8-1-2017

The Prevalence of Religious Service Attendance in America: A Review and Meta-Analysis

Chad S. Briggs
Southern Illinois University Carbondale, chad.briggs@judsonu.edu

THE PREVALENCE OF RELIGIOUS SERVICE ATTENDANCE IN AMERICA:
A REVIEW AND META-ANALYSIS

by

Chad S. Briggs

B.A., Wartburg College, 1997

M. A., Southern Illinois University Carbondale, 2001

A Dissertation

Submitted in Partial Fulfillment of the Requirements for the

Doctor of Philosophy Degree

Department of Psychology

in the Graduate School

Southern Illinois University Carbondale

August 2017

DISSERTATION APPROVAL

THE PREVALENCE OF RELIGIOUS SERVICE ATTENDANCE IN AMERICA: A REVIEW AND META-ANALYSIS

by

Chad S. Briggs

A Dissertation

Submitted in Partial Fulfillment of the Requirements for the

Doctor of Philosophy Degree

in the field of Psychology

with a concentration in Applied Psychology

Approved by:

Meera Komarraju, Ph.D., Chair

Eric Jacobs, Ph.D.

Paul Etcheverry, Ph.D.

Steve Karau, Ph.D.

Graduate School

Southern Illinois University Carbondale

June 30, 2017

AN ABSTRACT OF THE DISSERTATION OF

CHAD S. BRIGGS, for the Doctor of Philosophy degree in APPLIED PSYCHOLOGY, presented on June 30, 2017, at Southern Illinois University Carbondale.

TITLE: THE PREVALENCE OF RELIGIOUS SERVICE ATTENDANCE IN AMERICA: A REVIEW AND META-ANALYSIS

MAJOR PROFESSOR: Dr. Meera Komarraju

The Gallup Poll and General Social Survey have asked Americans about their religious service attendance since 1939 and 1972, respectively. With remarkable consistency, these two surveys have estimated that just over 40% of the American population regularly attends religious services. Yet, recent research has called this “gold standard” into question, citing three sources of bias in these estimates: (a) ambiguous item wording, (b) an ambiguously specified time frame; and (c) data collection methods that lend themselves to socially desirable responding. Several lines of research have developed to eliminate or minimize these sources of bias, but these efforts have yielded a wide variety of results, with some estimates being half as much as the gold standard! Methodological and psychometric differences are not the only source of variation, however. The characteristics of those sampled into studies also introduces variability. Given that attendance estimates are likely influenced by variations in both methodology and sampling, this study uses meta-analytic techniques to estimate the extent of their influence and to estimate the attendance rate after controlling for their influence. The findings indicate that efforts to reduce socially desirable responding have had the greatest impact on the attendance rate, followed by efforts to overcome the ambiguously specified time-frame. In addition, attendance rates are positively related to the proportion of African Americans, Whites and married respondents sampled, as well as mean years of education. Attendance rates are also negatively

related to the proportion of 18 to 30 year-old respondents sampled. After controlling for these methodological and socio-demographic study characteristics, the prevalence of weekly attendance in America was variously estimated as 41.4% for the gold standard items, 43.1% for items measuring attendance in the past week, 27.8% when asking respondents what they did yesterday (i.e., on Sunday via the time-use methodology) and 22.7% when attendance was counted manually.

DEDICATION

Dedicated to my Dad, Stephen Delos Briggs, whose life is my example, and my Mom, Bonnie Mae Briggs, who means everything to me.

ACKNOWLEDGEMENTS

I would like to thank Dr. Alan Vaux, Dr. Margaret Stockdale and Dr. Meera Komarraju for their patience and guidance on this project. Without your unswerving belief and patience in me, I would not have been able to achieve this goal that I have long pursued. I came into the program with the intention of doing my best, and you allowed me the opportunity to dig deep into a literature that was unknown to me, and to increase my knowledge exponentially. For that, I will be forever grateful. I would like to thank the Applied Psychology faculty for providing outstanding instruction, guidance and learning opportunities along the way. You prepared me well to pursue my dreams.

I would like to thank my brother and sister for always being there for me; for pushing me to be better; for showing and telling me how to be a better person; and for sharing your lives with me. Mike, you will forever be my "Bikey." Karen, you will forever be my big sister who will take care of me, and also make me "tougher." I love you both!

I would like to thank my Dad for showing me how to work hard and be a man; for showing me how to persist and not give up; for showing me how to do things the right way; for encouraging me, holding me accountable and pushing me; but most of all for loving and supporting me. I agree with my sister: You are my hero, and this is for you.

I would like to thank my Mom who is one of my best friends. Thank you for setting the expectation that I would go to college, and for helping me get through that first year! Thank you for supporting me in every way and for taking my calls at all hours of the day and night. Thank you for encouraging and believing in me, and making me laugh. Thank you for providing a love that makes me overflow. Thank you for setting an example of Christian love. You too are my hero! And, you know what, "it's a mystery, but it all worked out."

Finally, I would like to thank God and His Son, Jesus Christ. Thank you for setting my heart on fire for other people, for research, for your Church and for the things in life that bring us joy. Thank you for the people you have allowed me to meet and be blessed by along the way! May this be another step into something larger, and whatever lies ahead, let it be with You.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A Brief History of the Social Scientific Study of Religion

Religion is an integral facet of American life. Various nationally representative polls indicate that over 90% of Americans believe in God or a "higher power" (Baylor Religion Survey, 2005; Gallup Organization, 2007),[1] 85% report feeling close to God (Argyle, 2000), 89% "desire to be *closer* to God or in union with Him" (Davis & Smith, 2004), 84% state that religion is either "fairly important" or "very important" in their lives (Gallup Organization, 2007), 75% indicate that their approach to life is centrally grounded in religion (see Larson, Swyers & McCullough, 1998), 79% pray at least once a week, 86% ask for God's help in the midst of daily activities, 84% feel that they are guided by God at least once in a while, 88% find strength in their religion or spirituality (Davis & Smith, 2004), 85% or more affiliate themselves with a religion (Gallup Organization, 2007; Hout & Fischer, 2002),[2] 63% are members of a church or synagogue, 77% believe that the "Bible is the actual or inspired Word of God" (Gallup Organization, 2007) and 76% report that they have read the Bible, Torah, Koran or other sacred text (37% do so at least once a month; Baylor Religion Survey, 2005). Taken together, these statistics indicate that religion plays an integral part of everyday life for a large proportion of Americans.

---

[1] Most religion surveys (including the Gallup Poll and the General Social Survey) use items that ask respondents if they believe in "God" or if they believe in "God or a higher power." In almost all cases, more than 90% of respondents answer affirmatively to these types of questions. However, the Baylor Religion Research Survey attempted to tease out respondents who believe in God from those who believe in some sort of higher power or cosmic force. Their results are revealing, with 78% of respondents indicating that they believe in God, and an additional 14% choosing the higher power/cosmic force option. Although these latter results still suggest that over 9 in 10 Americans believe in God or a higher power, the actual rate of "belief in God" is much lower than what has been indicated in other surveys.

[2] Estimates for specific religious affiliations are as follows: 49% Protestant, 24% Catholic, 2% Jewish, 2% Mormon, 11% Other, 11% None and 1% unspecified (Gallup Organization, 2007).

Given the centrality of religion among Americans and its importance in shaping cognitions, affect and behavior, one would assume that religion would be a prime area of research in the social sciences. Unfortunately, the history of social scientific religious inquiry does not bear this out. Although several of the "founding fathers" in psychology and sociology (e.g., Durkheim, 1897/1951, 1915/1961; Freud, 1907/1962, 1927/1961; Galton, 1872; James, 1902) devoted a significant amount of time to studying religion and its psychosocial effects, social scientific inquiries into religion were largely absent for the better part of the twentieth century. This religious research void is likely due to a combination of factors.

One likely contributor to the research void is the relatively low level of religiosity found among social scientists. For example, while over 90% of Americans believe in God or a higher power, just over 40% of psychologists share the same belief (Callahan, 2001).[3] Furthermore, data from the General Social Survey (GSS) indicate that while just 7% and 16% of the American public rejects religion and the Bible as "God's Word," 19% and 38% of social scientists do so (Stark, Iannaccone and Finke, 1997), respectively. Social scientists and mental health professionals are also much more likely to endorse "spirituality" and reject traditional religion (Shafranske & Malony, 1985; Zinnbauer, Pargament, Cole, Rye, Butter, Belavich et al., 1997).[4] For example, Shafranske and Malony (1985) reported that 71% of psychologists describe spirituality as relevant, but just 9% report high levels of religious involvement and 74% indicate that religion is not their primary source of spirituality. Even when compared to professionals in other scientific fields, social scientists, especially those in psychology and anthropology,

---

[3] The article by Callahan was published in a magazine (*Commonweal*); consequently, the source for the statistics on "belief in God" was not provided to the reader. Thus, it's possible that the sample from which the statistics were drawn was unrepresentative of all psychologists in America. Yet, Callahan's reported estimate for the general public (i.e., over 90%) is consistent with estimates obtained from nationally representative samples, which adds some credibility to the statistic that was reported for psychologists.

[4] Zinnbauer, Pargament and colleagues (1996, 1997) found that individuals who identified themselves as "spiritual but not religious" were more likely to hold agnostic beliefs and to participate in non-traditional activities such as healing groups, meditation and yoga than those who identified themselves as "spiritual and religious."

"emerge as towers of unbelief" (Stark et al., 1997, p. 22). According to Stark et al., faculty in psychology and anthropology are nearly twice as likely to disavow religion and to never attend church as faculty in the physical sciences. Taken together, these findings suggest that part of the religious research void may have been due to the relative irreligiousness of theorists, researchers and practitioners in the social sciences.

This level of irreligiousness may stem from a couple of factors. First, irreligious persons may be self-selecting themselves into careers in the social sciences at a greater rate than religious persons. For example, Stark et al. (1997) reported that less religious persons were more likely to pursue advanced degrees and careers in the social sciences than religious persons. Thus, the social sciences may be relatively unappealing to those who would otherwise be most likely to engage in religious research.

Second, a number of prominent and influential early figures in the field were notorious for their anti-religious sentiments. Perhaps there is no better example of this than Sigmund Freud. In his 1927 treatise on religion, entitled *Future of an Illusion*, Freud referred to religion as an "obsessional neurosis," "illusion," "poison," "intoxicant" and "childishness to be overcome" (Freud, 1927/1961, p. 43 – 62), and claimed that "the encounter with death [is] the source of religious beliefs" (Freud, 1907/1962; cited in Lazar, 2006, p. 180). Although Freud correctly observed that religion could be problematic, he has been criticized for drawing his conclusions based on data that suffer from selection bias (e.g., see Koenig, McCullough & Larson, 2001). Specifically, Koenig et al. argue that Freud's conclusions were based on data obtained from mental health patients who happened to be religious, rather than on data obtained from representative samples of religious and non-religious persons. As a result, Freud likely overestimated the pervasiveness of religion's darker side, especially in terms of its effects on mental health (cf. Koenig et al., 2001). Nevertheless, Freud's clout within psychology was considerable during the early-to-mid twentieth century, and his views (along with others) no doubt played a role in the abandonment of research on religion.

Additionally, both the paradigmatic influence of positivism and the belief structures emanating from secularization theory likely played a role in the abandonment of religious research among social scientists. Positivism, which was formulated by Auguste Comte in the nineteenth century, is an epistemology that emphasizes the acquisition of knowledge solely through observable and verifiable processes (Trochim, 2001). This school of thought, which was predominant in the social sciences during the early-to-mid twentieth century, necessarily precludes the study of matters pertaining to the meta-physical or supernatural. It is no great surprise, then, that the study of religion waned during the same period in which positivism reigned in the social sciences. In a related vein, secularization theory, which is a product of the enlightenment period, posits that modernization is necessarily accompanied by an epistemological shift from religion and faith to science and reason (Argyle, 2000). Although perhaps not as influential as positivism, secularization theory, coupled with other timely historical events (e.g., the publishing of Darwin's theory of evolution), likely dissuaded social scientists from even perceiving that religion was a field of study worth pursuing.

The confluence of these anti-religious forces resulted in a veritable "40 years in the desert" for the social scientific study of religion (cf. Glock, 2009). With a few notable exceptions (e.g., Allport, 1950), there was very little social scientific research on religion between Freud's last main religious work in 1927 until the 1960s (see, for example, Berger, 1967; Glock & Stark, 1965; Greeley, 1963; Hirschi & Stark, 1969; Moberg, 1961; Stace, 1960; Stark & Glock, 1965, 1968)[5] and 1970s (e.g., Benson & Spilka, 1973; Bouma, 1979; Carroll & Roozen, 1975; Comstock & Partridge, 1972; Gorsuch & McFarland, 1972; Greeley, 1972, 1976, 1979; Hoge & Roozen, 1979; Kelley, 1972; Stark & Bainbridge, 1979; Wuthnow, 1977, 1978, 1979). At this point, however, the devices that once held interest in religious research at bay began to wane. Freudian theory proved largely untenable and his conclusions concerning religion were re-framed given the nature of his (unrepresentative) samples and the emergence of data that contradicted Freud's claims (cf. Koenig et al., 2001). The influence of positivism also lessened as social scientists began to realize that human behavior is much less deterministic than the objects of study in the physical world, and as new paradigms (e.g., post-positivism) emphasizing causal complexity began to emerge. Additionally, the accuracy of secularization theory was called into question by 70 years of religious stability (and even growth) in the world's most developed nation (e.g., see Gallup Organization, 1985, 2007). These paradigmatic shifts helped legitimize the pursuit of religious studies in the social sciences.

At first, the research on religion remained limited. The 1960s and 1970s were characterized by a smattering of psychosocial studies on religion. In the 1980s, however, the rate of religious research production increased noticeably, and by the end of the millennium the social scientific literature on religion had grown exponentially. One of the hallmarks of this growth is a tome on religion and health published by Koenig et al. (2001), in which they summarize findings from over 2,000 studies and 400 review articles. The number of studies on religion has continued to grow since Koenig et al.'s review,[6] and there are now major contributions to the social scientific study of religion from a number of fields: criminology (e.g., Stark, 1996), economics (e.g.,

[5] Three leading journals on the scientific study of religion were also first published in the late 1950s and early 1960s: the *Review of Religious Research* in 1959, the *Journal for the Scientific Study of Religion* in 1961 and the *Journal of Religion and Health* in 1962.

[6] In fact, Koenig, King and Carson (2012) just recently published a second volume that summarizes an additional 3,000 studies and review articles that were published between 2000 and 2010.

Iannaccone, 1984, 1988, 1990, 1995a, 1995b, 1997, 1998), epidemiology (e.g., Levin, 1996), health and medicine (e.g., Koenig et al., 2001), nursing (e.g., Baldacchino & Draper, 2001; Mickley, Carson & Soeken, 1995), psychology (e.g., Batson, Schoenrade & Ventis, 1993; Frederickson, 2002; McCullough, Pargament & Thoresen, 2000; Pargament, 1997) and sociology (e.g., Hadden, 1987; Iannaccone, 1988, 1994; Kelley, 1972).

As the social scientific literature on religion developed, the measurement of religious variables also became increasingly sophisticated. Whereas early studies typically included single-item measures (e.g., religious affiliation, religious service attendance, religious beliefs), reliable and valid multidimensional measures of religion are now both available and widely recommended. For example, a national working group supported by the National Institute on Aging (NIA; a division of the National Institute of Health) and the Fetzer Institute assembled a multidimensional measure of religion and spirituality that included 12 domains relevant to the study of religion and health (Fetzer Institute/NIA, 1997). A number of multidimensional measures have also been developed to tap individual constructs (e.g., see Batson et al., 1993; Ellison, 1983; Hood, 1975; Ladd & Spilka, 2006; Pargament, 1997; Piedmont, 2008; Poloma & Pendleton, 1989; Roof & Roof, 1984), and a compendium of 124 multi-item, multi-factor religious and spiritual measures have been assembled by Hill and Hood (1999). All in all, great strides have been made in the measurement of religious and spiritual variables.

Despite these psychometric advances, a single item measure of religious service attendance still stands out as one of the most consistent and robust predictors of a variety of biopsychosocial outcomes.[7] For example, attendance at religious services has been robustly associated with voting preferences and behavior (Green, 2007; Newport, 2005, 2007), volunteerism (Wuthnow,

[7] That religious service attendance has emerged as such a general and powerful predictor is fascinating, especially considering the statistical limitations associated with single-item measures (e.g., see Nunnally, 1978).

