
		1. Institutions
		2. Community / home setting
		3. Both
		Identify whether or not the data were collected from individuals living in an institutionalized context or within a non-institutionalized community setting (i.e., outside of the context of being admitted to an institution for medical care). Independent living arrangements in facilities are considered a community setting if no medical care provided.
		Special populations
SPPOPGEN		% General population
SPPOPACS		% Acute surgical
SPPOPACM		% Acute medical
SPPOPCHM		% Chronic medical / disability
SPPOPPAL		% Palliative
SPPOPPSC		% Psychiatric
SPPOPSUB		% Substance abuse
SPPOPHIV		% PWAIDS / HIV
SPPOPTH		% Other
SPPOPDES		Describe
		It is important to know the type of population from which the sample was drawn to be able to make valid generalizations and to assess whether different populations might affect the effect size between spirituality and quality of life. Enter percentages related to each general descriptor of the type of population from which the sample is drawn. Other descriptors and percentages can be added in the "describe" column below.
DISRES		Discipline of researcher(s)
		1. Medicine
		2. Nursing
		3. Social Sciences
		4. Psychology
		5. Anthropology
		6. Multidisciplinary
		Unknown [999]
		Describe
		Only identify a discipline if the whole research team clearly represents one discipline. Record an other discipline in the describe column if a discipline is not listed. Lipsey and Wilson (2001) suggest that coding source descriptors, such as the discipline in which the study occurred and the training of the researchers, may serve a descriptive as well as an analytic purpose when variables such as the discipline of the researchers and the degree of the researchers systematically account for differences in the analysis.
DISDES		
		"Sometimes . . . variables in this category are proxies for a substantive or methodological variable that might not otherwise get reported and coded in the studies" (p. 84).
TRAINRSH		Training of researcher(s)
		1. BA
		2. MA
		3. M.D.
		4. PhD
		Record the "highest" degree amongst the authors of this study.

TRAINRSL		1. B.A. 2. M.A. 3. M.D. 4. PhD	Record the "lowest" degree amongst the authors of this study.
----------	--	---	---

3. Methodology			
RESDES		Research design 1. Descriptive correlation 2. Longitudinal - correlational Describe	The list of research designs is very limited because most studies that are relevant to this meta-analysis will be of a correlational or descriptive nature. Other designs can be added to the list and coded afterwards as needed. Again, coding this type of information serves a descriptive purpose and might provide an analytical purpose by allowing further analysis of potential variability between studies. Please enter a brief description of the methodology in the space provided.
RESDESD			
SAMMETH		Sampling method 1. Simple random sampling 2. Stratified sampling 3. Cluster sampling 4. Systematic probability sampling 5. Convenience sampling 6. Quota sampling 8. Purposive sampling 9. Systematic nonprobability sampling Describe	Identify the sampling method used. Use the describe option to enter sampling methods not included in the list.
SAMMETHD			
CONVOTE		Consent vote (%)	Enter the number of people who consented as a percentage of the number of people who were contacted.
ATTR		Attrition (%)	Enter the number of people who did not complete the study as a percentage of the total number of people who consented.

4. Coding of Primary Sample			
This section only codes information about the general sample. Statistics of each subsample related to a particular effect size will be coded in the effect size form. The primary sample refers to the complete sample involved in this study.			
		Age Youngest Oldest Average SD	Enter age related data here.
PSAGEYNG			
PSAGEOLD			
PSAGEAVE			
PSAGESD			
		Gender % Male % Female	Enter gender information. Leave blank if unknown.
PSPERMAL			
PSPERFEM			

		Culture / ethnicity	<p>It is important to most accurately retrieve the cultural description of the sample because culture and spirituality may be correlated. The culture/ethnicity list provides a very superficial overview of general cultural characteristics that are typically collected as part of the demographic data. However, many other cultural variables may be reported in a variety of different ways. If the reported cultural characteristics do not match the characteristics below, enter a detailed qualitative description of the cultural characteristics reported as part of the demographic data and include percentages of each specific cultural group represented in the sample. The data entered here will be coded after all the data are collected.</p>
PSCUA		% Caucasian	
PSASIA		% Asian	
PSAFR		% African American	
PSHISP		% Hispanic	
PSSTHAM		% South American / Latin American	
PSOTHER		% Other / unknown	
PSETHDES		Describe	
		Religious affiliation / orientation	<p>Enter the percentage of people who state they are religious or are explicitly affiliated with a religious community as indicated below. Religious data can be collected in a variety of different ways. Use the description box at the bottom to enter a description and percentage of each religious affiliation that does not fit any of these categories. Also use the description box to enter religious data that describe religious characteristics other than religious affiliation. For example, if the question "do you consider yourself as being religious" is asked, then you would record this question with the associated percentage responses.</p>
PSRELCAT		% Catholic / Orthodox	
PSRELPRO		% Protestant	
PSRELJEW		% Jewish	
PSRELHIN		% Hindu	
PSRELBUD		% Buddhist	
PSRELISL		% Islam	
PSRELOTH		% Other	
PSRELNON		% No religious affiliation	
PSRELUNK		% Unknown	
		Description of other religious variables.	
PSSIZE		Sample size	Enter the size of the sub sample.

<p align="center">A Meta-Analysis of the Relationship Between Spirituality and Quality of Life Effect Size Level Code Book Rick Sawatzky - Revised June 15, 2002</p>			
CODE	VALUE		Explanation / comments
EFFECTID		Effect Size Identification	To calculate the mean effect size it is important to be able to identify which effect sizes are independent from one another. Effect sizes that are derived from the same sample or from the same instruments are not independent and should therefore not be included in the same calculations (Lipsey & Wilson, 2001). However, ES's of subsamples and subscales may still be used for other comparative calculations.
AUTHORS		Author name(s)	
YEAR		Year of publication	
PRIMES		Primary ES identification	Selection of the primary ES is based on the following criteria:
		Primary Effect Size (enter [1])	i. If different operationalizations of spirituality and/or quality of life are used in relation to the same sample, then the primary effect size is selected based on those operational definitions that most closely reflect the conceptualizations of spirituality and quality of life underlying this meta-analysis (see operational definitions criteria). ii. If there are two equally representative measures of spirituality or quality of life, then the instrument with the highest reliability and validity claims will be selected as the one used in calculating the primary effect size (i.e., multi-item measures are generally preferred over single-item instruments). iii. The primary effect size relates to a unique sample (i.e., no other primary effect sizes have been coded in relation to the same sample). iv. If the above criteria do not clearly distinguish the primary ES: 1. Select the ES that relates to the primary instrument as opposed to a subscale. 2. Use a conservative approach if there are more than one possible primary ES's (i.e., select the smallest effect size).
		Secondary Effect Size (enter [0])	Secondary ES's are those effect sizes of interest that are not independent of a primary ES because: i. they are derived from the same sample or sub-sample, or ii. they are based on or derived from a subscale of an instrument used in the calculation of the primary effect size.

STUDID		Study ID	
CALLNUM		Report ID	
Coder ID		Coder ID	
Operational definitions: Spirituality variable			
SPMEAS		Record the name of this measure / variable	Record the abbreviation first, followed by the complete name. It is important to be able to analyse the results in relation to different operational definitions. Each operational definition will receive a unique identifier once the list is completed.
SPAUTH		Record the complete reference to this measure (author names only)	For documentation and validity purposes. Some references come with similar names. Having reference information will also help in retrieving instrument validity and reliability information.
SPDECR		Record a detailed description of this measure	The description should include a description of the operational definition and related subscales. It would be helpful to include a list of the items if this is available.
SPSUBDEF		Select one of the following:	Select one of the descriptions pertaining to the degree of fit between the spirituality instruments and the conceptualization of spirituality underlying this meta-analysis.
		1. Explicit measures of existential and relational/transcendent dimensions	The instrument contains items that clearly measure the existential as well as the relational or transcendent dimensions of spirituality.
		2. Primarily existential measures of spirituality	This category refers to broad instruments containing items that primarily measure the existential dimension while also containing items that have relational connotations. Examples include those instruments that clearly measure the existential dimension while using less direct indicators of the relational dimension such as frequency of prayer.
		3. Primarily relational measures of spirituality	This category refers to broad instruments containing items that primarily measure the relational dimension of spirituality while also containing items that have existential connotations. Examples include measures of closeness to God, prayer and meditation that also refer to existential meaning and purpose.
		4. Ambiguous measures of spirituality	This category refers to instruments that ask people to rate their degree of spirituality or religiosity. It is unclear exactly what is being measured because we do not know how the term spirituality or religiosity is interpreted by the participants. The instrument may also include one or more items pertaining to specific religious or spiritual practices. Nevertheless, these instruments are of interest based on the assumption that participants would intuitively consider the existential and the relational dimensions of spirituality when responding to these types of instruments.

SPSUBES		Were additional effect sizes calculated for any subscales?	If the measure contains a subscale, indicate whether or not an effect size was calculated for this subscale by indicating 1 for yes, 0 for no or 999 for don't know.
SPMOD		Were any modifications to the original instrument made for this study? [0] = no; [1] = yes.	Sometimes researchers modify the original instruments for the purposes of their study. Tracking these modifications may provide important data when the modifications actually result in conceptual changes.
SPMODDES		Describe any modifications that were made.	
SPITEMS		Number of items that make up this scale.	Record the number of items that make up this scale.
SPITUSED		Number of items of the scale that were used in this study.	Record the number of items used in this study. Sometimes authors will choose to exclude particular items.
		If this measure is based on a subscale from another measure:	Some instruments are subscales of other measures. Identifying these relationships is important as they affect the analysis.
SPPRIM		Record the primary measure	Record the abbreviation first, followed by the complete name of the primary instrument.
SPPRIMAU		Record the reference to the primary measure	Record the complete reference to the primary instrument.
SPSCTYPE		Record the type of scale used 1. Ordinal / Likert 2. Dichotomous 3. Interval [999] Not known	Likert scales are considered ordinal even when statistical techniques are used afterwards to estimate their continuous equivalent.
SPRELSTR		Test-retest (Pearson's r) value	Record reliability measures calculated in this study
SPRELSTA		Internal consistency (Cronbach's alpha) value	
		Reference to reliability reported in other studies	Only report reliability of the instrument as discussed in the present study in relation to its current use. If reliability measures are only reported of the overall instrument but the effect size relates to a subscale then do not record this measure. Similarly, if the reliability measure reported only refers to the subscales but you are using the overall instrument then do not record this measure.
SPRELSTRH		Highest test-retest (Pearson's r) value reported	
SPRELSTRL		Lowest test-retest (Pearson's r) value reported	
SPRELSTAH		Highest internal consistency (Cronbach's alpha)	
SPRELSTAL		Lowest internal consistency (Cronbach's alpha)	
SPRELSO		Source of reliability measures	Report the entire reference to these reliability measures.
SPVALSO		Source of validity 1. Validity is established in the present study 2. Validity is established in other study to which the primary researcher(s) refer(s) 3. both	Record the documented source of validity to which the researcher refers. Leave blank [-] if validity was not discussed.

	Type of validity	Enter [1] for each degree of validity indicated by the researcher(s), enter [999] for those types of validity that are not discussed or reported. Only report the types of validity as discussed by the researchers of the present study.
SPVALEX	1. Expert validity / face validity	
SPVALCON	2. Content validity	
SPVALCRI	3. Criterion-related validity	
SPVALCNS	4. Construct validity	
Operational definitions: Quality of life variable		
QLMEAS	Record the name of this measure / variable	Record the abbreviation first, followed by the complete name. It is important to be able to analyse the results in relation to different operational definitions. Each operational definition will receive a unique identifier once the list is complete.
QLAUTH	Record the complete reference to this measure (author names only)	Some references come with similar names. Having reference information will help in retrieving instrument validity and reliability information.
QLDECR	Record a detailed description of this measure	The description should include a description of the operational definition and related subscales. It would be helpful to include a list of the items if this is available.
QLSUBDEF	Substantive descriptor of this instrument:	This category is included to be able to examine the relationship between spirituality and different types of operational variables associated with perceived quality of life.
	1. multidimensional measurement of quality of life	
	2. unidimensional measurement of the perception of quality of life	
	3. measurement of subjective well being	
	4. measurement of life satisfaction	
	5. other (describe)	
QLSUBES	Were additional effect sizes calculated for any subscales?	If this measure contains a subscale indicate whether an effect size was calculated for the subscale by indicating 1 for yes, 0 for no, or 999 for don't know.
QLITEMS	Number of items that make up this scale.	Record the number of items that make up this scale.
QLITUSED	Number of items of which this measure consists	Record the number of items used in this study. Sometimes authors will choose to exclude particular items.

		If this measure is based on a subscale from another measure:	Some instruments are subscales of other measures. Identifying these relationships is important as they affect the analysis.
QLPRIM		Record the primary measure	Record the abbreviation first, followed by the complete name of the primary instrument.
QLPRIMAU		Record the reference to the primary measure	Record the complete reference to the primary instrument.
QLSCTYPE		Record the type of scale used	
		1. Ordinal / Likert	Likert scales are considered ordinal even when statistical techniques are used afterwards to estimate their continuous equivalent.
		2. Dichotomous	
		3. Continuous	
		[999] Not known	
		Type of reliability reported in relation to current sample	Record reliability measures calculated in this study.
QLRELSTR		Test-retest (Pearson's r) value	
QLRELSTA		Internal consistency (Cronbach's alpha) value	
		Reference to reliability reported in other studies	Only report reliability of the instrument as discussed in the present study in relation to its current use. If reliability measures are only reported of the overall instrument but the effect size relates to a subscale then do not record this measure. Similarly, if the reliability measure reported only refers to the subscales but you are using the overall instrument then do not record this measure.
QLRELSTRH		Highest test-retest (Pearson's r) value reported	
QLRELSTRL		Lowest test-retest (Pearson's r) value reported	
QLRELSTAH		Highest internal consistency (Cronbach's alpha) value reported	
QLRELSTAL		Lowest internal consistency (Cronbach's alpha) value reported	
QLRELSO		Source of reliability measures	Report the entire reference to these reliability measures.
QLVALSO		Source of validity	Record the documented source of validity to which the researcher refers. Enter 999 if validity was not discussed.
		1. Validity is established in the present study	
		2. Validity is established in other study to which the primary researcher(s) refer(s)	
		3. both	
		[999] Unknown / not discussed	
		Type of validity	Enter [1] for each degree of validity indicated by the researcher(s).
QLVALEX		Expert validity	Enter [1] for each degree of validity indicated by the researcher(s), enter [999] for those types of validity that are not discussed or reported. Only report the types of validity as discussed by the researchers of the present study.
QLVALCON		Content validity	
QLVALCRI		Criterion validity	
QLVALCNS		Construct validity	
		[999] Unknown / not discussed	