1991), prejudice (Allport, 1987; Altemeyer, 2003), crime (DiIulio, 2002; Johnson, Larson, Li & Jang, 2000), delinquency (Johnson et al., 2000; Stark, 1996), drug and alcohol use (Musick, Blazer & Hays, 2000; Richard, Bell & Carlson, 2000; Stark, 1996), marital happiness and functioning (Glenn & Weaver, 1978; Hummer, Rogers, Nam & Ellison, 1999), divorce (Larson, 1985; Mahoney, Pargament, Tarakeshwar & Swank, 2001), well-being, mental health, physical health (Koenig et al., 2001), suicide (Comstock & Partridge, 1972; Nisbet, Duberstein, Yeates & Seidlitz, 2000) and mortality (Idler et al., 2009; Strawbridge, Cohen, Shema & Kaplan, 1997). Despite the consistency with which attendance is associated with these constructs, however, many questions remain.

One of the primary questions that cuts across each of the attendance associations mentioned above stems from the way in which it is measured. Typically, attendance is measured with a single item, worded similar to the following: "How often do you attend religious services?" At face value, this item simplistically measures the frequency with which respondents come into contact with a religious event. But, because the act of attending a religious service can be characterized by a wide variety of experiences and exposures (e.g., social interactions and relationships, prayer and meditation, music and the arts, instruction and guidance, prescription and proscription, coping skills, accountability, leadership opportunities, group membership and identification and so on), the item actually taps into a plethora of constructs. When attendance is measured as a single item, then, it is essentially treated as a "black box" construct. That is, when associations with attendance are found, we are left to speculate about which of the "experiences and exposures" associated with attendance are responsible for the association. Numerous researchers have attempted to use theory to deductively identify and model the "active ingredients" (i.e., the moderators and mediators) of religious services. Yet, in most of these models, the intervening variables are unable to completely explain the associations with attendance mentioned above, suggesting that attendance either has direct effects or that unexamined mediators have yet to be identified and modeled.

Addressing the unexplained variance issue will require research that attempts to inductively identify, measure and model the various components of religious services. Such efforts would make it possible to discover additional active ingredients that have not already been identified via theory or other deductive strategies. This line of research would also help determine if religious service attendance exerts its biopsychosocial effects simply by exposing individuals to a host of mediating variables (i.e., an indirect effects model), or if attendance offers a unique "Gestalt effect" in which attendees receive something more than the sum of the parts (i.e., a direct effects model). Most importantly, such a line of research would help point the way to the design and implementation of interventions that would help promote the health and well-being of Americans (e.g., see Lundahl, Taylor, Stevenson & Roberts, 2008).

Before performing in-depth studies on the attendance construct, however, basic facts about attendance (e.g., exposure rates) need to be established. For example, it has long been assumed that just over 40% of the American population regularly attends religious services (see Hadaway & Marler, 2005; Iannaccone, 1998, 2003). Yet, recent research has called this "gold standard" into question, citing three sources of upward bias in the original estimates. Specifically, it has been argued that the original attendance item estimates (from the Gallup Poll and GSS) are inflated because of: (a) ambiguous item wording, (b) an ambiguously specified time frame; and (c) data collection methods that lend themselves to socially desirable responding. Several lines of research have developed in an attempt to eliminate or minimize these sources of bias, but these efforts have yielded a wide variety of results (cf., Gallup Organization, 2007; Hadaway, Marler & Chaves, 1993; Presser & Stinson, 1998; Smith, 1998), with some estimates being half as much as the gold standard estimate (e.g., see Hadaway & Marler, 2005)!

Methodological and psychometric differences are not the only source of variation in the attendance estimates, however. Given the well documented differences in attendance across a number of sociodemographic variables (e.g., see American National Election Study, 2007; Davis & Smith, 2004; Ferraro & Kelley-Moore, 2001; Koenig et al., 2001; Stark, 2002; Strawbridge et al., 1997) and religious (e.g., see Iannaccone, 2003; Lazerwitz, 1961; Smith, Lundquist Denton, Faris & Regnerus, 2002; Stark & Iannaccone, 1997), it seems evident that sampling also introduces a degree of variability into the attendance estimates. For example, Hays et al. (1998) collected data from a stratified random sample of 2,971 elderly persons living in 5 contiguous counties in North Carolina, and found a regular religious service attendance rate of 76.8%. This rate is much higher than the gold standard estimate; yet, the sampling frame was focused on elderly persons living in North Carolina, and both the elderly and those living in the South are more likely to attend than younger individuals (e.g., see ANES, 2007) living in other regions of the country (Gallup & Jones, 1989; Hoge & Roozen, 1979; Iannaccone, 2003), respectively. Furthermore, the sample obtained by Hays et al. was over-representative of highly religious (e.g., close to 60% reported praying daily) African American (54.2%) women (66.7% were female) who are all more likely to attend than those who are less religious (e.g., see Musick, Traphagan, Koenig & Larson, 2000), White (e.g., see Stark et al., 1997) or male (ANES, 2007; Davis & Smith, 2004; Stark et al., 1997), respectively. As the Hays et al. study demonstrates, then, both the scope of the sampling frame and the composition of the sample are likely to add systematic variability to the population of religious service attendance estimates.

Given that attendance estimates are likely influenced by variations in methodology and sampling, it would be useful to estimate both the extent of their influence, and the attendance rate after controlling for their influence. To this end, meta-analytic techniques were employed to: (a) guide a systematic search of the literature for all published and unpublished reports of religious service attendance frequency; (b) extract all attendance data and information on potential moderating variables; (c) estimate the amount of systematic variation in the attendance estimates that is attributable to methodological and sampling differences across studies; and, (d) estimate the prevalence of religious service attendance in America while controlling for these systematic sources of variance.

Before pursuing these ends, I review the literature on the various lines of research that have attempted to estimate the prevalence of religious service attendance in America. In discussing these lines of research, methodological and item differences across studies were identified and presented as potential moderators of attendance frequency. Subsequently, the literature describing variations in attendance across a number of sociodemographic variables is discussed, with each being presented as a potential moderator of attendance frequency.

CHAPTER II

LITERATURE REVIEW

Prevalence of Religious Service Attendance in America

There are approximately 331,000 churches, synagogues, temples, mosques and other places of worship in the United States (Hadaway & Marler, 2005)[8] that are served by over 300,000 clergy (Koenig et al., 2001) and attended by as many as 127 million people every week. The latter figure is based on the most current U.S. population estimate (Census Bureau, 2009)[9] and a religious service attendance rate of slightly over 40% (i.e., 42%; see Hadaway & Marler, 2005). The "slightly over 40%" attendance rate is the most commonly accepted estimate in the literature (see Hadaway & Marler, 2005; Iannaccone, 1998, 2003), and is derived from decades of data from two well-known nationally representative surveys: the Gallup Poll and the GSS.

The Gallup Poll has primarily employed two items over the past seven decades to assess the prevalence of attendance among Americans. The first item employed by Gallup dates back to 1939, and assesses attendance within the past week by asking respondents, "Did you, yourself, happen to attend church or synagogue in the last seven days, or not?"[10] Then, in 1992, the Gallup Poll introduced a new item that assesses regular patterns of attendance by asking respondents, "How often do you attend church or synagogue?"[11] The first two response options for this item ("At least once a week" and "Almost every week") are combined to provide an estimate of regular religious service attendance. In a recent Gallup Poll (May, 2009), 45% of Americans indicated that they attended within the past week, while 46% indicated that they

[8] The 95% confidence intervals for this estimate are 309,000 and 356,000 (Hadaway & Marler, 2005). Of the 331,000 places of worship in the U.S., 82,183 were categorized as Protestant, 178,672 as Conservative Evangelical, 21,975 as Catholic or Orthodox, 36,450 as Other Christian and 11,720 as non-Christian.
[9] As of 2008, the U.S. population estimate was 304,059,724.
[10] Response options for the original Gallup item are "Yes" and "No."
[11] Response options for the new Gallup item include: "At least once a week," "Almost every week," "About once a month," "Seldom" or "Never")

attended regularly (F. Newport, personal communication, October 26, 2009; Gallup & Jones, 1989). The GSS item, which was first introduced in 1972, also assesses regular patterns of attendance, but with a slightly different question: "How often do you attend religious services?"[12] The first four response options for the GSS item ("Several times a week," "Every week," "Nearly every week" and "2 to 3 times a month") are combined to form an estimate of regular attendance. In the two most recent GSS (i.e., in 2004 and 2006), 43% and 40% of Americans, respectively, reported that they attend regularly (Davis, Smith & Marsden, 2007). Thus, the most recent Gallup and GSS data confirm that over 40% of the American population attends a place of worship on a weekly or near weekly basis.

Importantly, the current Gallup and GSS data not only converge on a similar attendance rate, but they have also done so for many years. The GSS, for instance, has consistently reported attendance rates that are equal to or greater than 40% since 1972 (Davis, Smith & Marsden, 2007), while the new Gallup item has done so since 1992 (F. Newport, personal communication, October 26, 2009; see Figure 1). Even more remarkable, the original Gallup item has yielded attendance rates of 40% or more since the late 1930s (Gallup & Jones, 1989; F. Newport, personal communication, October 26, 2009; see Figure 2). The consistency with which these two surveys have converged on the same attendance rate over time has helped establish an oft-cited "gold standard" in the attendance literature. Even a brief perusal of the wider literature on religion and spirituality will lead one to encounter numerous opening statements citing the "slightly over 40%" attendance figure.

[12] Response options for the GSS item include: "Several times a week," "Every week," "Nearly every week," "2 to 3 times a month," "Once a month," "Several times a year," "Once a year or less" and "Never."

Instrumentation Criticisms and Measurement Alternatives

Given that a consensus on the prevalence of religious service attendance in America seems to have developed in the literature, it comes as somewhat of a surprise that recent research has called the veracity of the Gallup and GSS findings into question. This questioning, however, stems from three well-founded criticisms. The first two criticisms focus on the language used in the attendance items. Specifically, the items are criticized for including both an ambiguous time frame and ambiguous (and over-inclusive) terminology, each of which potentially introduce a degree of upward response bias. The third criticism stems from the use of data collection methods that have been shown to introduce social desirability bias. In the last two decades, several lines of research have developed in an attempt to both estimate and minimize these shortcomings, while providing a more accurate estimate of the prevalence of religious service attendance in America. Below, I take a closer look at the three criticisms, and the lines of research that have developed to address these issues.

Ambiguous Time Frame

Although the response options for the original GSS and new Gallup items are time-oriented (e.g., the GSS options range from "Several times a week" to "Never"), the respondent is still allowed a degree of discretion in terms of which part of their attendance history to report. Consider, for example, a person who, for most of their lives has been an irregular attender, but who has recently taken to attending every week.[13] Given the wording of the question, there is no guidance given to the respondent about which of their two "histories" to report. If respondents tend to choose the time period in which they attend more often, as social desirability theory would suggest (e.g., see Presser & Stinson, 1998), then the overall attendance estimate will be

[13] The opposite scenario also applies (i.e., where a person has been a regular attender for most of their lives, but has recently taken to attending irregularly).

biased upward. Thus, it is possible that the attendance estimates yielded by the original GSS and new Gallup items are inflated because there is no direction given to the respondent about the time period to which they should refer.

The original Gallup item, however, provides a specific time frame for the respondent (i.e., “in the last seven days”); and, because the original Gallup item shares the same general methodology and language as the original GSS and new Gallup items (i.e., data for all three are collected via an interviewer, and all three include ambiguous terminology), it should be possible to estimate the effect of providing a specific time frame on the attendance rate by comparing the item results. Using data from the GSS (see Davis, Smith & Marsden, 2007), it is possible to estimate the weighted average regular attendance rate across all 26 years of the GSS (see Figure 1). From 1972 to 2006, the mean GSS regular attendance rate was 43.0% (range = 38.2%, 50.0%). Unfortunately, sample sizes were unavailable from Gallup, thereby making it impossible to compute weighted averages for the original and new Gallup items. It is possible, however, to approximate what these averages would be by using the median.[14] From 1992 to 2009, the median attendance rate for the new Gallup item was 44% (range = 41%, 46%; see Figure 1).[15] Thus, the longitudinal GSS and Gallup Poll data indicate that rates of regular attendance range from approximately 43% to 44%, respectively. Turning to the original Gallup item, however, we find that the median attendance rate since 1939 is a couple of percentage points lower, at 41% (range = 38%, 50%; see Figure 2).[16] Comparing the two sets of estimates indicates that socially desirable responding (due to the lack of a specific time frame) biases the

[14] The mean could also be used, but there are two outliers in the original Gallup item time series. Thus, the median is more appropriate.
[15] Decimals were not provided for the Gallup data (F. Newport, personal communication, October 26, 2009); hence, they are not reported here either.
[16] The original Gallup item data series was also characterized by a couple of outliers; thus, the median is a more appropriate measure of “average” for these data than the mean.

original GSS and new Gallup item estimates upward by 2% to 3% (or by a factor of 1.05 to 1.07),[17] respectively.

It's not at all clear, however, that the differences between the estimates obtained from the regular attendance items and the original Gallup item are due solely to socially desirable responding. Specifically, it's possible that the response option differences between the two sets of items would lead to a similar difference given that not all respondents who report attending "2 to 3 times a month" on the original GSS item, for example, would report attending "in the past seven days" on the original Gallup item. In fact, if the response option categories for the original GSS item are weighted by a probability of attending in a given week,[18] then the average attendance estimate for the GSS time series falls from 43% to 40.6%, which is in line with the original Gallup item estimate of 41%. When the response options for the new Gallup item are weighted in a similar fashion, however, the attendance estimate actually increases from a median of 44% to 45.7%.[19] It isn't clear why the attendance estimate for the new Gallup item increases when the time range is restricted, but perhaps the relatively limited response options at the upper end of the scale prevent respondents from adequately spreading out (i.e., the new Gallup item provides only two response options for those who attend once a month or more, whereas the original GSS item provides four). In sum, if only the GSS item is considered, the manifest conclusion would be that the specification of a more limited time frame (and not socially

---

[17] These figures were obtained by dividing the base attendance rates (i.e., 43% and 44%) by the reduced attendance rate (41%). This method is used throughout this paper, and is hereafter referred to as the inflation rate.

[18] Each response option for the GSS attendance item was assigned a probability weight based on the likelihood that respondents in that category would attend in a given week. These weights were multiplied by the frequency with which a response option was chosen, and the products were summed across all response option categories to arrive at a percentage estimate of Americans expected to attend in a given week. The weights used for the GSS item were adapted from Presser and Stinson (1998), and are as follows: "More than once a week" = .99; "Every week" = .99; "Nearly every week" = .75; "2 to 3 times a Month" = .58; "Once a month" = .23; "Several times a year" = .08; "Once a year" = .02; "Less than once a year" = .01; "Never" = 0.

[19] The weights used for the new Gallup item are as follows: "Once a week" = .99; "Almost every week" = .75; "About once a month" = .23; "Seldom" = .05; "Never" = 0.

desirable responding) is responsible for the drop in attendance observed between the original GSS and Gallup items. Yet, the results from the new Gallup item call this conclusion into question. Regardless of the reason, however, the data presented here do suggest that the provision of a clearly defined time frame tends to lower the attendance rate by a few percentage points.

Ambiguous Terminology

A degree of ambiguity also exists in the terminology used in the Gallup and GSS attendance items. Specifically, when formulating a response to these items, participants must decipher what is meant by "religious services," "church" and "synagogue." Although on the surface these terms don't seem to present a problem, one can quickly see that they can be over-inclusive if respondents include in their answers such events as prayer meetings, scripture studies, covenant groups, choir practices, social events or other events held at a place of worship (Newport, 2006).[20] In fact, Edgell Becker and Hofmeister (2001, p. 711) lend credence to this concern given that close to one-fifth of their random sample of 1,000 New Yorkers reported "some involvement" in these "other" types of religious activities.