Sample			
SAMPLE		Does this effect size relate to the whole sample or to a subsample? 1. Whole sample 2. Subsample [999] Unknown	Enter #1 if this effect size relates to the general sample as described earlier. You do not need to code the rest of section 4 (sub-sample) if this effect size pertains to the general sample which was coded in the report level coding form.
SSCHAR		Sub-sample characteristics 1. Age group 2. Gender 3. Culture or ethnicity 4. Religion / religiosity 5. Medical condition 6. A mixture of diverse characteristic Other (enter description)	Select the primary characteristic that distinguishes this subsample from the general sample.
		Age	Enter age related data here. Enter 999 if unknown.
SSAGEYNG		Youngest	
SSAGEOLD		Oldest	
SSAGEAVE		Average	
SSAGESD		SD	
		Gender	Enter gender information. Enter 999 if unknown.
SSPERMAL		% Male	
SSPERFEM		% Female	
		Culture / ethnicity	It is important to most accurately retrieve the cultural description of the sample because culture and spirituality may be correlated. The above list provides a very superficial overview of general cultural characteristics that are typically collected as part of the demographic data. However, many other cultural variables may be reported in a variety of different ways. If the reported cultural characteristics do not match the characteristics below, enter a detailed qualitative description of the cultural characteristics reported as part of the demographic data and include percentages of each specific cultural group represented in the sample. The data entered here will be coded after all the data is collected. Please ensure that all the culture/ethnicity percentages add up to 100%.
SSCUA		% Caucasian	
SSASIA		% Asian	
SSAFR		% African American	
SSHISP		% Hispanic	
SSSTHAM		% South American	
SSOTHER		Describe	

		Religious affiliation / orientation	Enter the percentage of people who state they are religious or affiliated with a religious community as indicated below. Use the description box at the bottom to enter a description and percentage of each religious affiliation that does not fit any of these categories. Please ensure that all the religious affiliation percentages add up to 100%.
PSRELCAT		% Catholic / Orthodox	
PSRELPRO		% Protestant	
PSRELJEW		% Jewish	
PSRELHIN		% Hindu	
PSRELBUD		% Buddhist	
PSRELISL		% Islam	
PSRELOTH		% Other	
PSRELNON		% No religious affiliation	
PSRELUNK		% Unknown	
SSRELDES		Description of other religious variables.	Religious data can be collected in a variety of different ways. Use this description box to enter religious data that describes religious characteristics other than religious affiliation. For example, if the question "do you consider yourself as being religious" is asked then you would record this question with the associated percentage responses.
SSSIZE		Sub sample size	Enter the size of the sub sample.
Effect size data			
SAMPLEN		Enter the sample size from which this ES was calculated	Sometimes the final sample size from which a particular ES is calculated varies slightly from the original sample size. Record the sample size (or estimated sample size) that relates to this ES. Be sure to select the appropriate confidence rating below if the exact sample size is not reported (e.g. often researchers will report that sample sizes associate with different combinations of variables varied within a certain range).
ESDATA		Type of data effect size based on	Instead of indicating the type of statistical analysis employed by the primary researchers, it is more important to indicate the type of data that was used for the effect size calculation. Select one of the following:
		1. means and standard deviations	Effectsize is directly calculated from means and SD's provided by the researchers.
		2. bivariate correlation / covariate matrix of Pearson's r	Bivariate Pearson's r is reported.
		3. frequencies or proportion	Effect size is estimated from frequency tables or proportions (Chi-square)
		4. t-value or F-value	
		5. chi-square (df=1)	
		Other	If none of the above applies, record the statistics you used to calculate the ES.
ESPAGE		5.3 Page number where the data for this effect size was found.	
		Calculated / reported effect	

		size	
CALESVAL		Enter the value of the calculated or reported ES (Pearson's r).	Reported ES's are standardized bivariate correlations between the two variables (i.e. Pearson's r). Use the ES Calculator to calculate ES's based on other statistics.
ESRELP		Record the reported p-value associate with this relationship.	Record the reported p-value (significance value) associated with the relationship between the two variables upon which the ES is based.
ESCONF		Confidence rating in effect size and Fisher's Zr computation	
	1. highly estimated		"Have N and crude p-value only, such as $p < .10$, and must reconstruct via rough t-test equivalence" (Lipsey & Wilson, 2001, p. 227).
	2. moderate estimation		"Have complex but relatively complete statistics, such as a multivariate ANOVA, as basis for estimation" (Lipsey & Wilson, p. 227).
	3. some estimation		"Have unconventional statistics which need to be standardized (e.g. equivalent t-values) or have conventional statistics but incomplete, such as exact p-level" or exact n for computing Fisher's Z(r) (Lipsey & Wilson, p. 227).
	4. slight estimation		"Must use significance testing statistics rather than descriptive statistics, but have complete statistics of conventional sort" (Lipsey & Wilson, p. 227).
	5. no estimation		The ES was either reported by the researchers as a Pearson correlation or you "have descriptive data such as means, standard deviations, frequencies, proportions, etc. from which an the effect size can be calculated directly" (Lipsey & Wilson, p. 227).
* several sections of this form were adapted from Lipsey, M. W., & Wilson, D. B. (2001). <i>Practical meta-analysis</i> . Thousand Oaks: Sage.			

Appendix D

Reliability Coefficients for Spirituality and Quality of Life Instruments

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
124-1	Coke, M. M. (1992)	Religiosity: Self-Rated Religiosity (Coke, 1992)	4	0.62	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	3 Average of test-retest scores across 2 to 10 weeks as reported in a review by Pavot and Diener (1993).	0.726
124-2	Coke, M. M. (1992)	Religiosity: Self-Rated Religiosity (Coke, 1992)	4	0.62	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	3 Average of test-retest scores across 2 to 10 weeks as reported in a review by Pavot and Diener (1993).	0.726
138-1	Cotton, S. P. et al. (1999)	FACIT-SP (Cella, 1997)	2	0.87	FACIT-B: Functional Assessment of Chronic Illness Therapy - Breast (Cella, 1997)	3 Cella et al. (1997).	0.85
142-1	Coward, D. D. (1991)	STS: Self-Transcendence Scale (Reed, 1991)	1	0.77	Emotional Well-being (Bradburn, 1969; Campbell et al., 1976)	3 Average the test-retest scores of the GWB (0.74 as reported by Coward, 1990) and the test-retest scores of the ABS (0.76 as reported by Bradburn (as cited in Bowling, 1997)).	0.75
145-1	Coward, D. D. (1996)	STS: Self-Transcendence Scale (Reed, 1991)	1	0.85	CWB: Cognitive Well-Being (Campbell et al., 1976)	1	0.74
147-1	Coward, D. D. (1998)	STS: Self-Transcendence Scale (Reed, 1991)	3 Average Cronbach's alpha from Coward (1991), Coward (1996) and Mellor et al. (1997)	0.83	CWB: Cognitive Well-Being (Campbell et al., 1976)	3 Test-retest reported by Coward (1996).	0.74

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
196-1	Fabricatore, A. N. et al. (2000)	Spiritual Life Integration (SLI) (subscale of Spiritual Involvement Scale) (Fenzel, 1996)	1	0.96	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	1	0.84
217-1	Ferrell, B. R. et al. (1998)	Spiritual subscale of QOL-BC (Ferrell et al., 1995).	2	0.71	Spiritual subscale of QOL-BC (Ferrell et al., 1995).	2	0.89
218-1	Ferrell, B. R. et al. (1995)	Spiritual Wellbeing Domain of QOL-CS (Ferrell et al., 1995)	1	0.9	FACT-G (Cella, 1993)	2	0.92
240-1	Fry, P. S. (2001)	Spirituality Health (subscale of "Domain Specific Efficacy") (Fry, 2001)	1	0.7	LSIA: Life Satisfaction Index (Neugarten et al., 1961)	1	0.74
240-2	Fry, P. S. (2001)	Spirituality Health (subscale of "Domain Specific Efficacy") (Fry, 2001)	1	0.7	LSIA: Life Satisfaction Index (Neugarten et al., 1961)	1	0.74
255-1	Genia, V., & Cooke, B. A. (1998)	SEI: Spiritual Experience Index (Genia, 1991)	2	0.82	Life Satisfaction Index- A (Neugarten et al., 1961)	3 Average split-half reliability coefficients reported in a review of the instrument by Bowling (1997).	0.695
316-1	Holland, J. C. et al. (1998)	SBI- System of Belief Inventory (Kash et al., 1995 ; Holland et al., 1998)	1	0.97	MOS: Medical Outcomes Study (Stewart et al., 1988)	2	0.845
317-1	Holland, J. C. et al. (1999)	SBI: System of Belief Inventory (Kash et al., 1995 ; Holland et al., 1998)	2	0.95	MOS: Medical Outcomes Study (Stewart et al., 1988)	2	0.76