In 1996, the GSS tested the possibility that respondents were using over-inclusive definitions of "religious services" in formulating their responses by implementing a branching question. The question stem asked respondents if they had done any of four activities during the previous seven days, the last of which was "attend religious services" (see Smith, 1998). Respondents who indicated attending in the last seven days (30.6%) were then asked three follow-up questions: whether they had (a) attended a "regular, weekly worship service at a

[20] The Gallup item, which asks about "church" and "synagogue" also omits mention of the Muslim and Hindu places of worship (i.e., "mosques" and "temples," respectively). This omission might lead to an underestimate of the actual attendance rate in America. Yet, there are relatively few Muslims and Hindus in the U.S., and there is evidence to suggest that adherents of non-Judeo-Christian faiths tend to re-interpret Judeo-Christian terms such as "church" and "synagogue" into their own vernacular (see Underwood & Teresi, 2002).

church/synagogue (e.g., mass or Sunday morning services)" (not including viewing a service on TV or listening to one on the radio); (b) watched "a religious program on television or listened to a religious program on the radio;" or, (c) attended "some other type of religious event or meeting (e.g., prayer breakfasts, Bible-study groups, choir practices, church-sponsored lectures, adult fellowship meetings)?" Approximately one-tenth (11.2%) of the respondents who initially indicated attending a religious service either meant something else (i.e., a religious media program or other event) or were uncertain (Smith, 1998). Excluding the latter respondents reduced the overall attendance estimate to 28%. Thus, problems with ambiguous wording were responsible for inflating the attendance rate by 2.6%, or a factor of 1.09. Because the baseline attendance rate in this study is much lower than the gold standard rate (i.e., 41% for attendance within the past week), however, the actual inflation percentage we could expect with the original Gallup item is closer to 3.4%.

In another study, Hadaway et al. (1993) used a random sample to conduct a phone survey with Protestants living in Ashtabula County, Ohio. Among the 606 respondents, 35.8% initially indicated attending a religious service in the last seven days. In this study, however, just four respondents meant something other than a worship service. Removing these four respondents yields an attendance estimate of 35.2%, which is nearly identical to the original estimate (inflation factor of 1.02). Here again, the base attendance rate is lower than the gold standard rate so the actual inflation percentage we could expect with the original Gallup item is 0.8%. In a later study, Marler and Hadaway (1999) surveyed 300 members of a large Evangelical Church, of which 209 (69.7%) reported attending in the past seven days, but 26 meant something other than a worship service. Removing these persons from the number of attendees drops the attendance rate to 61.0% (inflation factor of 1.13). Adjusting for the relatively high rate of attendance among this sample of church members, the inflation percentage for the gold standard estimate is expected to be 5.5%. Together, the data presented here suggest that the clarification of ambiguous item wording can reduce attendance estimates by approximately 1% to 5%.

Social Desirability Bias

The third, and perhaps most frequently cited criticism of the Gallup and GSS attendance estimates is that they are inflated by socially desirable responding (e.g., see Hadaway et al., 1993, 1998). This concern seems well warranted for two reasons. First, religion is important to the vast majority of Americans (Gallup Organization, 2007), and frequent religious service attendance is often viewed as central to a religious life (Mockabee, Monson & Grant, 2001). Therefore, the respondent who is guided by social desirability will, all else equal, select the more "religious" answer (in this case, more frequent attendance). Second, both the Gallup Organization and GSS rely on self-report data collection methods (phone and face-to-face interviews, respectively). Self-report instruments are notoriously associated with social desirability bias (Schwarz, 1999), and the presence of an interviewer (whether over the phone or in-person) tends to further distort responses (Tourangeau & Smith, 1996). Thus, both the outcome in question and the methodology lend themselves to socially desirable responding. Accordingly, it may be more prudent to ask "how much" rather than "if" the Gallup and GSS attendance estimates are biased upward by socially desirable responding. Three lines of research have developed to estimate "if" and "how much" upward bias is present in the original attendance estimates.

Time-Use Items. Presser and Stinson (1998) focused their efforts on an item that attempted to eliminate or minimize all three of the problems mentioned so far (i.e., both forms of ambiguous language and social desirability bias). The item in question asks respondents how they use their time on Sundays without ever mentioning “religious services,” “church” or “synagogue.” This “time-use” item was first introduced in a nationally representative longitudinal survey (1992 to 1994) initiated by the Environmental Protection Agency (EPA), and asks respondents the following: “I would like to ask you about the things you did yesterday—from midnight Saturday to midnight last night. Let’s start with midnight Saturday. What were you doing? What time did you finish? Where were you? What did you do next?” Because the item does not contain any religious language, respondents should feel free to omit any mention of their religious behavior without perceiving a loss of credibility with the interviewer (for an example interview guide, see Appendix A). Furthermore, the time-use item eliminates the problems associated with the original attendance items, namely ambiguous wording (by omitting mention of “religious services,” “church” and “synagogue”) and an ambiguous time-frame (by asking respondents to describe what they did yesterday).

To test the utility of the time-use item, Presser and Stinson (1998) compared the EPA survey results to GSS and Gallup Poll data (all three items were used) collected during the same time period. According to five Gallup and GSS polls conducted in 1993 and 1994, an average of 41% of Americans reported attending religious services “last week.”[21] Results from the time-use item, however, revealed that just 26%[22] of Americans spontaneously reported attending on a given Sunday. Because not all religious services are held on Sundays, however, this estimate was adjusted upward to 29%.[23] Thus, the EPA time-use item produced an attendance estimate

[21] Although the original GSS and new Gallup items don’t measure attendance “in the past week,” Presser and Stinson (1998) multiplied the proportion of respondents in each response category by a probability of attendance (generated by the authors; see p. 139), which essentially converted these items into measures of weekly attendance.
[22] When sampling weights were applied to the time-use data (i.e., to correct for unequal selection probabilities), this figure rose slightly to 26.9%.
[23] The 1996 GSS asked respondents, “On what day or days did you attend religious services during the last seven days” (Davis & Smith, 1996; cited in Presser & Stinson, 1998)? Approximately 3% indicated that they attended on a day other than Sunday; hence, Presser and Stinson adjusted their attendance estimate upward by 3%.

that was at least 12% lower than the GSS and Gallup item estimates, which translates into an inflation rate of 1.41.

In 2003, the U.S. Census Bureau began adding to this line of research by collecting time-use data as part of the American Time Use Study (ATUS). ATUS interviews are conducted on each day of the week with a nationally representative sample of Americans. Interviews that were conducted on Mondays[24] with adults age 18 and over were filtered by Presser and Chaves (2007), and numbered 4,675 in 2003, 3,098 in 2004 and 2,817 in 2005. The percentages of ATUS respondents who spontaneously reported attending religious services on Sunday were comparable to the EPA study results (i.e., before the latter results were adjusted upward for attendance on days other than Sunday). Specifically, 27% of the ATUS respondents reported attending a religious service on a given Sunday in 2003, 26.8% in 2004 and 26.3% in 2005 (Presser & Chaves, 2007).[25] Remarkably, these estimates are between 15% and 16% lower than the established rate for attendance in a given week (i.e., 41%), with inflation rates ranging from 1.52 to 1.56. Taken together, the EPA and ATUS studies yielded attendance estimates that were between 12% and 16% lower than the gold standard estimates reported by Gallup and the GSS (inflation rates ranged from 1.41 to 1.56)!

Before accepting the time-use figures, however, we must consider that relying on participants to spontaneously report their religious behavior (as the time-use item does) in response to an open-ended question may lead to an underestimate of the actual attendance rate. In an informal comparison, Pargament (1997, p. 137) found that behavioral frequency estimates tend to be lower when participants are given open-ended questions than when they are given

[24] "The interviews were distributed roughly evenly across the weeks of the year" (Presser & Chaves, 2007, p. 419).
[25] The ATUS attendance estimates reported by Presser and Chaves (2007) were adjusted for sampling error. Applying a similar adjustment to the initial 1992-1994 EPA time-use attendance data brought the original attendance estimate up to 26.9%, which is right in line with the ATUS findings.

closed-ended questions with a pre-defined set of response options. Below, I consider two reasons why this might be so, and what the implications are for the time-use attendance item estimates.

First, and perhaps most important, respondents may encounter memory retrieval problems. The provision of response options for closed-ended items provides respondents with prompts and context clues that help "jog" the respondent's memory. With open-ended items, however, respondents are not given these prompts or clues, and must rely on their memories to construct a response. Open-ended items also provide less guidance to the respondent as to what, exactly, the researcher is looking for in the response. Thus, behavioral frequency estimates will suffer to the extent that respondents are unable to both remember the details of the behavior in question and to "guess" what the researcher is looking for in the response. In defense of the time-use item, however, it does *not* seem to be a true open-ended question, and therefore, may not be subject to the same criticisms. Specifically, the reference period for the time-use item (i.e., within the last 24 hours), along with the salience of the behavior being considered (i.e., actively traveling to, and participating in a religious service), should minimize problems with memory recall, and the nature of the response format (what did you do first, what did you do after that?) should provide prompts and context clues for the respondent (assuming they can remember the first thing they did yesterday, and associate that with subsequent events). Thus, the time-use item may suffer from some of the problems that characterize true open-ended items (even remembering what happened yesterday can be difficult), but the nature of the item should minimize these problems and yield attendance estimates that are relatively close to the true attendance rate in a given week.

Second, some religious respondents may *not* be willing to talk about their religious behavior. Koenig et al. (2001, p. 78) argue that religious respondents may be less willing to be forthright with those who are perceived to be antagonistic toward religion (e.g., scientists). Religious respondents may also fail to volunteer their religious behavior if they perceive that the interview situation falls under scriptural proscriptions to keep one's religious behavior private (e.g., see Matthew 6:5-6, NIV). Alternatively, if respondents reveal that they engaged in socially undesirable behavior on Saturday night (after midnight), they may be less willing to reveal that they also attended a religious service on Sunday morning for fear that they would be perceived as hypocritical. Because the time-use item asks respondents to account for all 24 hours of the day, however, the opportunity to omit mention of religious behavior should be minimal. For example, if a respondent reports driving to town at 9am and returning home at 11am, the interviewer can easily detect that the reason for going to town and the time between 9am and 11am were omitted. Thus, underreporting for social and religious reasons should also be minimal with the time-use item.

In sum, the time-use item not only minimizes the three sources of upward bias associated with the Gallup and GSS items (i.e., ambiguous language and social desirability bias), but also minimizes memory retrieval problems (via the proximal reference period and built-in memory prompts and context clues) and underreporting due to social or religious reasons (via the hour-by-hour account of the previous day's activities). The lone caveat for the time-use item is that it refers only to the previous Sunday, and fails to assess typical patterns of attendance over time. Given these characteristics, we can conclude that the time-use item provides an excellent method of obtaining attendance rates on *any given Sunday*, and that these rates are much lower than those yielded by the original Gallup item.

Self-Administered Items. Because the time-use item simultaneously minimizes all three sources of upward response bias, it is not possible to determine how much bias is attributable solely to social desirability. In their second line of research, however, Presser and Stinson (1998) investigated an item that isolates at least part of the social desirability effect. Specifically, Presser and Stinson examined an attendance item that was used in the Monitoring the Future Survey (MFS), which is a nationally representative survey of high schoolers and young adults. The wording of the MFS item was identical to the original GSS item,[26] therefore it still includes problems with ambiguous wording (i.e., "religious services") and an ambiguously specified time-frame. The Monitoring the Future Survey, however, is a *self-administered* instrument, which allows for the removal of a large source of social desirability bias (i.e., the presence of an interviewer). Because the self-administered MFS item reduces only one source of response bias, we should be able to estimate the amount of upward bias in the original attendance estimates that is due to the presence of an interviewer.

A total of 6,700 young adults (ages 19 to 28) were given the self-administered MFS item between 1993 and 1994. Because this sample is limited to young adults, Presser and Stinson (1998) use comparison data from interviewer-administered GSS and Gallup surveys (the original items were used) conducted between 1993 and 1994 with the same age group. Presser and Stinson also utilized time-use data collected from this age group by the University of Michigan's Survey Research Center (SRC) between 1992 and 1994. The results from Presser and Stinson's comparison are provocative. Using the interviewer-administered attendance items, the GSS and Gallup Poll found attendance rates of 30% and 36%, respectively. Consistent with previous findings, the SRC time-use item yielded a much lower attendance rate of 21% (inflation rate =

[26] The response options were slightly different, however: "Never," "Rarely," "About once or twice a month" and "About once a week or more."

1.43 to 1.71). Somewhat surprisingly, however, the self-administered attendance item also yielded a much lower attendance rate (23%; inflation rate = 1.30 to 1.57) than the GSS and Gallup items.[27] This latter figure was duplicated a decade later by Presser and Chaves (2007), who used 2003 and 2004 MFS data, and found that 24% of 19 to 28 year-olds reported attending regularly. These results suggest that socially desirable responding (due to the presence of an interviewer) is responsible for biasing the attendance estimate upward by a factor of 1.30 to 1.57. If these inflation rates generalize to the adult population, then the self-administered item would be expected to yield attendance estimates for the past week that range from 26.1% to 31.5%, which are much lower than the regular attendance estimates yielded by the Gallup and GSS items.

Before moving on to the next line of research, it is important to note that the self-administered item has several key advantages over the other attendance items discussed thus far. First, the elimination of an interviewer from the data collection process removes a big source of response bias. Accordingly, the resultant attendance estimates are likely to be more accurate than the estimates obtained from the three Gallup and GSS items. Second, the elimination of an interviewer from the data collection process makes the administration of the item much more feasible for researchers who lack the means to conduct one-on-one interviews (via phone or in-person). Third, the response options provided for the self-administered item make it possible to estimate *regular* patterns of religious service attendance. Yet, the lack of a timeline for the self-administered item potentially allows a degree of social desirability bias to leak back into the responses. In sum, the self-administered item shows promise, but future research would do well to: (a) provide a defined time-frame (ala the original Gallup item, or an alternative time frame,

[27] Both the GSS and MFS estimates were obtained by weighting the response category frequencies by probabilities of attending in a given week, and are therefore directly comparable.

such as "in the last year"); (b) provide clarification on terms such as "religious services," "church" and "synagogue;" and, (c) provide further data on the effects that ambiguous language and the presence of an interviewer have on attendance estimates.

Count Data. In the third line of research, Hadaway and colleagues (Hadaway & Marler, 1997, 2005; Hadaway et al., 1993, 1998; Marler & Hadaway, 1999) investigated a method of estimating the prevalence of attendance that also circumnavigates all three criticisms leveled against the Gallup and GSS items. Specifically, Hadaway et al. (1993) obtained count data from religious congregations and denominations and compared these with survey data. Part one of their study was limited to Protestants in Ashtabula County, Ohio. Using a random sample of all active telephone exchanges, which includes listed and unlisted numbers, Hadaway et al. implemented a phone survey[28] to obtain both an estimate of the number of Protestants living in the county and the self-reported attendance rate for this group. Just under two-thirds (66.4%) of the 606 phone respondents indicated that their religious affiliation was Protestant, which translated into a population estimate for the county of 66,565 Protestants. Among the Protestant phone respondents, 35.8% reported attending religious services in the last seven days, although this reduced to 33.2% after removing those respondents who meant something other than a worship service (n = 4) and those who attended on a day other than Saturday or Sunday (n = 6). Given the population of Protestants in Ashtabula County and the rate at which they reportedly attend, Hadaway et al. estimated that approximately 22,100 individuals should be in attendance on any given Sunday. Through extensive and exhaustive efforts, Hadaway et al. identified all of the Protestant churches in Ashtabula County (n = 159) and requested average attendance figures (i.e., count data) from these places of worship. Where count data were unavailable, attendance was estimated by the number of cars in the parking lot (churches of comparable size and

[28] The response rate was 71% (Hadaway et al., 1993).

composition were used to estimate an attendance count from the number of cars in a parking lot) or by actual head counts during services over a two-month period. From the count data, just 13,080 individuals, or 19.6% of the Protestant population in Ashtabula County were found to be in attendance on an average Sunday, which is much lower (by a factor of 1.69) than the estimates obtained from the phone survey.

In part two of their study, Hadaway et al. (1993) compared self-report and count data for Catholics living across the nation. The self-reported attendance rate for Catholics was obtained from the Gallup Organization, and count data were obtained from 18 dioceses across the United States. The results were again quite striking. According to the Gallup Poll, 51% of Catholics reported attending religious services in the last seven days, while just 28% of Catholics living in the 18 dioceses were counted present (inflation rate = 1.82).[29] The disparity between these two figures (23%) may be somewhat inflated given that no count data were included from Catholics living in the Southeast, the region with the highest rates of religious service attendance, whereas the nationally representative Gallup Poll included Catholics from this region. Any inflation in the disparity between the self-report and count data, however, is likely small given that fewer Catholics live in the Southeast (between 5% and 8% according to Hadaway et al.) than in any other region of the country. Moreover, Hadaway et al. located four surveys that provided geographically matched self-report data from Catholics living in 4 of the 18 dioceses. In each of the four dioceses, the self-report estimates were much higher (between 1.17 and 2.19 times greater) than the estimates obtained from the count data. Thus, for both Protestants and Catholics, the self-report method of estimating attendance yielded rates that were nearly twice as great as those yielded by the count method!

[29] Attendance at Catholic Mass is usually averaged over a four-week period during the fall.

Yet, count procedures are not free from criticism either. Iannaccone and Everton (2002) cautioned researchers from "placing much faith" (p. 192) in aggregated count-based estimates of attendance given that congregations and denominations vary greatly in their counting procedures (see also Zech & Gautier, 2004). Moreover, Iannaccone and Everton state that, in their experience, "most counts fall well short of actual attendance" (p. 192). Smietana (2006) reports that counts may capture as few as 60% of those in attendance, stating that persons in Sunday school and those who are late may not be counted. Furthermore, Hout and Greeley (1998), in their review of the Hadaway et al. (1993) article, point out that there was substantially more variability in the Catholic count data (a 2 to 1 ratio was found for the dioceses with the highest and lowest attendance estimates, respectively) than in the Catholic survey data (where ratios were 1.6 to 1 for census regions and 1.3 to 1 for cities, respectively; see p. 114); and, as Hout and Greeley (1998, p. 114 – 115) correctly observe, "error-ridden data vary more widely than clean data." Adding weight to these arguments is the omission of detailed information on the count procedures employed by the various congregations and dioceses included in the Hadaway et al. (1993) study. Thus, the reader has no way of assessing the amount of variability that may have been introduced in the count-based attendance estimates as a result of differing count procedures. Taken together, these arguments suggest that aggregated count data tend to be error-laden, and, at best, provide a lower-bound estimate of the actual attendance rate.

Anticipating these criticisms, Hadaway and Marler conducted three additional studies (Hadaway & Marler, 1997, 2005; Marler & Hadaway, 1999). In the first study, Hadaway and Marler (1997) personally counted the number of Catholics attending Mass at all Catholic parishes in Ashtabula County, Ohio and in Oxford County, Ontario (Canada), and compared these figures with geographically matched survey-based estimates of attendance. In both cases, the survey-based estimates were more than twice as high as the count-based estimates.

In the second study, Marler and Hadaway (1999) collected individual-level count and self-report data from members of a large evangelical Protestant church. On a single Sunday, they counted attendees at worship, obtained Sunday school attendance records and then conducted phone interviews over the next week (Sunday to Saturday) with 300 adult members of the church. Amazingly, just 115 of 181 (63.5%) respondents who indicated that they attended Sunday school on this particular Sunday were actually counted present (inflation rate = 1.57). Results from the worship service attendance comparison were not provided by Hadaway et al. (1998),[30] but the authors did state that the persons most likely to overstate their attendance were those who self-reported attending every week.

In the third study, Hadaway and Marler (2005) used a slightly different method of obtaining count data. Specifically, they utilized data collected as part of the U.S. Congregational Life Survey (USCLS; Woolever & Bruce, 2001), which included a nationally representative "hypersample" of religious congregations. The hypersample was drawn by asking GSS respondents to identify the religious congregation or group they attended; religious leaders from each of the identified congregations were then contacted and asked to provide average annual attendance figures over a five-year period. Because the GSS respondents were part of a random sample of Americans, the congregations they named were also assumed to comprise a nationally representative sample of U.S. congregations. The overall average weekly attendance figure provided by the USCLS congregations (188.5 persons) was then multiplied by the estimated total number of U.S. congregations (331,000; see Hadaway & Marler, 2005) to obtain an estimate of the total number of Americans attending religious services. This product was then divided by the

[30] The Marler and Hadaway (1997) study had not been published.

total number of persons in the U.S. to obtain a percentage of Americans attending religious services. The attendance estimate that Hadaway and Marler arrived at (21.1%) was approximately half the gold standard rate of "slightly over 40%" (inflation rate = 1.94). Thus, Hadaway and colleagues' count data consistently show attendance estimates that are nearly half of those obtained by the original attendance items.

Summary

Clearly, the research investigating the prevalence of religious service attendance has advanced in the last two decades. Prior to the early 1990's, the Gallup and GSS survey estimates were the gold standard. They not only converged on the same rate (slightly more than 40%), but had done so for the better part of the twentieth century. As researchers began to point out problems with the gold standard items, however, efforts were made to develop and test alternative items and methods that could produce more accurate prevalence estimates of attendance. Although the research in this area is still young and in need of further work, there are some revealing findings.

As expected, the sources of ambiguity associated with the Gallup and GSS items are responsible for relatively trivial amounts of inflation. Specifically, the provision of a clearly defined time frame (i.e., "in the past week") reduced the attendance estimate by 2% to 3% percent (inflation rates = 1.05 and 1.07), while controlling for ambiguous terminology reduced the attendance estimate by approximately 1% to 5% (inflation rates = 1.02 to 1.13; see Table 1). Thus, the available evidence suggests that the use of ambiguous language is responsible for inflating the Gallup and GSS attendance estimates by a factor of 1.02 to 1.13 (*median* = 1.07),[31] which translates into a median regular attendance estimate of 41%, and a median attendance estimate for the past week of 37.6% (see Table 1). Further research is needed to determine if

[31] The following inflation rates were used to calculate the median: 1.02, 1.05, 1.07, 1.09 and 1.13.

these estimates are accurate, and whether the inflation figures associated with each type of ambiguity are additive.

The inflation estimates associated with social desirability bias, on the other hand, are much higher and more complex than expected. The self-administered items—which reduce social desirability bias (due to the presence of an interviewer), but still contain problems with ambiguous language—yield inflation rates ranging between 1.30 and 1.57 (*median* = 1.44[32]), which translate into attendance rates of 26.1% to 31.5% (*median* = 28.8%; see Table 1).[33] The time-use item—which minimizes all three sources of upward response bias, but still contains a degree of downward bias due to the limited time frame—yields somewhat higher inflation rates (1.41 to 1.71; *median* = 1.53[34]) and lower attendance estimates (24.0% to 29.0%; *median* = 26.9%; see Table 1). The count-based method—which also minimizes the three sources of upward bias, but is likely affected by downward bias due to the introduction of error via the variable count procedures used by different congregations and denominations—yields even higher inflation rates (1.17 to 2.19; *median* = 1.82)[35] and lower attendance estimates (18.7% to 35.0%; *median* = 22.5%; see Table 1). Taken together, the three lines of research suggest that social desirability bias (and ambiguous wording for the time-use item and count-based method) is responsible for inflating the attendance estimates by a factor of 1.17 to 2.19 (*median* = 1.56),[36] which translates into attendance estimates that lie between 18.7% and 35.0% (*median* = 26.3%).

---

[32] There were just two inflation rate estimates for the self-administered item: 1.30 and 1.57. Also, these figures apply to 19 to 28 year-olds, and may not generalize to the larger adult population.
[33] Although the self-administered item generally assesses regular religious service attendance, Presser and Stinson (1998) weighted the response category frequencies to arrive at an estimate of attendance within the past week. Thus, the self-administered item estimates were compared to the original Gallup item median estimate (41%), which measures attendance within the past week.
[34] The inflation rates used to calculate the median were as follows: 1.41, 1.43, 1.52, 1.54, 1.56 and 1.71.
[35] Five inflation rates were used to calculate the median: 1.17, 1.69, 1.82, 1.94 and 2.19.
[36] The following inflation rates were used to calculate the median: 1.17, 1.30, 1.41, 1.43, 1.52, 1.54, 1.56, 1.57, 1.69, 1.71, 1.82, 1.94 and 2.19.

Exactly which line of research yields the most accurate attendance estimates is debatable. The self-administered item, which minimizes problems with social desirability, seems to be the best method for estimating regular attendance rates,[37] but it still includes ambiguous wording. Yet, the actual amount of inflation due to ambiguous wording is not likely to be much (i.e., between 2% and 3.4%; see Table 1). Furthermore, the self-administered item eliminates the time-intensive methods of data collection required by the other methods, and is therefore, the most convenient method for both minimizing social desirability bias and estimating regular religious service attendance rates. Thus, the self-administered item has considerable utility, and should be given serious consideration. In fact, if wording and time-reference improvements are made, it may yet prove to be the most effective and efficient means of estimating the prevalence of regular religious service attendance in America.

On the other hand, if researchers are interested in estimating the attendance rate on any given Sunday, the time-use item and count-based method both seem capable of generating relatively accurate estimates. Yet, the differences between the two sets of estimates are non-trivial. Specifically, the time-use item yielded a median estimate of 26.9%, while the count-based method yielded a median estimate of 22.5%. And, here again, there is no clear answer about which set of estimates is more accurate. The time-use item, for instance, is administered via an interviewer, which may introduce social desirability and inflate the resultant attendance estimates (although inflation should be minimized given the nature of the response format). The count-based method, on the other hand, relies on data collection methods that vary across congregations and denominations, which likely introduces a degree of measurement error and

[37] It should be mentioned again, however, that this item wasn't used as a measure of regular attendance by Presser and Stinson (1998); nor do the results reported here reflect regular attendance. Yet, the response options provided with the item lend themselves most readily to measuring regular rates of attendance.

downward bias. Thus, it's likely that the count-based method and time-use item represent the lower and upper bounds of the true attendance rate for any given Sunday.

In sum, we are left with (at least) three imperfect methods of estimating the prevalence of attendance. The self-administered item seems to be the best available method for estimating regular religious service attendance rates, whereas the count-based method and time-use item seem best suited for estimating the attendance rate on any given Sunday. Until further research can both improve upon these methods and provide clarification on some of the gray areas (e.g., by determining how much bias remains for each item and method), however, it seems prudent to continue studying each of the estimation methods covered here. Accordingly, one of the aims of this study is to accumulate all available attendance estimates in the literature, and arrive at a mean estimate for each item and method. This will not only allow for a better understanding of the variability between these methods (especially after controlling for other moderator variables), but it will also provide a basis for comparing the utility of future efforts to improve on the current methods of estimation. Before doing this, I draw on the literature reviewed here to make a few predictions about what to expect from a meta-analysis on attendance estimates.

Hypothesis 1. Studies that utilize items and methods that (explicitly or implicitly) provide a specific time frame will yield lower attendance estimates than studies that fail to do so.

Hypothesis 1a. The length of the time frame specified by an item or method will be positively related to estimates of attendance (e.g., "regular" attendance > attendance "in the past week" > attendance "yesterday").

Hypothesis 2. Studies that utilize items and methods that control for, or eliminate the use of ambiguous terminology (e.g., "religious services," "church" or "synagogue") will yield lower attendance estimates than studies that fail to do so.

Hypothesis 3. Studies that utilize items and methods that minimize or eliminate opportunities for socially desirable responding will yield lower attendance estimates than studies that fail to do so.

Hypothesis 3a. Minimizing or eliminating opportunities for socially desirable responding will suppress attendance estimates to a greater extent than controlling for, or eliminating ambiguous language (i.e., ambiguous time frame frames and terminology).

Hypothesis 4. Among the items and methods reviewed here that attempt to reduce socially desirable responding, the count-based method will yield lower attendance estimates than the time-use item, which will yield lower estimates than the self-administered item.

Sociodemographic Sources of Variation in Religious Service Attendance

Next, I examine systematic, group-level sources of variation in religious service attendance across a number of sociodemographic variables. The specific sociodemographic variables included in this review are: gender, race and ethnicity, age, familial status (i.e., marital and parental status) and socioeconomic status.

Gender

That women are more religious than men is one of the oldest and most consistent findings in the literature on religion and spirituality. Stark (2002) commented that, perhaps only in the area of crime and delinquency, are there comparable gender differences. In the literature review for this paper, a total of 138 gender comparisons were found. In 123 (89.1%) of these comparisons, females demonstrated higher levels of religiosity and spirituality, whereas men exceeded women in just 8 comparisons (6.0%), 5 of which were from samples of Jewish immigrants, with the other being from a sample of religiously unaffiliated adults. Examples of religious and spiritual outcomes where women have been found to exceed men include: holding a religious or spiritual world view (King, Speck & Thomas, 2001), viewing faith and religion as important in their lives (Brown, 1994; Evans, Cullen, Dunaway & Burton, 1995; Koenig et al., 2001; Meisenhelder & Chandler, 2000; Pew Research Center, 2016; Wallace & Forman, 1998), believing in God, the soul (Diener & Clifton, 2002) and life after death (Miller & Stark, 2002; Stark et al., 1997), reporting a religious affiliation (Adlaf & Smart, 1985; Pew Research Center, 2016; Van Tubergen, 2006), church membership (Argyle & Beit-Hallahmi, 1975; Gallup, 1990; Stark, 2002), having confidence in organized religion (Gallup, 1990), participating in youth group programs (Smith et al., 2002), reading the Bible (Barna Research Group, 2000; Stark, 2002), listening to or watching religious programs (Hays et al., 1998), engaging in prayer (Argyle & Beit-Hallahmi, 1975; Barna Research Group, 2000; Bergan, 2000; Brown, 1994; Koenig et al., 2001; Miller & Stark, 2002; Pew Research Center, 2016; Stark, 2002; Stark et al., 1997), relying on religion as a coping resource (Dunn & Horgas, 2000; Koenig et al., 2001; Marks, 2005), having religious or spiritual experiences (Back & Bourque, 1970; Cavendish, Welch & Leege, 1998; Underwood & Teresi, 2002) and reporting higher levels of general religiosity (Adlaf & Smart, 1985; Chatters, Levin & Taylor, 1992; Evans et al., 1995; Idler et al., 2009; Koenig et al., 2001; Schnoll, Harlow & Brower, 2000), commitment (Taylor & Chatters, 1991) and involvement (Barna Research Group, 2000; Evans et al., 1995; Idler et al., 2009).

Women also attend religious services more frequently than men. Historical data point to a gender gap in attendance dating back at least to the mid-seventeenth century, where women reportedly outnumbered men in the pews in male-led Puritan congregations (Lummis, 2004). Current national data indicate that the gap between male and female adult attendance fluctuates between 5% and 15% (American National Election Study, 2007; Barna, 2002; Barna Research Group, 2007; Davis & Smith, 2004; Gallup, 1990; Gallup & Castelli, 1989; Lazerwitz, 1961;

Mahoney et al., 2001; Pew Research Center, 2016; Stark et al., 1997). For example, data from a nationally representative sample collected by the Pew Research Center (2016) in 2014 indicate that women (40%) are 8% more likely to attend religious services at least once a week than men (32%). Using data from the 2004 GSS, Davis and Smith (2004) reported a 12% gap between females (54.1%) and males (42.1%) who reported attending once-a-month or more. Similarly, Stark et al. (1997) used GSS data collected from approximately 30,000 adults over a 28-year period (1972 through 1990), and found a 14% attendance gap (51% for females, 37% for males) among those who reported once-a-month or more attendance. Data from the 2004 American National Election Study (ANES) yielded an 8.5% gap in weekly or near weekly attendance between women (57.4%) and men (48.9%; ANES, 2007). Finally, the Barna Research Group collected data from a nationally representative sample of Americans in 1999 and 2000 and found a 10% gap in attendance during the past week between females (45%) and males (35%; Barna Research Group, 2000), but this difference shrunk to 6% (50% for females, 44% for males) in 2006 (Barna Research Group, 2007). Importantly, Presser and Stinson (1998) found that the gender gap in attendance did not change when they compared results from the direct Gallup item with results from the time-use item. Regardless of methodology, then, it appears that women outnumber men in the pews by approximately 10%.

Although the gender gap in attendance is nearly ubiquitous among general population samples, there is evidence to suggest that the gap varies for a few subsets of the population (see Hoge & Roozen, 1979). For example, de Vaus (1984) found that although the traditional gender gap existed for those in low and medium status jobs (regular attendance rates for males and females in "laborer and service worker" jobs were 39% and 47%, respectively), the gap shrunk considerably for those in high status jobs (49% and 53% for males and females in "professional and managerial" jobs, respectively), and actually reversed for those with graduate degrees (56% and 53% for males and females, respectively). The Pew Research Center (2016) also found that the gender gap shrinks when controlling for employment status, with full-time employed women being less likely to attend than part-time or unemployed women.

Additionally, Van Tubergen (2006) analyzed data from a sample of immigrants, and found that women were more likely to be affiliated with a religion; but, among the affiliated, men were more likely to attend religious services. Similarly, Lazerwitz (1970) analyzed data from a random sample of Jewish men and women living in or near Chicago, and found that men attended at a higher rate than women. Interestingly, this latter attendance gap was greater among first- and second-generation immigrants than it was among third-generation immigrants. Relatedly, the Pew Research Center (2016) found that while Christian women were more likely to attend than Christian men, the pattern is actually reversed among Orthodox Jews and Muslims. In contrast, Mueller and Johnson (1975) used data from a nationally representative sample of Jewish men and women, but found the typical general population pattern of attendance (i.e., women attended synagogue more often than men). In making sense of these conflicting findings, Hoge and Roozen (1979) suggested that it might take a few generations before the American cultural pattern of attendance begins to emerge among U.S. immigrants. For most Americans, however, the female-dominated pattern of attendance is the norm.