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
320-1	Hong, S. M., & Giannakopoulos E. (1994)	Religiosity (Mol, 1970)	4	0.62	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	3 Average of test-retest scores across 2 to 10 weeks as reported in a review by Pavot and Diener (1993).	0.726
351-1	Kennedy, J. E. et al. (1998)	Spirituality Scale (Kennedy et al., 1998)	1	0.85	Well-being (Several items taken from Stewart & Ware, 1992)	1	0.89
37-1	Baetz, M. et al. (2002)	Duke Religion Index - Intrinsic Religiousness Subscale (Koenig et al., 1997)	1	0.87	SWLS - Satisfaction With Life Scale (Diener et al., 1985)	2	0.82
394-1	Landis, B. J. (1996)	SWBS: Spiritual Well-Being Scale (Paloutzian & Ellison, 1982 ; Ellison, 1983)	1	0.96	Psychosocial Adjustment to Illness Scale. Self Report (PAIS-SR) (Derogatis, 1986)	1	0.96
406-2	Levin, J. S., & Taylor R. J. (1998)	Subjective Religiosity (Jackson, 1991)	4	0.62	Life Satisfaction (Jackson, 1991)	4	0.56
439-1	Maton, K. I., & Zimmerman, M. A. (1992)	Spirituality (Maton, 1989)	1	0.8	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	1	0.81
462-1	Mellors, M. P. et al. (1997)	STS: Self-Transcendence Scale (Reed, 1991)	1	0.86	QLI: Quality of Life Index (Ferrans & Powers, 1992)	1	0.84
491-1	Neill, C. M., & Kahn, A. S. (1999)	Intrinsic Religious Motivation Scale (Hoge, 1972)	1	0.91	Life Satisfaction Index Z (Wood et al., 1969)	1	0.73
515-1	Peacock, J. R., & Poloma, M. M. (1999)	Prayer Experience Scale (Poloma & Pendleton, 1991)	1	0.88	Life Satisfaction (Peacock & Poloma, 1999)	4	0.56
52-1	Beery, T. A. et al. (2002)	SWBS: Spiritual Well-Being Scale (Ellison, 1983)	2	0.84	IWB: Index of Well-Being (Campbell et al., 1976 as cited in Beery et al., 2002)	1	0.96

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
521-1	Pfeifer, S., & Waelty, U. (1995)	Religious Commitment (Pfeifer & Waelty, 1995)	5	0.84	Life satisfaction (Pfeifer & Waelty, 1995)	5	0.82
521-2	Pfeifer, S. & Waelty U. (1995)	Religious Commitment (Pfeifer & Waelty 1995)	5	0.84	Life satisfaction (Pfeifer & Waelty, 1995)	5	0.82
522-1	Pippalla, R. S. & Chaar, M. D. C. (2001)	WHOQOL-100 Spiritual Domain (WHOQOL Group, 1998)	3 Bonomi et al. (2000)	0.87	WHOQOL-100 - Overall quality of life and general health subscale (WHOQOL Group, 1998)	3 The WHOQOL Group (1998)	0.84
527-1	Poloma, M. M., & Pendleton B. F. (1991)	Prayer Experience Scale (Poloma & Pendleton, 1991)	3 Poloma & Pendleton (1991)	0.87	Life Satisfaction (Poloma & Pendleton, 1990)	3 Poloma & Pendleton (1990)	0.91
54-1	Benight, C. C. et al. (2001)	SWBS: Spiritual Well-being Scale (Paloutzian & Ellison, 1982)	1	0.71	SF-36: Health Status Questionnaire, Short Form 36 (Ware & Sherbourne, 1992)	1	0.86
555-2	Ringdal, G. I. (1996)	Religiosity (Ringdal, 1996)	1	0.71	EORTC QOL-C30: Cancer Quality of Life Questionnaire (Aaronson et al., 1993)	3 Cronbach alpha reported by Ringdal and Ringdal (1993).	0.82
61-1	Bienenfeld, D. et al. (1997)	Religious Commitment (Adapted from Kauffman, 1979)	4	0.62	LSIA: Life Satisfaction Index A (Neugarten et al., 1961)	3 Average split-half reliability coefficients reported in a review of the instrument by Bowling (1997).	0.695
616-1	Simmons, Z. et al. (2000)	Idler Index of Religiosity (Idler, 1987)	2	0.67	MQOL-SIS: McGill Quality of Life - Single Item Scale (Cohen et al., 1995)	4	0.56
656-9	Tate, D. G. et al. (1997)	FACT-SP: Functional Assessment of Cancer Therapy - Spiritual Well-Being (Cella et al., 1993)	2	0.88	FACT: Functional Assessment of Cancer Therapy (Cella et al., 1993)	2	0.92

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
665-1	Thomas, M. E., & Holmes, B. J. (1992)	Religion	1	.628 2	A domain satisfaction scale (Original data from: Campbell et al., 1976)	1	0.812 4
665-2	Thomas, M. E., & Holmes, B. J. (1992)	Religion (Campbell, 1976)	1	.628 2	A domain satisfaction scale (Original data from: Campbell et al., 1976)	1	0.812 4
675-1	Tuck., I. et al (2001)	SHI: Spiritual Health Inventory (Highfield, 1992)	3 Average Cronbach's alpha from Highfield (1992)	0.83	FAHI: Functional Assessment of HIV Infection (Cella et al., 1993)	2	0.89
69-1	Bonomi, A. E. et al. (2000)	WHOQOL - Spirituality Subscale (The WHOQOL Group, 1998)	1	0.86	SF-36 (Ware & Sherbourne, 1992)	3 Average test-retest taken from McHorney et al. (1993).	0.82
703-1	Westlake, C. et al. (2002)	SPS: Spiritual Perspective Scale (Reed, 1986)	2	0.92	MOS-SF 36 (Ware & Sherbourne, 1992)	3 Average test-retest taken from McHorney et al. (1993).	0.82
724-1	Wyatt, G. et al. (1996)	LTQL-SP/PHIL: Long-term Quality of Life Scale - Spiritual / Philosophical subscale (Wyatt et al., 1996)	1	0.87	CARES (Schag & Heinrich, 1990)	3 The overall test-retest reliability across one week as reported by Schag et al. (1990).	0.91
73-2	Brady, M. J. et al. (1999)	FACIT-SP: Functional Assessment of Chronic Illness Therapy - Spiritual Well-Being (Peterman et al., 2002)	1	0.87	FACT-G QOL-SIS: Contentment with QOL (Cella et al., 1993)	4	0.56
739-1	McQuaide, S. (1998)	Spirituality (McQuaide, 1998)	4	0.62	Overall Well-Being (McQuaide, 1998)	5	0.82