Theoretical Explanations for the Gender Gap in Attendance

A number of theories have been offered to explain the gender gap in attendance (e.g., see Batson et al., 1993; Lummis, 2004; Miller & Hoffmann, 1995; Pargament, 1997; Stark, 2002), but most of these have failed to garner empirical support. Arousal theory, on the other hand, as received some empirical support. Arousal theory holds that individuals vary with respect to their natural need for neurological stimulation (Cochran, Wood & Arneklev, 1994; Baier & Wright, 2001). According to Cochran et al. (1994, p. 95), sub-optimally aroused persons "tend to be highly impulsive, risk-taking thrill seekers" … [who] are unlikely to find [normative social activities, including] religion and/or religious services neurologically satisfying" (see also Baier & Wright, 2001, p. 5; Miller & Hoffman, 1995). Empirical research has provided some support for this latter contention. Specifically, religious importance and attendance are negatively related to risky behaviors such as drunk driving, riding with a drunk driver, infrequent seat belt use, alcohol, tobacco and drug use, binge drinking, premarital and extramarital sex, truancy, carrying weapons to school, property damage and interpersonal violence (Cochran et al., 1994; Koenig et al., 2001; Wallace & Forman, 1998). Furthermore, the rejection of religion in-and-of itself has been characterized as a form of risk-taking given that it reflects a potential incurment of supernatural sanctions and the loss of supernatural rewards (e.g., see Miller & Hoffman, 1995).

Importantly, men tend to be greater risk takers than women (Sherkat, 2002). In fact, Miller and Hoffman (1995) state that "females perceive greater risk in virtually all aspects of life, and thus are less likely to have risk-taking attitudes and … [engage in risky] behaviors" (p. 65). For example, females are less likely than males to make risky financial investments (Powell and Ansic, 1997); avoid wearing seatbelts, speed while driving and drive without a license (Stark, 2002); skip school and miss work (Stark, 2002); approve of premarital sex (Earle & Perricone, 1986), engage in casual sex, risky sexual behavior and extramarital sex (Baldwin & Baldwin, 1988; Martins, Tavares, Lobo, Galetti and Gentil, 2004; Poppen, 1995; Stark, 2002); abuse alcohol, (Martins et al., 2004); use tobacco and drugs (Cochran et al., 1994; Stark, 2002); gamble compulsively (Stark, 2002) and engage in crime and delinquency (Evans, Cullen, Dunaway & Burton, 1995; Koenig et al., 2001; Stark, 2002). Thus, it may be that men are underrepresented in the pews because they perceive religion and religious services as too “safe” and “non-stimulating” to hold their interest.

Just a couple of studies have investigated the effect of controlling for risk preference on the religious gender gap. Miller and Hoffman (1995) used data from 2,408 high school seniors who were participating in the Monitoring the Future study. As expected, Miller and Hoffman found that, compared to males, females rated religion as more important in their lives while scoring lower on a two-item measure of risk preference. When risk preference was controlled for the coefficient representing the relationship between gender and religious importance was still significant, but was reduced by almost 40%. In addition, when Miller and Hoffman tested a structural equation model with these same data, they found that gender had both a direct and an indirect effect (via risk preference) upon religious importance. In a follow-up study, Freese (2004) used data from 1,400 adults who participated in the U.S. version of the 1990 – 1993 World Values Survey, and employed four dependent variables: religious affiliation status, religious service attendance, importance of religion in one’s life and finding comfort and strength in one’s religion. Again, women demonstrated more favorable forms of religiosity on all four indicators.[38] When risk preference was included in the respective models, however, each of the gender coefficients was reduced, but only by a median of 5.6% (range = 2.3% to 8.5%). Thus, while risk preference may indeed moderate the gender gap in religiosity, the magnitude of the effect is still in question. Furthermore, it’s important to remember that at least 60% of the gender gap in religiosity remained after controlling for risk preference in these two studies. Thus, risk preference is, at best, just one piece of the puzzle.

[38] Gender comparisons on the risk preference variable were not provided.

A more promising explanation for the gender gap in religiosity stems from gender orientation theory.[39] According to Bem (1981a, 2010), gender orientation consists of two distinct sex-types: feminine and masculine. The former sex-type refers to stereotypically feminine characteristics (e.g., affection, gentleness, understanding, sensitivity to the needs of others), whereas the latter refers to stereotypically masculine characteristics (e.g., ambition, self-reliance, independence, assertiveness). Both of these orientations are theorized to develop independently of gender[40] as a result of socialization and cultural norm assimilation processes (see Bem, 1981a). Accordingly, measures of gender orientation (e.g., the Bem Sex Role Inventory; Bem, 1981b) allow males and females to vary in the extent to which they identify with *both* the feminine and masculine orientations.

Importantly, there is evidence that the feminine orientation is more closely tied with religious outcomes than the masculine orientation. For example, compared to males and females who score high on the masculine orientation, those who score high on the feminine orientation tend to be more spiritual (Herman, 1996; cited in Thompson & Remmes, 2002), religious (Thompson, 1991; Thompson & Remmes, 2002), religiously committed (Thompson & Remmes, 2002; Francis et al., 2001; cited in Lummis, 2004), religiously involved (Francis & Wilcox, 1996, 1998; cited in Thompson & Remmes, 2002), have fewer religious doubts (Thompson & Remmes, 2002), have more favorable attitudes toward Christianity (Francis & Wilcox, 1998), engage in private religious activity and attend religious services more often (Thompson & Remmes, 2002) and are more likely to view religion as an end (instead of a means to other personal or social ends; Thompson & Remmes, 2002). Thus, the feminine orientation clearly

[39] Although gender orientation theory has been replaced in the literature by gender role theory (e.g., see Eagly, 2009), the relevant research on gender and religion utilizes the former theoretical framework.
[40] Men and women do typically score higher on their respective sex-typed factors, however (e.g., see Choi, Fuqua & Newman, 2009).

seems to be linked with greater levels of religiosity and religious involvement than the masculine orientation.

Even more important, a limited amount of research suggests that when gender and gender orientation are included in the same model, gender loses most of its predictive power, leaving gender orientation as the primary (or sole) predictor of the religious outcome in question. For example, Francis and Wilcox (1998) collected data from two samples of secondary students in the U.K. The first sample consisted of 687 ninth and tenth graders (aged 13 to 15), and the second consisted of 292 students attending a religious studies program (aged 16 to 18). Among both groups of adolescents, Francis and Wilcox found that females had more favorable attitudes toward Christianity. When gender orientation was added to the model, however, gender was no longer a significant predictor of attitudes toward Christianity. Similarly, Francis and Wilcox (1996) collected data from 159 college students in Wales, and found that gender was unable to explain additional variance in attitudes toward Christianity when added to a model that already contained the masculine and feminine orientations.

In the only known study involving an American population, Thompson (1991) collected data on five religious indicators from a convenience sample of 358 undergraduate students attending one of three colleges in the northeast. Gender initially predicted each of the five religious indicators (self-reported religiosity, religious devotionalism, general religious beliefs, orthodox religious beliefs and religious behavior). After adding the feminine and masculine orientations, however, the coefficient for gender became insignificant in each model, while the coefficient for feminine orientation became stronger.[41] Rather than discarding gender as an explanatory variable, however, Thompson noted that it still exerted an indirect effect on the religious indicators given the close tie between gender and gender orientation (i.e., 69% of

[41] The coefficient for masculine orientation was insignificant in both the partial and full models.

women scored above the mean on femininity, whereas just 39% of men did so). Although there is limited research on this topic (and even less that utilizes data from American samples), these findings indicate that the cognitive schemas people hold about how men and women are supposed to think, feel and behave may be more important than their biological sex in the prediction of religious outcomes.

Summary. Studies have demonstrated that both arousal theory and gender orientation theory are able to account for at least part of the gender gap in religiosity. Gender orientation theory, however, seems to have more explanatory power than arousal theory given that the former accounted for 100% of the gender gap, whereas the latter accounted for less than half of the gap. Both areas are understudied, however, with just a handful of studies devoted to investigating the explanatory power of these two respective theories. Thus, further research is needed to verify the findings reported here. In addition, research is needed that brings both the arousal and gender orientation theory variables together to determine whether they can jointly account for variance in the gender gap (directly or indirectly) or if one theory is, in fact, superior to the other.

Even if future research is able to account for the gender gap, however, it will be important to determine whether gender continues to exert indirect effects on religious outcomes. This was the case for gender orientation (e.g., females were more likely than males to score high on the feminine orientation), and it will likely be the case for other explanatory variables (e.g., because males are more likely to be risk takers, and risk takers are less likely to participate in religion, males are less likely to participate in religion). For now, then, gender remains an important variable in understanding variation in religious service attendance.

Hypothesis 5. Studies with higher percentages of female participants will yield higher attendance estimates, on average, than studies with higher percentages of male participants.

Race and Ethnicity

Race and ethnicity[42] is perhaps the only demographic variable that is able to produce a gap in religiosity and religious service attendance that rivals that of gender in terms of both its size and consistency. This gap, however, is largely limited to comparisons between White Americans and African Americans, with the latter group generally evidencing higher levels of religiosity and spirituality. In the literature review for this paper, a total of 55 comparisons were found on the religiosity and spirituality of African Americans and White Americans. In 42 (76.4%) of these comparisons, African Americans demonstrated significantly higher levels of religiosity and spirituality than White Americans. Example religious and spiritual outcomes where African Americans exceeded White Americans include: holding religion as important in their lives (Cameron, 1969; cited in Batson et al., 1993; McConahay, 1970; Wallace & Forman, 1998), religious involvement (Chatters & Taylor, 1994) and affiliation (Briggs, Malek, Davis, Davis & Lorentz, 2009; Wallace & Forman, 1998), holding traditional religious beliefs (Glenn, 1964; Sasaki, 1979), feeling strongly about those beliefs (Sasaki, 1979), daily devotional activities (Idler et al., 2009), daily prayer (Stark et al., 1997), belief that prayer and faith in God are able to bring about recovery from illness (Ell & Haywood, 1985), using prayer as a coping strategy (Dunn & Horgas, 2000), watching or listening to religious media (Levin, Taylor & Chatters, 1994; Taylor, Chatters, Jayakody & Levin, 1996), having religious or spiritual experiences (Back & Bourque, 1970; Sasaki, 1979; Underwood & Teresi, 2002), subjective religiosity (Chatters et al., 1992; Chatters & Taylor, 1994; Eliassen, Taylor & Lloyd, 2004), devotionalism (Cavendish et al., 1998) and organizational and non-organizational religiosity (Chatters et al., 1992).

[42] Race and ethnicity were treated as a single variable in this study.

Much less comparative data are available for other American ethnic groups. Mickley and Soecken (1993) found that Hispanic Americans scored higher than White Americans on a measure of intrinsic religious motivation (i.e., where religion is a master motive in one's life; see Allport, 1950; Allport & Ross, 1967). In a study of 576 open heart surgery patients, Idler et al. (2009) found no differences between Hispanic Americans and White Americans across ten behavioral, cognitive and affective indicators of religiousness, but did find that "Other" ethnic respondents (i.e., not African American, Hispanic or White) were more likely to participate in daily devotional activities than White Americans. Ell and Haywood (1985) studied a group of patients recovering from myocardial infarction, and found that Hispanic and African American patients were more likely than White patients to believe that their recovery was up to God; no difference was found between the Hispanic and African American patients. Wallace and Forman (1998), on the other hand, reported that African American youth were more likely than either Hispanic or White American youth to be affiliated with a religion and to hold religion as important in their lives. Data from the 2001 American Religious Identification Survey indicate that African Americans (85%) report higher rates of affiliation with Christianity than non-Hispanic Whites (79%), while American Indians (66%) report lower levels of affiliation (Garroutte et al., 2014). In a study of over 600 first-year college students, Briggs et al. (2009) found that religious affiliation rates were highest for African American students (87.1%), followed by Asian American students (84.2%), White students (78.4%) and Hispanic American students (77.7%). The sample sizes for the Asian and Hispanic American students, however, were both below 20; hence, solid conclusions about the relative religiosity of these groups cannot be drawn from this study.

A few studies have also compared the religiosity of one ethnic group with that of all other ethnic groups. For example, Brown (1994) reported that non-Whites were more likely than Whites to stress the importance of daily prayer. Koenig et al. (2001) noted that African Americans were more likely than persons from other ethnic groups to hold fundamentalist religious beliefs. Using a sample of nearly 3,000 older adults (average age = 73 years), Benjamins, Musick, Gold and George (2003) found that African American elders were more likely to report watching or listening to religious media than elders from other ethnic groups. Other studies report that African Americans are more likely than individuals from other ethnic groups to rely on religion (Koenig et al., 2001), and to report higher levels of subjective (Pargament, 1997), public and private religiosity (Bourjolly, 1998).

Turning to religious service attendance, we again see that African Americans attend at a higher rate than all other ethnic groups (Briggs et al., 2009; Davis & Smith, 2004; Iannaccone, 1998, 2003; Lazerwitz, 1961; McConahay, 1970; Mitchell & Weatherly, 2000; Musick, 1996; Sasaki, 1979; Stark et al., 1997; Wallace & Forman, 1998). Stark et al. (1997) claim that there is a 12% gap in once-a-month or more attendance between African American and White adults, with several additional studies supporting this contention (Briggs et al., 2009; Davis & Smith, 2004; Lazerwitz, 1961). In an older study, Lazerwitz (1961) combined data from three national surveys conducted in 1957 and 1958 by the Survey Research Center at the University of Michigan, and found that 84% of African American Baptists reported attending religious services regularly or often, while 69% of White Baptists reported doing so. In a more recent study, Briggs et al. (2009) found that 30.6% of African American first-year college students, 18.9% of White students and 14.8% of students from other ethnic groups reported attending religious services nearly every week or more. Nationally representative data from the GSS also reveal a 12% gap in attendance between African Americans and White Americans. Specifically, 43.2% of African Americans in the GSS reported attending religious services about once-a-week or more, while 31.1% of Whites and 29.9% of all other ethnic groups reported doing so (Davis & Smith, 2004).

Despite the consistency across the four studies noted above, two other nationally representative surveys found a much smaller gap. Specifically, the Barna Research Group (2007) found that 52% of African Americans reported attending religious services in the past week, whereas 49% of White Americans, 41% of Hispanic Americans and 29% of Asian Americans reported doing so. The 2004 ANES data also revealed negligible differences among African Americans (54%), Whites (53.6%) and Others (53.3%) who reported attending almost every week or more (ANES, 2007).

Although the latter two studies found relatively small gaps in attendance between African Americans and Whites, the majority of the research on ethnic differences in attendance suggests that there are consistent and significant differences between these groups. Moreover, Presser and Stinson (1998) found that the difference in attendance between African Americans and members of other ethnic groups was invariant across both the direct Gallup item and the indirect time-use item. Thus, it seems clear that African Americans attend religious services at a more frequent rate than White Americans, but the exact difference between these two groups is somewhat less certain. Furthermore, very little data are available regarding the attendance patterns of other relatively large American ethnic groups such as Hispanic and Asian Americans. Typically, these ethnic groups are lumped together with other minority groups when making comparisons, the results of which tend to suggest that *other* ethnic groups attend less frequently than White Americans, albeit not always significantly. This latter finding is in contrast with the research cited above on other religious indicators (e.g., intrinsic religious motivation, religious beliefs about health and the importance of prayer). Accordingly, it will be important for future research to draw samples that adequately represent other ethnic groups in order to gain a more accurate understanding of their attendance patterns. Such efforts would not only help inform the literature on how each ethnic group stacks up on a variety of religious indicators, including attendance, but would also provide the basis for the development of theory to explain any differences that might emerge.

Theoretical Explanations for the Attendance Gap Between African and White Americans

The origin of the attendance gap between African and White Americans dates back to the beginning of the Black Church in America. Keller (2000) reports that as slaves were brought to America, they were evangelized by White Americans, but, because of inimical treatment both within (e.g., being forced to sit in the back or balcony of a church) and outside of the Church, they eventually formed their own denominations and congregations. The Church then "became the one arena in which Blacks could exercise leadership without White interference" (Keller, 2000, p. 33). Because the Church provided this escape from the wider world, it not only served a religious function, but also began to take on a number of additional psychosocial, communal, civic and political functions. Mattis and Jagers (2001), for example, noted that the theology of the early Black Church, which grew out of, and reflected their struggle against the larger society, became primarily concerned with questions of oppression, liberation, hope, love and justice. This theological focus helped those living during slavery and the pre-Civil Rights era to develop a framework for understanding and coping with their plight (e.g., viewing themselves as modern-day Israelites in captivity, and believing that God will eventually grant them victory over their oppressors), to allow members of the Black Church to gain a sense of belonging and self-worth and to establish cohesive communities (Cavendish, Welch & Leege, 1998; Coleman & Holzemer, 1999; Thompson & McCrae, 2001). Over time, the Black Church also became a center for activities that are typically handled in White communities by civic, educational, health, social and political organizations (Armstrong & Crowther, 2002; Batson et al., 1993; Coleman & Holzemer, 1999; Jackson & Reddick, 1999; Taylor, Ellison, Chatters, Levin & Lincoln, 2000), with the most powerful example being the vital role that the Church played during the Civil Rights Movement. Given the central role that the Church has played in the lives of African Americans for the past few centuries, it is now widely believed to be *the* "central institutional pillar [next to the family] in the African American community" (Mattis et al., 2004, p. 126). Thus, it should also be no surprise that African Americans are more likely to participate in the life of their local congregation than are White Americans.

In addition to, and perhaps because of, the central role that the Church has played in the African American community, there is also some evidence that African Americans tend to get more out of their religious experience than White Americans. For example, Koenig et al. (2001) reports that religious participation is a greater determinant of well-being for African Americans than Whites, and, conversely, that the absence of a religious affiliation is related to depressive symptoms among African Americans, but not among Whites. Elderly African Americans have also been found to have greater life satisfaction than elderly Whites, reportedly because of their greater contact with church-related friends (Koenig et al., 2001). In a study of caregivers, African Americans reported receiving more comfort from religion and perceived higher levels of reward in their work (e.g., positive feelings, outcomes of caregiving) than White caregivers (Picot et al., 1997; cited in Dunn & Horgas, 2000). In addition, a number of studies have linked religious involvement with lower levels of blood pressure (see Koenig et al., 2001), with some finding that the relationship holds especially (Koenig, George, Cohen et al., 1998a; cited in Koenig et al., 2001) or only (Koenig et al., 1998; cited in Seeman, Dubin & Seeman, 2003) true for African Americans. One reason that African Americans may garner more biopsychosocial benefits from religious participation than Whites is that they are not only more likely than Whites to participate in religious activities, but they are also more likely to pray and engage in other forms of private religious behavior (e.g., Bible study, watching or listening to religious media; Dunn & Horgas, 2000; Ellison, 1991b; cited in Ellison, 1992) that are important in coping (Pargament, 1997), the development of hope (Coleman & Holzemer, 1999) and the establishment and maintenance of social support networks (Jang & Johnson, 2004).

Summary. Although there are no known studies that have directly linked the role of the Black Church in the African American community to participation rates, there is a preponderance of theoretical work to suggest that this is the primary reason why African Americans attend more frequently than White Americans. In addition, there is some evidence to suggest that attending religious services is more beneficial in terms of health and well-being for African Americans than it is for White Americans. Given these conditions, Sherkat (1997, p. 75) has described African Americans' religious choices as "semi-involuntary," suggesting that it is actually very difficult for African Americans "to leave the Church or to cease participating in church activities."

Hypothesis 6. The percentage of African American participants in studies will be positively related to the attendance estimates that these studies yield.

Age

The data on religious service attendance and age provide a more nuanced picture than what we saw for gender and race and ethnicity. With a few exceptions (e.g., ANES, 2007; Taylor, Chatters, Mattis & Joe, 2010), samples that cross-sectionally represent the age spectrum find *non*-significant (or at least attenuated) associations with attendance (e.g., Hoge & Roozen, 1979; Lazerwitz, 1961). Yet, there are clear and consistent trends in attendance if we look within certain periods of the lifespan. Below, I explore attendance patterns within four periods of life: adolescence, young adulthood, middle adulthood and late adulthood.

Adolescence (Ages 13 to 17). Most research on adolescent attendance suggests that it initially mirrors that of the adult population (Child Trends Databank, 2007; Smith et al., 2002; Wallace & Forman, 1998). For example, data from the Monitoring the Future Survey (MFS), a nationally representative survey of students in eighth, tenth and twelfth grades conducted every year from 1991[43] through 2004, suggests that approximately 44% of eighth graders[44] attend religious services at least once a week (see Figure 3; Child Trends Databank, 2007). As youth advance in age, however, their attendance levels become less frequent (Child Trends Databank, 2007; Regnerus, 2003; Regnerus & Uecker, 2006; Smith et al., 2002; Wallace & Forman, 1998). For instance, MFS data indicate that approximately 39% of tenth graders and just 32% of twelfth graders[45] attend at least once a week (Child Trends Databank, 2007). If the MFS data are, in fact, representative, then adolescence is characterized by a sizeable decrease (by over 10%) in the frequency of religious service attendance.

There are some data that qualify this conclusion, however. Specifically, Regnerus and Uecker (2006) used longitudinal data collected from 11,415 adolescents (7th through 12th grades) who participated in the National Longitudinal Study of Adolescent Health (Add Health), and found that children with more educated parents were the least likely to show changes in

---

[43] Monitoring the Future data on twelfth grade students dates back to 1975.
[44] This is the median weekly attendance rate for eight graders between 1991 and 2004 (range = 42.1% to 46.8%).
[45] These are the median weekly attendance rates for tenth (range = 36.5% to 42.1%) and twelfth graders (range = 30.8% to 34.9%), respectively, between 1991 and 2004.

attendance rates over a one-year period. Interpreting this finding, Regnerus and Uecker suggested that parental education played a stabilizing role in the lives of their children by providing them with an ability to be skeptical of claims made by those at the extreme ends of the religious spectrum. Thus, it appears that although adolescence is characterized by a decline in attendance, there is at least one variable (i.e., parental education) that may help offset this decline.

Another interesting observation is that while attendance may decline through adolescence, other measures of religiosity tend to remain constant or even increase through this period. For example, King, Elder and Whitbeck (1997) conducted a longitudinal study of high schoolers in Iowa, and found that while attendance decreased, involvement in other church-related activities actually increased. Additionally, Regnerus (2003) analyzed nationally representative data on over 11,000 adolescents in grades seven through twelve who participated in the Add Health study, and found that while religious service attendance decreased significantly with age, a latent measure of religiosity (importance of religion in one's life, frequency of prayer, frequency of attendance and frequency of attendance at church youth activities) remained constant. Thus, it may be that while attendance decreases in frequency as youth advance through adolescence, their religiosity, per se, remains the same or even intensifies.

Young Adulthood (Ages 18 to 30). Although the religious service attendance data for young adults is somewhat inconsistent, the majority of the available research suggests that attendance remains relatively low (and may even decline) through this period of life. For example, Presser and Stinson (1998) found that 30% and 36% of young adults between the ages of 19 and 28 reported regular attendance in response to the original GSS and Gallup items, respectively, while 21% reported attending in response to the time-use item. Recall, however, that the respective rates for all adults are greater than 40% in response to the original GSS and Gallup items (Davis & Smith, 2004; Gallup Organization, 2007), and between 26% and 29% in response to the time-use item (Presser & Stinson, 1998). Thus, these data indicate that attendance among young adults may be 5% to 10% lower than it is in the general population.

Several other studies corroborate Presser and Stinson's (1998) findings. Cross-sectional data from the 2004 ANES, for example, indicate that attendance may decline through the early thirties (see Figure 4). Specifically, the ANES data reveal attendance rates of 42.1% for 18 to 24 year-olds, 37.3% for 25 to 29 year-olds and 35.2% for 30 to 34 year-olds, but for those age 35 and over, the attendance rate is markedly higher (median = 57.6%; ANES, 2007). Similarly, data from the 1979-1980 National Survey of Black Americans (NSBA) indicates that attendance is at its lowest point for those between the ages of 25 to 34, but then gradually increases through the age of 74 (see Figure 5; Chatters & Taylor, 1989). In a nationally representative study of 784 immigrants to the U.S., Cadge and Ecklund (2006) found that young adults (under age 25) were less likely to attend religious services than older adults, but significantly so only when compared with those between the ages of 36 and 45.[46] Toussaint, Williams, Musick and Everson (2001), who used data collected from a nationally representative sample, found that not only do young adults attend less frequently than middle-aged and older adults, they also pray less and report lower levels of religiosity and spirituality. In addition, Hoge, Johnson and Luidens (1993) asked 500 middle-aged (aged 33 to 42 years) who had been confirmed in the Presbyterian Church as teenagers to retrospectively report trends in their attendance (thereby ruling out cohort effects). Three-quarters of the respondents said that they had "dropped out" (i.e., attended less than six

[46] Cadge and Ecklund (2006) used data from the 12-month follow-up New Immigrant Survey-Pilot (NIS-P), but failed to provide detailed information on the sample's religious affiliation composition. Jasso, Massey, Rosenzweig and Smith (2002) provide a proxy, however, as they reported the religious affiliation breakdowns for the 976 immigrants who participated in the *baseline* NIS-P: Christian (64.7%), No Religious Affiliation (15.0%), Muslim (8.0%), Buddhist (4.0%), Hindu (3.4%), Jewish (2.6%), other (1.4%) and no response (1.2%).

times a year) at some point since being confirmed, and nearly 70% of these individuals reported dropping out between the ages of 17 and 22.

There are also longitudinal data that corroborate a general decline in attendance among young adults. Specifically, Hoge and Petrillo (1978a, 1978b; cited in O'Connor, Hoge & Alexander, 2002) collected religious data during the 1970s from a group of 451 religiously involved Catholic, Baptist and Methodist tenth graders (representing 35 congregations) living in the Maryland suburbs of Washington, D.C. Twenty-two years later, O'Connor et al. (2002) followed up with 285 of the original respondents, and found that attendance rates had declined between the ages of 16 and 38 for members of all three denominations.

Despite the consistently low attendance rates reported above for the young adult group, a few studies have found different results. Data from the 2004 GSS, for example, indicate that there are only trivial differences in attendance between those aged 18 to 30 (33.5%), 31 to 44 (31.2%) and 45 to 64 (33.9%; Davis & Smith, 2004). Additionally, Ferraro and Kelley-Moore (2001) used a sample of 2,869 adults age 25 and over from the Americans Changing Lives Survey (ACLS), and found that younger individuals were actually more likely to increase their attendance over a two-and-a-half year period (Wave I was conducted in 1986), even after including controls for non-organizational religious behavior, religious salience and affiliation, health status, mental health history and frequency of participation in voluntary activities. Because this latter finding is not broken down across age groups, however, it could just reflect a common trend in attendance as people move into middle adulthood (see below).

Additionally, Lazerwitz (1961) reported findings that were more nuanced than the other studies included in this review. Specifically, Lazerwitz summarized data collected from Protestants, Catholics and Jews in three national surveys that were conducted between 1957 and

1958 (see Figure 6). Among Protestants, Lazerwitz found that the attendance rate among 21 to 24 year-olds (27%) was relatively low, but this rate jumped by 10% among 25 to 29 year-olds (37%), which was still lower, but more in line with the attendance rate for middle-aged and older adults (median attendance = 40%). Among Catholics, there were trivial differences between young adults (21 to 24 year-olds = 70%, 25 to 29 year-olds = 69%) and older adults (median attendance = 72%), but these non-differences are likely due to the pre-Vatican II emphasis on church attendance. With regard to Jewish respondents, the attendance data also suggest that young adults (21 to 34 year-olds) attend synagogue at about the same rate as middle-aged adults (35 to 49 year-olds; 6% and 8%, respectively). The attendance rate among Jews age 50 and over, however, was much higher (25%), which Lazerwitz attributed to the higher prevalence of orthodoxy among the older Jewish immigrants. Taken together, these data suggest that differences in attendance rates between young and older adults may exist only for Protestants (and perhaps Jews), although current data are needed to confirm these findings.

Although there are some contradictory findings here, the majority of the attendance data for young adults suggests that this age group attends less frequently than their middle-aged and elder counterparts. Additional longitudinal research is needed to determine if these trends are merely due to cohort effects or if attendance actually does remain low or even decline during young adulthood. Furthermore, the study conducted by Lazerwitz (1961) raises an intriguing question: do lifespan trends in attendance vary across religious groups and denominations? Further research is needed to answer this question, and to identify other potential moderators of the age-attendance associations reported here.

Middle Adulthood (Ages 31 to 64). Although both adolescence and young adulthood are characterized by declining rates of religious service attendance, the research on middle adulthood suggests that this is a period where adults "return to the pew." Ferraro and Kelley-Moore (2001), for example, used data from the ACLS to study changes in attendance rates over the adult lifespan. Specifically, two waves of data were collected—the first in 1986 and the second two-and-a-half years later—yielding a sample of 2,869 adults age 25 and over. Ferraro and Kelley-Moore found that younger individuals (i.e., older "young adults" and middle-aged adults) were more likely to increase their religious service attendance over the study period than older adults. Recall, however, that these attendance data were not broken down by age category; if they had been, it would have been possible to determine exactly which age groups saw the largest jumps in attendance.

The 2004 ANES data allow for a more precise determination of when attendance frequency increases (see Figure 4). Specifically, the ANES data indicate that attendance increases markedly for those in their late 30's, with "Almost every week or more" attendance jumping from 35.2% for those in their early thirties (30 to 34 years old) to 50.6% for those in their late thirties (35 to 39 years old; ANES, 2007). The attendance rate then gradually increases for those in their 40's (56.8% for 40 to 44 year-olds, and 57.6% for 45 to 49 year-olds), plateaus for those in their 50's (54.2% for 50 to 54 year-olds, and 54.7% for 55 to 59 year-olds) and spikes again, this time to 70.0%, for those in their early 60's (60 to 64 years old; ANES, 2007). Thus, according to the ANES data, attendance rates spike upward for those in their late 30's and early 60's, but remain relatively stable for those in their 40's and 50's. The relatively high attendance rates yielded by the ANES data may invite a degree of skepticism about the veracity of the data, but given that the ANES methodology is constant across age groups, it may only be the *magnitude* of the attendance rates that is questionable, with the *trends* being fairly representative of the underlying variation in attendance across the lifespan. In fact, cross-sectional data from the 1979-1980 NSBA verify that attendance does, in fact, begin to increase during the late thirties (see Figure 5; Chatters & Taylor, 1989).

Data from Lazerwitz (1961) and Albrecht, Cornwall and Cunningham (1988) also suggest that attendance increases during middle adulthood, but only for some religious groups and denominations. Recall that Lazerwitz (1961) used data from three national surveys (conducted in 1957 and 1958) to report on attendance rates across the lifespan for three religious groups: Protestants, Catholics and Jews (see Figure 6). Among Protestants, the rate of regular religious service attendance actually increased during the latter years of young adulthood (27% for 21 to 24 year-olds, 37% for 25 to 29 year-olds), with another upswing during middle adulthood (range = 37% for 40 to 44 year-olds to 42% for 50 to 54 year-olds). As previously mentioned, the data for Catholics revealed a stable attendance rate throughout the adult life-span; hence, there doesn't appear to be any growth in this outcome during mid-life. The data for Jewish respondents also suggests a lack of growth during mid-life. Specifically, 6% and 8% of Jewish respondents aged 21 to 34 and 35 to 49, respectively, reported regular attendance. In a study of over 1,800 Mormons randomly sampled from 27 different wards across the U.S., Albrecht et al. (1988) found that among those who reported periods of infrequent attendance (i.e., they reported period[s] of a year or more where they had attended less than once a month), most returned between the ages of 20 and 35, whereupon the rate of return diminished substantially, but remained positive through age 65. Taken together, these findings suggest that attendance patterns may vary during middle adulthood for some religious groups, but not for others.

Contrary to previous findings, data from the 2004 GSS, which is a pooled dataset of persons from all religious affiliations, indicate that attendance frequency is fairly constant throughout young and middle-adulthood with no apparent upswing in the frequency of this behavior. Recall that the regular attendance rate among young adults (18 to 30 years) was 33.5%; this rate then actually dipped slightly for those aged 31 to 44 (31.2%) before returning to baseline for those aged 45 to 64 (33.9%; Davis & Smith, 2004). If age-attendance trends do, in fact, vary across religious groups, then perhaps these aggregated data mask any affiliation-specific, mid-life attendance upswings that may have been present. Given the general regard given to the GSS, it is important to determine if this is the case, especially considering that the GSS data are in stark contrast to the findings summarized above.

Taken together, the evidence presented here suggests that middle adulthood is characterized by increases in attendance. Yet, it is important to recognize that the data are limited and not all together clear. For instance, one nationally representative dataset yielded no differences in attendance between young and middle adulthood (see Davis & Smith, 2004), while another (although dated) indicated that increases in attendance may hold only for Protestant Christians (see Lazerwitz, 1961). Clearly, more research is needed to obtain a more accurate assessment of the attendance trends that characterize middle adulthood, and to determine whether these trends vary across religious groups and denominations.