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
89-3	Byrd, K. R. et al. (2000)	ROI-IR: Religious Orientation Inventory - Intrinsic Religiosity Subscale (Allport & Ross, 1967)	1	0.86	Satisfaction with Life (Poloma & Pendleton, 1990)	1	0.84
63-1	Blaine, B., & Crocker, J. (1995)	Religiosity Salience-Cognition scale (King & Hunt, 1975)	1	0.94	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	1	0.76
223-1	Fitchett, G. et al. (1999)	R-Cope - positive subscale (Pargament et al., 1998)	1	0.82	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	3 Average of test-retest scores across 2 to 10 weeks as reported in a review by Pavot and Diener (1993).	0.726
359-1	Kim, J. et al. (2000)	SWBS: Spiritual Well-Being Scale (Paloutzian & Ellison, 1982)	2	0.84	LSQ: Life Satisfaction Questionnaire (Fugl-Meyer et al., 1991)	2	0.94
460-1	Meller, S. (2001)	Religious Faith (Meller, 2001)	1	0.89	Satisfaction with Life	4	0.56
554-1	Riley, B. B. et al. (1998)	SWBS (Spiritual Well-Being Scale) and FACT-SP combined (Paloutzian & Ellison, 1982; Cella et al., 1993)	3 Average Cronbach's alpha reported in psychometric evaluations by Bufford, Paloutzian, & Ellison (1991) and Cella et al. (1993).	0.86	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
569-1	Ruffing Rahal, M. A. (1991)	SWBS: Spiritual Well-Being Scale (Paloutzian & Ellison, 1982)	3	0.84	II-II: Integration Inventory II (Ruffing Rhahal, 1991)	1	0.91

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
770-1	Lewis, C. A., et al. (1997)	Francis Scale of Attitudes towards Christianity - FSAC (Francis & Stubbs, 1987)	3 Average Cronbach's alphas reported by Lewis & Maltby (1995) and Francis (as cited in Aycock, 1999)	0.97	SWLS Satisfaction With Life Scale -(Diener et al., 1985)	3 Average of test-retest scores across 2 to 10 weeks as reported in a review by Pavot and Diener (1993).	0.726
761-1	WHOQOL Group (1998)	WHOQOL-100 Spiritual Domain: Spirituality/ Religion/ Personal Beliefs (WHOQOL Group, 1998)	1	0.85	Overall Quality Of Life (and general health perceptions) domain of the WHOQOL (WHOQOL Group, 1998)	1	0.84
263-1	Gioiella, M. E., et al. (1998)	SWBS: Spiritual Well-Being Scale (Paloutzian & Ellison, 1982)	3 Bufford, Paloutzian, & Ellison (1991)	0.89	FLIC: Functional Living Index: Cancer (Schipper et al., 1984)	3 Conner-Spady et al. (2001) referred to a review of the FLIC by Clinch stating that "reliability coefficients have ranged from 0.68 to 0.89 using Cronbach's alpha and test-retest reliability at 1-month intervals" (p. 480). An average reliability coefficient of 0.785 has been entered.	0.785
794-1	Pargament, K. I. (1999)	R-COPE (Brief): Positive Religious Coping Subscale (Pargament, 1999)	1	0.87	Quality of Life (Spitzer et al., 1981)	1	0.72
169-8	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82

ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
169-7	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
169-6	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
169-5	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
169-4	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
169-3	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
169-2	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
169-1	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974)	2	0.85	SWLS: Satisfaction with Life Scale (Diener et al., 1985)	2	0.82
ID	Study	Spirituality Instrument	Rel. Source ¹	Rel. Coef	QOL Instrument	Rel. Source ¹	Rel. Coef.
529-1	Post-White, J. et al. (1996)	Spirituality Index (Post-White, et al., 1996)	1	0.78	Quality of Life tool (QOL tool) (Nelson et al., 1990)	1	0.66

¹The reliability source was coded as follows:

- 1: Reliability data provided by the authors of the present study.
- 2: Reliability data from secondary sources as cited by the authors of the present study.
- 3: Reliability data from other sources as indicated in the next column.
- 4: No reliability data could be found for these single-item instruments. They received the lowest reliability scores of the sample of instruments.
- 5: No reliability data could be found for these instruments. Multi-item instruments received an average reliability score.

Appendix E

Description and Classification of Operational Definitions of Spirituality

ID	Study	Instrument	Description
Instruments explicitly representing existential and relational domains			
491-1	Neill, C. M. & Kahn, A. S. (1999)	Intrinsic Religious Motivation Scale (Hoge, 1972)	This scale is based on ten items from the Intrinsic Religious Motivation Scale (Hoge, 1972). Sample items include: "My faith involves all of my life" (intrinsic), "One should seek God's guidance when making every important decision" (intrinsic), "In my life I experience the presence of the Divine" (Intrinsic), "My religious beliefs are what really lie behind my whole approach to life" (Intrinsic), "It doesn't matter so much what I believe as long as I lead a moral life" (Extrinsic), "Although I believe in my religion, I feel there are many more important things" (Extrinsic) (Basset, 1999, p. 137). The scale ranges from 0 to 40 with higher values indicating increased spirituality.
89-3	Byrd, K. R., Lear, D., & Schwenka, S. (2000)	ROI-IR: Religious Orientation Inventory - Intrinsic Religiosity Subscale (Allport & Ross, 1967)	Representative items of the Intrinsic Religiosity Subscale include: "It is important for me to spend periods of time in private religious thoughts and mediation", "I try hard to carry my religion over into all my other dealings in life", "Quite often I have been keenly aware of the presence of God or the Divine Being", "My religious beliefs are really what lie behind my whole approach to life", "Religion is especially important because it answers many questions about the meaning of life" (Burris, 1999, pp. 144-153).
724-1	Wyatt, G. et al. (1996)	LTQL-SPiR/PHIL: Long-term Quality of Life Scale - Spiritual / Philosophical subscale (Wyatt et al., 1996)	The philosophical / spiritual view of life factor consists of 12 items which have been labeled by the authors as follows: guiding energy, inner direction helps me, follow inner voice, receive subtle cues, (5) appreciate time with family / friends, I have intuitive experiences that reassure me about health care choices, notice things in nature, become closer to family / friends, don't take things for granted, better idea about serious illness, sympathetic with major illness (pp. 161-162). All items are rated on a five-point Likert-type scale (not at all - very much). The overall QOL scale was specifically designed to include the spiritual domain. However, several confounding items associated with social and psychological domains are also represented.
316-1	Holland, J. C. et al. (1998)	SBI- System of Belief Inventory (Holland et al., 1998)	This scale consists of 54 items representing four domains which are described by the authors as follows: "(1) the degree to which persons felt that they derived meaning from an existential perspective (i.e. ethereal, of an immaterial nature, or a sense of meaning of life); (2) the use of certain religious practices and rituals, such as meditation and prayer; (3) the relationship to a superior being or a perceived higher power, such as God; and, (4) the level of social support derived from a community of individuals sharing similar beliefs" (p. 462). Items were rated on a 4-point Likert-type scale.
317-1	Holland, J. C. et al. (1999)		