Late Adulthood (Age 65 and Over). The literature on attendance in late adulthood predominantly suggests that the elderly attend more frequently than any other age group. For example, the 2004 ANES data (see Figure 4) yield median attendance rates of 39.7% and 54.7% for adults younger than 30 and for those between the ages of 30 and 64, respectively, but the median attendance rate for adults age 65 and over is 72.2% (ANES, 2007). Thus, the ANES data suggest that there is a 15% gap between the young adult and middle-aged attendance rates, and a 17.5% gap between the middle-aged and late adulthood attendance rates. Given that the ANES attendance figures are so much higher than other national surveys, however, it seems wise to question whether the magnitude of the differences between the age groups are accurate. Yet, given that random sampling is used by the ANES, we can be reasonably confident that there are differences across the lifespan.

George Barna (1991) also collected data from a nationally representative sample, and found a large gap in attendance frequency between young adults (aged 18 to 25), middle-aged adults (aged 26 to 64) and those aged 65 and over. Specifically, the attendance rate for the elderly group (69%) was 19% higher than it was for the young adult group (50%), and 17% higher than the middle-aged group (52%). Data from the 2004 GSS (Davis & Smith, 2004) provide further support for a jump in attendance late in life, with 45.3% of those over age 65 and less than 34% of middle-aged adults attending regularly. Citing data from the Gallup Poll, Koenig and Larson (1998) reported a similar disparity. Specifically, Koenig and Larson noted that while 43% of all Americans attend weekly or more, 53% of those aged 65 and over reported doing the same. Lazerwitz (1961), on the other hand, failed to find an increase in attendance among both Protestant and Catholic respondents, but did find a sizeable increase in attendance among Jewish respondents (see Figure 6). Specifically, 25% of Jewish respondents age 50 and over reported regular attendance at synagogue, compared with 6% of those aged 21 to 34 and 8% of those aged 35 to 49. Lazerwitz attributed much of the attendance differences to the higher level of orthodoxy among the older group of respondents, yet it's possible that their age had something to do with the attendance increase. Taken together, the studies cited above indicate that late adulthood is generally associated with a sizeable increase in attendance (of approximately 14%)[47], but more research is needed to determine if these trends are consistent across religious groups and denominations, and to rule out any cohort effects.

[47] Seven attendance comparisons were possible within the studies cited here. The differences ranged from 10% to 18%, with a median of 14.1%.

Looking within late adulthood also reveals an interesting attendance pattern. Specifically, studies that examine attendance within late adulthood generally find that attendance frequency increases through a certain age, after which it tends to fall off slightly, but still remains relatively high. Oman and Reed (1998), for instance, examined attendance patterns in late adulthood using data collected in 1990 and 1991 from 2,025 adults age 55 and over who were living in Marin County, California. They found that the attendance rate for those between the ages of 55 and 64 was the lowest at 20.3%; this rate then rose to its highest point for those between the ages of 65 and 74 (27.6%) before dipping slightly to 26.1% for those aged 75 to 84 and to 23% for those aged 85 or older. Despite the dip in attendance late in life, it is important to note that those who were 85 or older still reported attending at a slightly higher rate than those aged 55 to 64.

Data from the 1979-1980 NSBA show a similar pattern of attendance within late adulthood (see Figure 5; Chatters & Taylor, 1989). Specifically, religious service attendance peaks for both African American males and females between the ages of 65 and 74, before declining slightly for those past the age of 75. It should be noted, however, that the reported attendance rate for those over the age of 75 is still relatively high (i.e., it remains the second highest period of attendance for males, and the fourth highest period for females). Data from the ANES (2007) provide further confirmation of the up-and-down pattern of attendance within late adulthood (see Figure 4). Specifically, for those aged 55 to 59, the "Almost every week" or more attendance rate was 54.7%. This rate then fluctuated between 63.4% and 72.2% for those between the ages of 60 and 74 before dropping off sharply to 56.1% for those aged 75 to 79. Surprisingly, however, the attendance rate then swung dramatically upward to 84.0% for those aged 80 to 84 before taking another downturn to 76.9% for those aged 85 or older. These data indicate that attendance rates follow a more "jagged" pattern than expected in late adulthood, but the general trend is still upward through the latter years of life. The slight downturn in attendance during the last period of life reported in the studies cited here (i.e., ANES, 2007; Chatters & Taylor, 1989; Oman & Reed, 1998) may reflect a decline in health and functional status, but it should be noted that attendance during this period is still higher than in most stages of life.

Even if declining health and functional status limit the ability to participate in religious services, there is evidence that non-organizational indicators of religiosity tend to remain high in late adulthood. Specifically, Idler, Kasl and Hays (2001) analyzed longitudinal data from 2,812 elderly persons participating in the Established Populations for Epidemiologic Studies of the Elderly (EPESE) project (New Haven, Connecticut site),[48] and found that one's sense of religiosity as well as the strength and comfort received from religion does not vary as a function of time-to-death. Moreover, Mindel and Vaughan (1978), in a study of 106 elderly persons living in the Midwest, found that those who had poor functional status were more likely than healthier individuals to participate in non-organizational religious activities such as watching or listening to religious programs on the TV or radio. Similarly, Hays et al. (1998) used longitudinal data collected from a random sample of 2,971 elderly persons who participated in the Duke University site of the EPESE project, and found that the use of religious media remained constant in late life regardless of functional mobility.

In sum, late adulthood seems to be a time of relatively intense religious participation. Several studies (e.g., ANES, 2007; Barna, 1991; Davis & Smith, 2004; Koenig & Larson, 1998) indicate that attendance jumps in late adulthood by approximately 14% over earlier attendance levels. Declining health and functional status may eventually attenuate some of these gains, but even among those close to death, attendance levels remain high. Moreover, as declining

[48] Initial interviews were conducted in 1982, with annual follow-up interviews conducted through 1989 and an additional follow-up conducted in 1994. Re-interview response rates were exceptionally high (between 94% and 96%), which helps eliminate selection threats to internal validity (see Shadish, Cook & Campbell, 2002).

functional status limits participation in organized religious activities, older individuals tend to maintain their inner sense of religiosity (see Idler et al., 2001) and either maintain or increase their non-organizational religious behavior by tuning into religious media programs on TV or on the radio (Hays et al., 1998; Mindel & Vaughan, 1978).

Summary of the Age-Attendance Relationship. Generally, studies that attempt to examine the age-attendance relationship find non-significant (or attenuated) associations. Yet, the two variables *are* related, albeit non-linearly. As we have noted, religious service attendance among adolescents initially mirrors that of the adult population. As youth advance in age, however, their attendance tends to become less frequent. This downward trend lasts through young adulthood, but then increases sharply in middle adulthood. Attendance frequency then appears to plateau during middle adulthood before increasing sharply once again during late adulthood.

Although the age-attendance patterns identified here appear to be real, there are a few notable exceptions and caveats that should be considered. First, Regnerus and Uecker (2006) suggested that parental education can help offset declines in adolescent attendance by providing the latter with the ability to critically analyze extreme (anti-)religious claims made by peers. Further research is needed to verify this finding, and to determine if parental education continues to offset declines in attendance through young adulthood. Second, the 2004 GSS revealed an invariant age-attendance pattern through middle adulthood. This pattern is unexpected, and, given the quality of the data, should be investigated further. Third, Lazerwitz (1961) found invariant attendance patterns through middle adulthood for both Catholic and Jewish respondents. Although dated, this study brings up a potentially interesting interaction with religious affiliation that should be investigated further. Finally, longitudinal research is needed to determine if the age-attendance trends described here are due to cohort effects, or if they represent the actual underlying relationship between these two variables.

Theoretical Explanations for the Age-Attendance Trends in Adolescence and Young Adulthood

As a number of authors have alluded to (e.g., Koenig, 1994b; McGuire, 1981b; Smith, Lundquist Denton, Faris & Regnerus, 2002; Willits & Crider, 1989), the developmental challenges facing adolescents and young adults seem to play a key role in shaping religious behavior. Human development theory suggests that one of the key challenges in late adolescence and young adulthood is the establishment of emotional independence and a sense of personal identity (e.g., see Erickson, 1968; cited in Willits & Crider, 1989). The negotiation of this developmental challenge typically leads teens to distance themselves emotionally and behaviorally from their parents, a process that can involve questioning, critically examining and even rejecting (at least temporarily) their parents' religious beliefs and behaviors (Koenig, 1994b; Ozorak, 1989; Uecker, Regnerus & Vaaler, 2007; Willits & Crider, 1989). While this developmental process often leads adolescents and young adults away from religion, Koenig (1994b) suggests that it is a necessary step if one is to achieve a faith that the individual can call their own (i.e., a faith that is held for intrinsic reasons rather than one that is driven extrinsically by the authority figures in a young person's life). Thus, the decline in attendance during late adolescence and young adulthood may be, for many, a necessary step toward a deeper, more meaningful faith later in life.

Another explanation that should be considered is that adolescents and young adults are not necessarily choosing to avoid religion or diminish their participation, but rather, are facing a variety of transitional issues that are competing for their attention, time and energy. For example, as youth move out of the home, they are likely to begin new jobs or start postsecondary educations, both of which require time and energy and may disproportionately involve (school) work on days when religious services are held (i.e., on the weekends). Youth who are no longer under their parent's supervision are also likely to explore and engage in attractive behaviors that may have been previously unavailable to them (e.g., keeping a late night schedule). In addition, many late adolescents and young adults struggle with their newfound responsibilities. For example, Briggs et al. (2009) found that a commonly cited transition issue for first-year university students was taking responsibility for some of the simple daily tasks that their parents had previously performed for them (e.g., waking up or doing homework on time). It's possible, then, that factors such as an exploration of new freedoms (e.g., staying up late), coupled with a lack of supervision and ownership over one's life (e.g., not getting up on time) contribute to the decline in attendance during this period of life.

In partial support of this line of reasoning, Smith and Lundquist Denton (2005) reported on data from the National Study of Youth and Religion, and found that over half of youth who had once attended religious services at least a few times a year provided mostly passive reasons (e.g., "no reason," "don't know why," "life transition or disruption") for their current level of infrequent or non-attendance. In addition, Uecker et al. (2007) used data from a telephone survey of 500 middle-aged adults (ages ranged from 33 to 42 years) who had been confirmed in the Presbyterian Church, and found that 75% had experienced at least one period of infrequent attendance (i.e., fewer than six times a year) since confirmation. Approximately two-thirds of the reasons given for the periods of infrequent attendance were not anti-religious in nature, but instead, represented a passive, falling away from religious participation (e.g., "left home," "moved away from family," "too busy," "lack of interest," "lazy"). Similarly, in a study of over 1,800 Mormons, Albrecht et al. (1988, p. 68) reported that 54% of those who had disengaged from the Mormon Church for a year or more stated that they had just “found other interests and activities that led them to spend less and less time on church-related activities.” Thus, it’s plausible that as youth transition into adulthood, their declining participation in religious services can be partially attributable to the emergence of new responsibilities and more attractive alternatives, coupled with a passive attitude toward their religious involvement.

Another factor that is tied to the newfound independence that youth encounter when they move out of the home is that they are now free to engage in risky behavior, some of which may be at odds with their religious teachings. In such cases, youth are likely to experience cognitive dissonance created by the disparity between their beliefs and behavior. When dissonance occurs, youth may disengage from religion to avoid being or appearing hypocritical. For example, Albrecht et al. (1988) found that 42% of Mormons cited the difference between their beliefs and behavior as a reason for their non-participation. Additionally, Uecker et al. (2007) used data from Wave I (1994 – 1995) and Wave III (2001 – 2002) of the National Longitudinal Study of Adolescent Health to predict declines in religious service attendance using a number of proscribed behaviors. They found that cohabitation, premarital sex, alcohol use, a change in drinking behavior between waves, marijuana use and the onset of marijuana use by Wave III all predicted declines in attendance. Given these data, then, it seems likely that cognitive dissonance theory could explain a portion of the decline in attendance observed during young adulthood.

Theoretical Explanations for the Age-Attendance Trends in Middle and Late Adulthood

If the transition period between adolescence and young adulthood is characterized by a need to break away from the parental model of religiosity, then the latter years of young adulthood and early years of middle adulthood seem to be characterized by a need to establish and refine a personal religious identity (see Fowler, 1981). The establishment of a personal religious identity involves making an internally-driven religious commitment (Koenig, 1994b), and taking responsibility for one's religious beliefs and behaviors (Fowler, 1981). It follows, then, that individuals who are in the process of establishing their own religious identity would be less likely to offer passive or ambiguous reasons, and more likely to offer substantive reasons for their patterns of religious (non-) participation. Given that individuals typically go through the process of establishing a religious identity during the latter years of young adulthood or early years of middle adulthood, it is likely that this process explains at least a portion of the upswing in attendance observed during middle adulthood.

Middle adulthood is also a time when individuals tend to "slow down," get married, have children and engage in fewer risky behaviors. As summarized in the next section (on *Familial Status*), both marriage and the presence of children are associated with greater rates of religious participation, perhaps because these life events create a need to take on more responsibility and ownership over one's life, but also because parents have a need to provide a religious education for their children (e.g., see Nash & Berger, 1962; Stolzenberg, Blair-Loy & Waite, 1995). In addition, individuals, particularly males, undergo biological changes as they enter middle adulthood (e.g., decreasing testosterone levels; e.g., see Gray, 1991; Seidman, 2003; Simon, Nahoul & Charles, 1996) that are associated with lower levels of risky behavior. For example, the commission rates of most crimes, including alcohol, drug and sex violations, markedly fall off after the age of 30 (e.g., see Steffensmeier, Allan, Harer & Streifel, 1989). Given the inverse relationship previously noted between risky behavior and religious participation (e.g., see Uecker et al., 2007), it seems likely that both life events (e.g., marriage and parenthood) and biological changes explain a portion of the relatively high rates of religious participation observed during middle adulthood.

Another factor that may help explain the increasing rates of attendance observed throughout middle and late adulthood is the accumulation of religious capital. Religious capital is a term introduced by economists (e.g., see Azzi & Ehrenberg, 1975; Iannaccone, 1984, 1990, 1995b) to describe the resources individuals have available to them to "invest" in the production of desired religious outcomes (e.g., religious satisfaction). Iannaccone (1984, 1990, 1995b), and later Stark and Finke (2000), theorized that religious capital includes purchased goods (e.g., transportation, appropriate attire and tithing), time and energy (for engaging in private and public religious activities), human capital (e.g., religious knowledge, familiarity with religious ritual and doctrine, general education and innate skill or ability), interpersonal relationships with co-religionists (i.e., friendships or other social support networks) and emotional attachments (e.g., as a result of bonding with co-religionists, religious or mystical experiences or the performance of religious rituals over time). Together, these assets can be used to "produce" religious satisfaction, but the amount of satisfaction derived depends directly upon the amount and quality of religious capital (available to be) invested. For example, individuals who have transportation and appropriate attire are able to attend religious services, and fit in with co-religionists (from an aesthetic standpoint, at least). Those who are familiar with a particular religion's rituals and doctrine, which may contain symbolism and jargon, are able to glean meaning and knowledge from religious services that might otherwise be missed or misunderstood. Furthermore, those who attend religious services are likely to meet other like-minded individuals with whom they can develop friendships and social support networks. Last, those who participate in public religious activities, or who privately pray, meditate or study religious texts are likely to have emotional experiences from which interpersonal bonds and associations may be formed. Each of these aspects of religious capital, then, play an important role in determining the amount of satisfaction one is able to reap from their religious participation.

Importantly, the production of religious satisfaction has two important consequences for understanding religious participation: (a) the production process naturally leads to the accumulation of additional religious capital (i.e., capital gains); and, (b) the experience of religious satisfaction typically leads to more participation (see Iannaccone, 1990, 1995b). For example, those who invest their time in attending religious services are likely to gain additional knowledge, develop deeper and more meaningful relationships with co-religionists and develop emotional attachments, all of which lead to a more efficient production process (e.g., the establishment of larger social networks allows attendees to gain religious knowledge not only from religious sermons, but also from social interactions) and greater levels of religious satisfaction. It is easy to see, then, that the investment of religious capital and the production of religious satisfaction can become a cycle (or habit; see Iannaccone, 1984) where attendees invest their religious capital, experience religious satisfaction, accrue capital gains, re-invest their capital more efficiently, experience (greater levels of) religious satisfaction and so on. Thus, it is theoretically possible that the elevated attendance levels observed during middle and, especially, late adulthood are due to the cumulative effect of religious investing over the lifespan.