ID	Study	Instrument	Description
Instruments explicitly representing existential and relational domains (cont'd)			
255-1	Genia, V. & Cooke, B. A. (1998)	SEI: Spiritual Experience Index (Genia, 1991).	This scale consists of 38 items developed with the intent to measure "the degree of spiritual maturity" independent from particular spiritual beliefs or religious backgrounds (Genia, 1991, p. 339). Sample items include: "I often feel closely related to a power greater than myself", "My faith gives my life meaning and purpose", "I gain spiritual strength by trusting in a higher power", and "I feel a strong spiritual bond with all of humanity" (pp. 344-345). The scale also includes several reversed items such as: "I believe that there is only one true faith", and "I never challenge the teachings of my faith" (p. 345).
675-1	Tuck, I., McCain N. L., & Elswick, R. K. (1992)	SHI: Spiritual Health Inventory (Highfield, 1992).	This scale consists of 31 items representing feelings and behaviors associated with personal spirituality. Sample items include: "I believe God can help me", "I feel this illness has no purpose or meaning", "I have both strong negative and positive feelings toward my faith", "I worry about life after death", "My life has a purpose", and "I believe God accepts me even with my faults" (p. 3).
240-1	Fry, P. S. (2001)	Spirituality Health - subscale of the Domain Specific Efficacy Scale (Fry, 2001).	This scale consists of "four items reflecting spiritual health efficacy beliefs relating to one's perceived ability to generate spiritually based faith and inner strength, for example: "Even in hard times, the meaning and purpose I have found for my life provides me a sense of peace and harmony", "Even in a time of struggle I can go to a spiritual dimension within myself for guidance" (Fry, 2001, p. 798). Items are scored on a four-point Likert-type scale (strongly agree - strongly disagree).
240-2	Fry, P. S. (2001)		
351-1	Kennedy, J. E. et al. (1998)	Spirituality Scale (Kennedy et al. 1998)	This scale was designed to measure change in spirituality in relation to a particular event. Though it does not measure the degree of spirituality at a particular point in time, it is correlated with change of wellbeing. The items on the spirituality scale are: "My belief that there is a divine plan for the world has . . .", "My search for spiritual meaning has . . .", "My desire to understand events in spiritual terms has . . .", "My belief that it is important to follow a spiritual path has . . .", "My tendency to base my actions on guidance from a higher power has . . ." (Kennedy et al., 1998, p. 327). Participants were asked: "How much each of these statements describes your experience since the crime."
703-1	Westlake, C. et al. (2002)	SPS: Spiritual Perspective Scale (Reed, 1986).	This instrument consists of "10 items that measure the extent to which spirituality permeates one's life and one engages in spirituality related interactions" (Reed, 1986, p. 87). Each item is rated on a Likert-type scale from one to six. Sample items include "Having a spiritual outlook gives my life meaning and purpose", "God, or the Supreme Being, is very real to me", and "I consider myself to be a spiritual person" (Reed, 1982, p. 75).

ID	Study	Instrument	Description
Instruments explicitly representing existential and relational domains (cont'd)			
394-1	Landis, B. J. (1996)	SWBS: Spiritual Well-Being Scale	This instrument consists of two subscales that the authors labeled "existential" and "religious". The existential subscale consists of 10 items associated with meaning and purpose in life, and the religious subscale consists of ten items that represent one's relationship to a divine being. Representative items of the two subscales include: "I don't know who I am, where I came from, or where I am going", "I feel a sense of well-being about the direction my life is headed in", "I believe there is some real purpose for my life", "I believe that God loves me and cares about me", "I have a personally meaningful relationship with God", and "My relationship with God contributes to my sense of well-being" (as cited in Boivin, Kirby, Underwood & Silva, 1999, p. 385). The scale also includes nine items that are reverse scored. Each item is scored on a six point Likert-type scale.
52-1	Beery, T. A. et al. (2002)	(Paloutzian & Ellison, 1982).	
54-1	Benight, C. C. et al. (2001)		
359-1	Kim, J. et al. (2000)		
569-1	Ruffing Rahal, M. A. (1991)		
63-1	Blain, B., & Crocker, J. (1995)	Religiosity Salience-Cognition Scale (King & Hunt, 1975)	The following five items derived from the Religiosity Salience-Cognition scale were used in this study: "My religious beliefs are what lie behind my whole approach to life", "My religious beliefs provide meaning and purpose to life", "I am frequently aware of God in a personal way", "I allow my religious beliefs to influence other areas of my life", and "Being a religious person is important to me" (Blain & Crocker, 1995, p. 1034). Items were rated on a Likert-type scale ranging from one to seven.
223-1	Fitchett, G. et al. (1999)	R-Cope - Positive Subscale (Pargament, 1998)	The R-Cope consists of 21 items divided in two subscales: positive and negative religious coping. The scales measure behaviors, practices and feelings indicative of religious coping. The positive religious coping items clearly reflect the existential and relational attributes of spirituality. The negative religious coping items are more ambiguous and seem to reflect the author's biases more than that these items present a measure of the degree of spirituality. The Positive Subscale was therefore selected as the primary measure for the purposes of this meta-analysis. Representative items include: "Thought about how my life is part of a larger spiritual force", "Worked together with God as partners to get through this hard time", "Looked to God for strength, support, and guidance in this crisis", "Tried to find the lesson from God in this crisis", and "Asked God to help me find a new purpose in living" (Pargament, 1999, p. 717).
794-1	Pargament, K. I. et al. (1999)	R-COPE (Brief): Positive Religious Coping Subscale (Pargament, 1999).	

ID	Study	Instrument	Description
Instruments explicitly representing existential and relational domains (cont'd)			
554-1	Riley, B. B. et al. (1998)	SWBS (Paloutzian & Ellison, 1982) and FACT-SP (Cella, 1997) combined.	This study is unique in that it used two scales to assign people to one of three groups derived from a factor analysis. The authors labeled these three groups as "religious", "existential", and "nonspiritual" (p. 263). The religious group was scored high on items associated with one's relationship with God and saw their lives as both "fulfilling and meaningful" (p. 263). The existential cluster scored the lowest on items associated with one's relationship with God or prayer but scored highest on items pertaining to meaning and purpose in life. The non-spiritual cluster scored lowest on all items. The correlation for this meta-analysis is based on the difference in quality of life scores between the "religious" and "nonspiritual" groups.
Spirituality instruments that are primarily existential in nature			
138-1	Cotton, S. P. et al. (1999)	FACIT-SP: Functional	This scale consists of eight items measuring meaning and purpose in life and four items pertaining to one's faith and spiritual beliefs. Representative items include: "I have a reason for living", "I am able to reach down deep into myself for comfort", and "I have trouble feeling peace of mind" (Peterman et al., 2002, p. 49). Both the FACIT-SP and the FACT-SP contain the same items pertaining to spirituality.
656-9	Tate, D. G. et al. (1997)	Assessment of Chronic Illness Therapy - Spiritual	
73-2	Brady, M. J. et al. (1999)	Subscale (Cella, 1997) and FACT-SP: Functional Assessment of Cancer Therapy - Spiritual Subscale	
217-1	Ferrell, B. R. et al. (1998)	Spiritual Subscale of QOL-BC	The spiritual subscale of the QOL-BC consists of seven items pertaining to hope, purpose in life, positive change, spiritual change, importance of religious and spiritual practices, and uncertainty. Each item is rated on a scale from 0 to 10 (10 representing the best outcome).
218-1	Ferrell, B. R. et al. (1995)	(Ferrell, et al. 1995).	
145-1	Coward, D. D. (1996)	STS: Self-Transcendence	This 15-item scale is based on Reed's (1991) Theory of Self-Transcendence. The items relate to "intrapersonal, interpersonal, and temporal experiences characteristic of later life that reflect expanded boundaries of the self" (p. 6). Sample items cited by Mellors (1997) include: "At this time in my life, I see myself as: sharing my wisdom and experiences with others; letting others help me when I may need it; putting aside some things that I once thought were important; and finding meaning in my spiritual beliefs" (p. 63). The scale is rated on a 4-point Likert scale from "not at all" to "very much".
147-1	Coward, D. D. (1998)	Scale (Reed, 1991).	
462-1	Mellors, M. P. et al. (1997)		

ID	Study	Instrument	Description
Spirituality instruments that are primarily existential in nature (cont'd)			
522-1	Pippalla, R. S. & Chaar, M. D. C. (2001)	WHOQOL - Spirituality Subscale	The entire scale is presented as follows: "The following few questions are concerned with your personal beliefs, and how these affect your quality of life. These questions refer to religion, spirituality and any other beliefs you may hold. 1. Do your personal beliefs give meaning to your life? 2. To what extent do you feel your life to be meaningful? 3. To what extent do your personal beliefs give you the strength to face difficulties? 4. To what extent do your personal beliefs help you to understand difficulties in life?" (WHOQOL-group, 1998, p. 1578). Each item is rated on a scale from 1 (not at all) to 5 (extreme amount).
69-1	Bonomi, A. E. et al. (1999)	(WHOQOL Group, 1998).	
761-1	WHOQOL Group (1998)		
460-1	Meller, S. (2001)	Religious Faith (Meller, 2001).	Eight religiously oriented questions which were considered to be important to the well-being of caregivers. "The examinees were asked to mark 'believe', 'doubt', or 'don't believe' next to eight statements concerning belief in God, in supreme world-conducting power, in personal divine supervision, and so forth. The indicator was the examinee's mean answer" (Meller, 2001, p. 62). Note that though this may reflect a Likert-type scale, the range is very constricted and predefined. No other information about this scale is provided. Reference is made to a non-English article where, presumably, the scale might be further discussed. This scale is coded as a primarily existential measure of spirituality because of its focus on belief. The relational or transcendent dimension is, however, represented by items pertaining to one's relationship to the divine (e.g., divine supervision).
Spirituality instruments that are primarily relational / transcendent in nature			
37-1	Beatz, M. et al. (2002)	Duke Religion Index - Intrinsic Religiousness Subscale (Koenig, Parkerson & Meador KG, 1997).	The intrinsic religiousness was measured by the following three items derived from the Duke Religion Index: "In my life I experience the divine", "religious beliefs are my whole approach to life", and "I try hard to carry religion into all dealings in life" (Beatz et al. (2002, p. 161). Items were rated on a five point Likert-type scale.
515-1	Peacock, J. R. et al. (1999)	Prayer Experience Scale (Poloma & Pendleton, 1991).	The items of this scale are cited by Peacock (1999) as follows: "How often have you experienced the following during prayer ... never, once or twice, occasionally, or regularly? Felt divinely inspired or led by God to perform some specific action; received what you believed to be a deeper insight into a spiritual or biblical truth, received what you regarded as a definite answer to a specific prayer request, felt the strong presence of God, experienced as deep sense of peace and well-being" (p. 328).
527-1	Poloma, M. M. (1991)		

ID	Study	Instrument	Description
Spirituality instruments that are primarily relational / transcendent in nature (cont'd)			
196-1	Fabricatore, A. N. et al. (2000)	SLI: Spiritual Life Integration - subscale of Spiritual Involvement Scale (Fenzel, 1996).	This study is based on a 12-item subscale of the Spiritual Involvement Scale. A representative item cited in this study includes "I feel close to God" (Fabricatore et al., 2000, p. 223). According to the researchers higher scores are indicative of "greater integration of individuals' spirituality into their lives" (p. 223). Items are rated on a five-point Likert-type scale. No further information has been published about this scale.
439-1	Maton, K. I. and Zimmerman, M. A. (1992)	Spirituality (Maton, 1989)	The following three items measured on a 5-point Likert-type scale were used to operationalize spirituality: (1) "I experience a personal close relationship with God", (2) "I experience God's love and caring on a regular basis", (3) "My religious faith helps me cope during times of difficulty" (Maton & Zimmerman, 1992, p. 84).
169-1 to 199-8	Dorahy, M. J. et al. (1998)	Theism scale (Maranell, 1974).	The Theism Scale is a subscale of the Religious Attitudes Scale (Maranell, 1974) consisting of 12 items rated on a 5-point Likert-type scale. Representative items include: "God is always watching over us", "God is a divine spirit guiding my life", and "I do not feel that a belief in God is necessary" (reverse scored) (as cited in Wulff, 1999).
770-1	Lewis, C. A. et al. (1997)	FSAC: Francis Scale of Attitudes towards Christianity (Francis & Stubbs, 1987).	This scale explicitly pertains to Christian beliefs and values which constrains its generalizability. However, when applied to a Christian sample the scale is seen as a measure of spirituality representing primarily relational items but also items with existential connotations. Representative items include: "God is very real to me", "The idea of God means much to me", and "I find it hard to believe in God" (as cited in Aycock, 1999, p. 82). Items are measured on a 5-point Likert-type scale. Several items are reverse coded.
Ambiguous measures of spirituality / religiosity			
616-1	Simmons, Z. et al. (200)	Idler Index of Religiosity (Idler, 1987).	This instrument consists of four items measuring two dimensions of religiosity: organizational and subjective religiosity (also termed public and private religiosity). Idler (1987) described the items as follows: "An index of public religiousness was created from two items: attendance at religious services (6 levels) and the number of other congregation members (4 levels) known to the respondent An index of private religiousness was created from two items tapping subjective religious experiences, 'Aside from attendance at religious services, do you consider yourself to be ... (1) deeply religious (2) fairly religious (3) only slightly religious or (4) not at all religious' and 'How much is religion a source of strength and comfort to you? (1) none (2) a little (3) a great deal'". Codings for the first items were reversed, both were standardized and summed" (p. 231).

ID	Study	Instrument	Description
Ambiguous measures of spirituality / religiosity (cont'd)			
665-1	Thomas, M. E., and Holmes, B. J. (1992)	Religion (Campbell et al., 1976).	This scale is based on the average score of two items pertaining to attendance of religious services and degree of religious mindedness on a five-point Likert-type scale.
665-2	Thomas, M. E., and Holmes, B. J. (1992)		
320-1	Hong, S. M. and Giannakopoulos, E. (1994)	Religiosity (Mol, 1970).	A five-item religiosity scale rated from "strongly disagree" to "strongly agree" on a five-point Likert-type scale. The authors do not provide any further information about this scale. The source of the scale to which the authors refer was reviewed, but no further information was presented in the original source (Mol, 1970). The authors' conceptualization of religiosity is as follows: "Religiosity refers to the importance of religion, the interest in religion, or religion mindedness (Chamberlain & Zika, 1988)" (p. 548). Based on this description it was assumed that the scale could be interpreted as an ambiguous measure of spirituality.
555-2	Ringdal, G. I. (1996)	Religiosity (Ringdal, 1996).	"In this study the author developed the following questions that may be seen as rough indicators of religious disposition: What can you tell about your religious beliefs? (I believe in God, I do not believe in God, I do not know). Have your religious beliefs been of support to you after you became ill from cancer? (Very good support, Some support, No support)" (p. 201).
124-1	Coke, M. M. (1992)	Religiosity: Self-Rated Religiosity	Respondents were asked to rate their agreement with the statement "I am a religious individual" on a 5-point Likert Scale (p. 318).
124-2	Coke, M. M. (1992)	(Coke, 1992).	
521-1	Pfeifer, S. and Waelty, U. (1995)	Religious Commitment	The authors describe their scale as follows: "The level of religiosity was computed from 15 items which were weighted for their significance in expressing religiosity. Nine items that reflect broad social conventions (i.e. 'I am a religious person', 'I believe in God') were given 1 point each. Five items were multiplied by a factor of 2 as they showed more specific aspects of high religiosity and reflected special religious practices (e.g. regular church attendance, praying before eating, consultation of a Christian counselor, belief in demonic causes of disease)" (p. 71). Total scores were between 0 and 20.
521-2	Pfeifer, S. and Waelty, U. (1995)	(Pfeifer & Waelty, 1995).	

ID	Study	Instrument	Description
Ambiguous measures of spirituality / religiosity			
739-1	McQuaide, S. (1998)	Spirituality (McQuade, 1998).	This scale was based on a factor analysis of 122 statements about women's wellbeing at midlife designed by the author. The spiritual component of this scale was described as follows: "additional scales measured such issues as the women's sense of their own spirituality ..." (p. 24). No further information was provided.
406-2	Levin, J. S. and Taylor R. J. (1998)	Subjective Religiosity (Neighbors & Jackson, 1984).	This scale is based on a single item response to the question, "How religious would you say you are?" rated on a four-point Likert-type scale (Levin & Taylor, 1998, p. 702).
292-1	Harvey, C. D. H. et al. (1991)	Religiosity (Harvey et al., 1991).	Level of religiosity was measured by one item on a scale from "very religious" to "not at all religious".
61-1	Bienenfeld, D. et al. (1997)	Religious Commitment (Adapted from Kauffman, 1979).	The researchers state that "questions covered both the behavioral and experiential dimensions of religious commitment, including frequency of private prayer and perceived closeness to God" (p. 47).

Appendix F

Mean Effect Size Analyses of the Relationship between Spirituality and Quality of Life

Mean Effect Size Adjusted for Instrument Reliability¹

Model	Fisher's z(r)	SE	Pearson's r	- 95% CI	+ 95% CI	P
Fixed	.3202	.0087	.3097	.2942	.3250	.0000
Random	.3556	.0333	.3413	.2824	.3977	.0000

Mean Effect Size Not Adjusted for Instrument Reliability²

Model	Fisher's z(r)	SE	Pearson's r	- 95% CI	+ 95% CI	P
Fixed	.2383	.0067	.2338	.2214	.2463	.0000
Random	.2619	.0230	.2561	.2135	.2976	.0000

Mean Effect Size Prior to Removal of Outliers³

Model	Fisher's z(r)	SE	Pearson's r	- 95% CI	+ 95% CI	P
Fixed	.2406	.0067	.2361	.2237	.2390	.0000
Random	.2677	.0237	.2615	.2177	.3043	.0000

¹ N = 59 studies. Homogeneity Analysis: Total Q = 635.2891, df = 58, p = .0000

² N = 59 studies. Homogeneity Analysis: Total Q = 476.0618, df = 58, p = .0000

³ N = 62 studies. Homogeneity Analysis: Total Q = 536.4616, df = 61, p = .0000

Appendix G

Comprehensive Mixed Effects Model¹

Descriptives

Mean ES - Fisher's Z(r)	R-Square	N
.3555	.3242	59

Regression Coefficients

Variable	b	SE	- 95% CI	+ 95% CI	p	Beta
Constant	.0778	.0896	-.0978	.2534	.3852	.0000
Spirit. Dummy Var.1 ²	.1801	.1043	-.0243	.3845	.0841	.2231
Spirit Dummy Var. 2	-.2586	.1022	-.4589	-.0583	.0114	-.3413
Spirit Dummy Var. 3	-.0309	.1078	-.2422	.1805	.7747	-.0388
QOL Dummy Var. 1 ³	.3011	.1435	.0199	.5823	.0358	.2933
QOL Dummy Var. 2	.3976	.1012	.1992	.5960	.0001	.5467
Sampling Method ⁴	-.0536	.1086	-.2666	.1593	.6215	-.0585
Age Group ⁵	-.0445	.1029	-.2462	.1571	.6652	-.0494
Ethnic Black ⁶	.1371	.1391	-.1356	.4098	.3244	.1211
Ethnic White ⁷	-.2290	.1595	-.5417	.0837	.1511	-.1650

Homogeneity Analysis

	Q-value	df	P
Model	30.4568	9	.0004
Residual	63.4813	49	.0800
Total	93.9381	58	.0020

¹ Based on effect sizes adjusted for instrument reliability.

² Spirituality Instruments

Comparison group: Instruments containing explicit existential and relational items.

Spirit. Dummy Var. 1: Instruments primarily existential in nature

Spirit. Dummy Var. 2: Instrument primarily relational / transcendent in nature

Spirit. Dummy Var. 3: Ambiguous spirituality instruments

³ Quality of Life Instruments

Comparison group: General multidimensional instruments

QOL Dummy Var. 1: Disease specific multidimensional instruments

QOL Dummy Var. 2: All other instruments (general QOL, wellbeing, & life-satisfaction)

⁴ Probability compared to non-probability sampling.

⁵ Samples of people older than 55 compared to other samples.

⁶ Exclusively "Black" or African-American samples compared to other samples.

⁷ Exclusively "White" or Caucasian samples compared to other samples.

Appendix H

Trimmed Mixed Effects Model¹

Descriptives

Mean ES - Fisher's Z(r)	R-Square	N
.3562	.2719	59

Regression Coefficients

Variable	b	SE	- 95% CI	+ 95% CI	p	Beta
Constant	.0656	.0912	-.1132	.2443	.4723	.0000
Spirit. Dummy Var.1 ²	.1975	.1062	-.0106	.4057	.0629	.2438
Spirit Dummy Var. 2	-.2227	.0975	-.4139	-.0315	.0224	-.2931
Spirit Dummy Var. 3	.0080	.0985	-.1851	.2012	.9350	.0101
QOL Dummy Var. 1 ³	.2793	.1476	-.0100	.5687	.0585	.2706
QOL Dummy Var. 2	.3705	.1032	.1683	.5727	.0003	.5074

Homogeneity Analysis

	Q-value	df	P
Model	23.5982	5	.0003
Residual	63.1765	53	.1598

¹ Based on effect sizes adjusted for instrument reliability.

² Spirituality Instruments

Comparison group: Instruments containing explicit existential and relational items.

Spirit. Dummy Var. 1: Instruments primarily existential in nature

Spirit. Dummy Var. 2: Instrument primarily relational / transcendent in nature

Spirit. Dummy Var. 3: Ambiguous spirituality instruments

³ Quality of Life Instruments

Comparison group: General multidimensional instruments

QOL Dummy Var. 1: Disease specific multidimensional instruments

QOL Dummy Var. 2: Collapsed category of all other instruments (general QOL, wellbeing, & life-satisfaction)

Appendix I

Trimmed Mixed Effects Model *Not* Adjusted for Instrument ReliabilityDescriptives

Mean ES - Fisher's Z(r)	R-Square	N
.2613	.3549	59

Regression Coefficients

Variable	b	SE	- 95% CI	+ 95% CI	p	Beta
Constant	-.1801	.3320	-.8308	.4707	.5876	.0000
Spirit. Dummy Var.1 ¹	.1424	.0706	-.0040	.2808	.0437	.2687
Spirit Dummy Var. 2	-.1569	.0641	-.2826	-.0312	.0144	-.3115
Spirit Dummy Var. 3	-.0666	.0852	-.2337	.1004	.4342	-.1311
QOL Dummy Var. 1 ²	.2140	.0992	.0197	.4084	.0309	.3175
QOL Dummy Var. 2	.2864	.0675	.1540	.4187	.0000	.5976
SPREL ³	-.0707	.3198	-.6975	.5561	.8250	-.0363
QLREL ⁴	.3638	.2440	-.1144	.8420	.1360	.1796

Homogeneity Analysis

	Q-value	df	P
Model	32.8996	7	.0000
Residual	59.8078	51	.1863
Total	92.7073	58	.0026

¹ Spirituality Instruments

Comparison group: Instruments containing explicit existential and relational items.

Spirit. Dummy Var. 1: Instruments primarily existential in nature

Spirit. Dummy Var. 2: Instrument primarily relational / transcendent in nature

Spirit. Dummy Var. 3: Ambiguous spirituality instruments

² Quality of Life Instruments

Comparison group: General multidimensional instruments

QOL Dummy Var. 1: Disease specific multidimensional instruments

QOL Dummy Var. 2: Collapsed category of all other instruments (general QOL, wellbeing, & life-satisfaction)

³ Reliability coefficients of spirituality instruments.⁴ Reliability coefficients of quality of life instruments.