The data available to test the religious capital model, however, are slim. The cross-sectional age-attendance figures presented previously are one source of data available to test the model, but they provide mixed support. Specifically, the age-attendance data provided by Lazerwitz (1961; see Figure 6) for Catholics, Protestants and Jews indicate that only the latter group demonstrates any type of gradual increase in attendance across the lifespan (as predicted by the theory). Even here, however, the most plausible explanation for the increase in attendance among the older Jews is that they are more orthodox (see Lazerwitz, 1961). Data from the 1979-1980 NSBA (see Figure 5; Chatters & Taylor, 1989), however, are more in line with the predictions of the religious capital model. Specifically, the NSBA data for both African American males and females show a consistent growth pattern across the lifespan (starting in middle adulthood for males) that fits nicely with the predictions of the religious capital model. Data from the ANES study (2007; see Figure 4) provide further support. Starting at around age 35, the ANES data show a trend of increasing rates of attendance throughout the remainder of the lifespan. It is worth noting, however, that after age 50, the series has more of a sawtooth pattern, first decreasing, then increasing throughout the remainder of the lifespan, which would suggest that there are additional factors at work besides the processes described in the religious capital model.

Although longitudinal studies are capable of providing more substantive support for the religious capital model, there are only a few studies that have reported on the utility of religious investing as a predictor of future attendance. Iannaccone (1990) reported on data collected from three large national and regional surveys: (1) the 1963 and 1974 Catholic American Surveys (see Greeley et al., 1976); (2) a 1963 survey of Northern California church members (see Glock & Stark, 1966); and, (3) the 1978 to 1987 GSS. The Catholic American Survey data revealed that both a childhood religious education and parental levels of attendance were able to predict future levels of attendance in young adulthood. The other two surveys revealed similar findings, with childhood religious education predicting young adulthood attendance in the church member study, and parental attendance predicting young adult attendance in the GSS. In an independent study, Stolzenberg et al. (1995) tested the predictive utility of the religious capital model using data from the National Longitudinal Study of the High School Class of 1972. Consistent with the findings reported by Iannaccone (1990), Stolzenberg et al. found that religious participation at age 20 predicted church membership at age 22, 25 and 32, but the relationship became weaker over time, again suggesting that there are other factors at work besides the religious capital investment process.

Longitudinal data on older adults is even more limited. In the only study found involving older adults, Strawbridge et al. (1997) followed 6,928 individuals, aged 16 to 94, over a 27-year period. Unfortunately, Strawbridge et al. did not report a coefficient of association between attendance at time one and time two, but they did report that 58% of those who attended once a week or more at baseline were still attending just as frequently 27 years later; and, conversely, 86% of infrequent baseline attenders were still attending less than once a week at follow-up. Despite the tendency for the longitudinal studies to corroborate the religious capital model, the available data are too sparse for an adequate test of the model. Furthermore, the findings reported here could just reflect a selection bias. That is, individuals who gravitate toward religion at time one are expected to remain connected to religion at time two simply because they demonstrated a prior interest. Even so, the gradually increasing rates of attendance from middle through late adulthood (see Figures 4 and 5) suggest that the religious capital model is not without merit.

Another explanation that should be considered is that the high levels of attendance observed in late adulthood could simply reflect differential mortality rates between religious service attenders and non-attenders (see Koenig et al., 2001 for a review). For example, Comstock and Partridge (1972) collected mortality data over a three-to-six-year period following a 1963 census of over 90,000 residents of Washington County, Maryland. They found that infrequent attenders (i.e., those who reported attending less than once a week on the census) were two to four times more likely than frequent attenders to die of arteriosclerotic heart disease,[49] pulmonary emphysema, cirrhosis of the liver and suicide over the follow-up period. In a more recent study, Strawbridge et al. (1997) followed 6,928 people, aged 16 to 94, living in Alameda County over a 28-year period (from 1965 to 1994), and found that frequent attenders (i.e., those who reported attending at least once a week at baseline) had lower all-cause mortality rates than less frequent attenders. Importantly, this relationship held even after the researchers controlled for demographics, health conditions, health practices, body mass index and social connections, and represents a 23% reduction in mortality risk compared to infrequent attenders. Similarly, Musick, House and Williams (2004) used longitudinal data on over 3,600 participants in the 1986 Americans' Changing Lives Survey, and found that those who attended once a month or more had a 30 to 35 percent reduced risk of mortality over a seven-year-follow-up period, even after controlling for demographics, health status, health behaviors, social integration and religious factors and beliefs. McCullough, Hoyt, Larson, Koenig and Thoresen (2000) conducted a meta-analysis on the relationship between mortality and religious involvement,[50] and were able to glean 42 effect sizes from 29 studies. They found that those who were highly involved in religion had survival odds that were 29% greater than the less involved, even after controlling for sample demographics, health behaviors, physical health, social support and a number of study and measurement characteristics.[51] Given that the relationship between religious service attendance and mortality has been shown to be consistently negative and robust,

[49] This outcome applied only to females.

[50] "Religious involvement," as defined in the McCullough et al. (2000) study included a variety of organizational and non-organizational measures of religiosity and religious involvement (e.g., religious service attendance, self-rated orthodoxy, prayer frequency spending spare time in church activities and religious coping).

[51] When only public measures of religious involvement were considered, the highly involved were 43% more likely to survive than the less involved.

it's plausible that the high attendance levels observed in late adulthood are, in part, a function of differential mortality rates between frequent and less frequent attenders, such that the former are comprising a larger percentage of the elderly population simply because of their longevity.

Regardless of the reasons underlying the age-attendance trends, the data are clear in that the middle-aged and elderly attend more frequently than the young. Accordingly, the following hypothesis can be formed:

Hypothesis 7. The average age of a sample will be non-linearly related to religious service attendance estimates, such that studies of young adults (mean age = 18 to 30 years) will yield the lowest estimates, followed by adolescents (13 to 17 years), middle-aged adults (31 to 64 years) and elderly adults (65 years and older), respectively.

Familial Status

Several researchers (Bahr, 1970; Glock, Ringer & Babbie, 1967; Hout & Fischer, 2002; Lazerwitz, 1961; Wilson & Sherkat, 1994) have suggested that levels of religious participation change along with significant events in the family life-cycle. Specifically, they explain that as adolescents and young adults gain independence and eventually leave their family-of-origin, they also tend to leave or disengage from the religion in which they were raised. As these individuals marry and begin forming their own families, however, they tend to re-engage with the religion of their youth. These assertions are consistent with the age-attendance data reviewed above (see Figures 4, 5 and 6). Recall that attendance rates begin at a relatively high rate early in adolescence (when children are still dependent upon their parents), but then decline through the remainder of this period (as they gain independence) and through most of young adulthood (as they leave their family of origin) before increasing again somewhere during the late-twenties to mid-thirties (corresponding loosely with marriage and the presence of school-age children).[52] Thus, the age-attendance data provide tentative support for the family life-cycle model.

A number of empirical studies have also tested the family life-cycle model more directly. Generally, these studies have found support for the family life-cycle model with religious involvement varying in the expected manner as a function of both marital and parental status. Below, I enumerate the evidence for these two variables.

Marital Status. Most of the available empirical data suggest that frequent attenders are more likely to be married and less likely to be separated or divorced than infrequent attenders. For example, Shrum (1980) combined data from six GSS studies (1972 – 1977), and found that frequent attendance (once a month or more) was positively associated with marital status, and negatively associated with a history of separation and divorce, even after controlling for education, family income, age at first marriage and marital cohort. In fact, those who reported attending less than once a year were twice as likely to have a history of separation or divorce (34.5%) than those who reported attending once a month or more (18.1%). Barna (1991) collected data from a nationally representative sample, and found that 57% of married respondents had attended a religious service in the past seven days, while just 39% of unmarried respondents had done so. In a more recent study, Thompson and Remmes (2002) collected data from an elderly sample of 214 men (average age = 75 years) living in Massachusetts, and found that married men attended religious services more often than unmarried men, even after controlling for gender orientation and ideology. Additionally, Koenig et al. (1999) found that frequent attenders were not only more likely to be married, but were also more likely to have

[52] See Stolzenberg et al. (1995). In addition, Census Bureau (2008) data indicates that males and females get married for the first time at an average age of 24.2 and 21.9 years, respectively. The average age when the first child is born is 24.6 (CDC, 2008). Importantly, the average age at which each of these events occurs has increased over the last few decades, a trend that Hout and Fischer (2002) attribute to an increase in the percentage of American young adults pursuing advanced educational degrees.

large social networks and to have confidants within their social networks than less frequent attenders. Thus, the data from these empirical studies suggest that frequent attenders are more likely to enjoy larger social networks, have more close ties within those networks, be married and are less likely to separate or divorce than less frequent attenders.

Several reviews have also reported that frequent attendance is commonly associated with marital stability. Gartner et al. (1991), for example, reported that all five studies in their review found a negative relationship between attendance and divorce. Ellison, Boardman, Williams and Jackson (2001) concluded from their review that religious involvement was negatively related to marital discord, intergenerational conflict and divorce. Chatters et al. (1992) cited three studies that found a positive relationship between religious participation and marital status (Beeghley, Van Velsor & Bock, 1981; Cornwall, 1989; Taylor, 1988b), but also cited one study that found a negative relationship among young adults (Glock et al., 1967). Finally, Mahoney et al. (2001) conducted a meta-analysis, and found that frequent attendance was associated with greater levels of marital satisfaction and commitment, and a lower likelihood of divorce,[53] even after controlling for a number of relevant demographic, marital and familial characteristics.

A number of longitudinal studies have also linked religious service attendance to marital stability. McCarthy (1979), for example, used data from the 1973 National Survey of Family Growth, and found that separation rates after 5, 10 and 15 years of marriage were approximately twice as high among Protestants and over four times as high among Catholics who attended less than once a year than among those who attended once a month or more. Similarly, Fergusson, Horwood and Shannon (1984) found that baseline attendance levels predicted separation rates

---

[53] In an attempt to make the attendance-divorce association more interpretable, Mahoney et al. (2001) made the assumption that 55% of married couples attend religious services at least once a month, and 50% of all marriages end in divorce. Using these assumptions, Mahoney et al. estimated that infrequent attenders would have a divorce rate of 60%, whereas frequent attenders would have a rate of 44%.

five years later, with mothers and fathers who reported never attending being two and three times more likely to separate, respectively, than those reporting at least monthly attendance. Additionally, Booth, Johnson, Branaman and Sica (1995) found that married couples who increased their attendance over a four-year period were less likely than those who did not increase their attendance to experience divorce prone cognitions (e.g., thoughts that the marriage might be in trouble) and behaviors (e.g., consulting with clergy, a counselor or an attorney), although marital satisfaction was unaffected. Finally, Strawbridge et al. (1997) followed a group of 5,286 adults living in Alameda County over a 28-year period, and found that those who reported weekly attendance in 1965 were 79% more likely to be married to the same person 28 years later than those who attended less frequently, even after controlling for demographics, religious affiliation, and a number of health indicators.

Importantly, there is also evidence to suggest that religious involvement increases following marriage. For example, Stolzenberg et al. (1995) used data from the National Longitudinal Study of the High School Class of 1972, which included a sample of 19,000 seniors who were re-interviewed at five different time points after graduation (in 1973, 1974, 1976, 1979 and 1986). Stolzenberg et al. found that getting married was associated with an increase in church membership of about four to eight percent for both men and women in 1976, 1979 and 1986. On the other hand, youth who had cohabitated evidence a reduction in church membership equal to about fifteen percent. In another study, Wilson and Sherkat (1994) used data from a representative sample of over 1,100 high school seniors who participated in three waves of the Youth-Parent Socialization Panel Study (i.e., in 1965, 1973 and 1983 when they were 18, 25 and 35 years old, respectively). While all of the seniors retained in Wilson and Sherkat’s final sample reported a religious affiliation at age 18, those who married prior to 1973 were more likely to continue reporting a religious affiliation at age 25 than those who did not marry. In addition, married respondents who "dropped out" by age 25 (i.e., they no longer reported a religious affiliation) were more likely to re-affiliate by age 35 than unmarried respondents. Interestingly, when Wilson and Sherkat re-examined the marriage-affiliation associations by gender, they found that while women were less likely to drop out in the first place, marriage seemed to only affect men's re-affiliation rates.

Turning to religious service attendance, Williams and Lawler (2001) used nationally representative data on over 1,200 married Christians who were asked to self-report both their current attendance and their attendance at the time of their engagement. They found that respondents reported significantly higher levels of attendance at the time of their interview than at their engagement, suggesting an increase in attendance following marriage. In another study, Petts (2009) used longitudinal data on 2,472 youth who first participated in the National Longitudinal Survey of Youth in 1979 (when they were between 14 and 21 years old), and then again between 1988 and 2004 (when they were at least 20 years old). He found that, for a subset of participants who attended at least moderately throughout adolescence, those who married were more likely to attend at follow-up than youth who did not marry. Conversely, regular attenders who cohabitated were more likely to have lower levels of attendance at follow-up than youth who did not cohabitate. In a study with similar findings, Thornton, Axinn and Hill (1992) interviewed a representative sample of young adults from the Detroit area in both 1980 and 1985 (when they were 18 and 23 years old, respectively). Respondents who had been married between 1980 and 1985 reported significantly higher levels of attendance at follow-up, while those who had cohabitated reported lower levels of attendance. Thus, the available data suggest that not only are frequent attenders likely to have positive marital outcomes, but marriage itself seems to have a boosting effect on levels of religious participation, while cohabitation has an attenuating effect.

Parental Status. Although only a few studies have focused on the relationship between parental status and attendance per se, a number of empirical studies have found positive associations between parental status and religious involvement. For example, in a review of the literature, Hoge and Roozen (1979) reported that parents were more likely than non-parents to become members of a church (Nash & Berger, 1962), participate in the life of a church (Anders, 1955; Glock, Ringer & Babbie, 1967) and attend religious services (Argyle & Beit-Hallahmi, 1975; Carroll and Roozen, 1975; Davis, 1962; Hoge & Carroll, 1978; Lazerwitz, 1961; Metz, 1965; Mueller & Johnson, 1975). In more recent research, Argue, Johnson and White (1999) and Stolzenberg et al. (1995) found that parenthood was positively related to membership in a church or synagogue; and, O'Connor et al. (2002) found that having children was related to greater levels of religious involvement.

A few studies have also focused on irreligious outcomes. For example, Wilson and Sherkat (1994) studied rates of apostasy[54] by following over 1,100 high school seniors over a 16-year period. They found that married fathers were less likely to leave the religion in which they were raised, and were more likely to return if they had left than either unmarried or fatherless males. Similarly, Smith (2006) and Stolzenberg et al. (1995) found that marriage and having children were associated with a return to religion. In another study, Bainbridge (2005) used data from *Survey2001* and the GSS, and found that, among adults age 40 or older, non-parents were more likely to self-report as atheist (3.1%) than same-aged adults with at least one child (2.2%). Furthermore, this rate disparity increased as the number of children increased, and was larger for

[54] Wilson and Sherkat operationally defined apostates as those who self-identified with a religious affiliation at a particular point in time, but then later disavowed any association with religion.

females than males (i.e., men and women without children were 1.3 and 2.7 times as likely to self-report as atheist than were men and women with two or more children, respectively).

Interestingly, some evidence suggests that the relationship between parental status and religious involvement is conditioned by the age of the child(ren). For example, Carroll and Roozen (1975) used a national sample of adults, and found that those who had children younger than age five were less likely to attend religious services, while those with school-age children were more likely to attend than the average American. In another study, Argue et al. (1999) found that parents' religiosity significantly increased when their children were aged two to four, and then again when they were aged five to ten, but the strength of the association between parental status and membership in a church or synagogue didn't reach its zenith until the children were age 10. Similarly, Stolzenberg et al. (1995) used data from the National Longitudinal Study of the High School Class of 1972, and found that, for 32-year-old parents, a child's age has a monotonically increasing effect on rates of parental church membership up until about age five, where the effect peaks until about age eight or nine and then begins to decline as the child approaches adolescence. Thus, it appears that levels of parental religious involvement may not increase immediately (or substantially) upon having children, but only after a time-lag of at least a few years.

Although the studies summarized above all found that parents tend to be more religious and religiously involved than non-parents, a couple of studies have reported different findings. Glock et al. (1967), for instance, found that married persons with children were less likely to participate in organized religious activities than either single persons or married persons without children. Additionally, Cadge and Ecklund (2006) studied a nationally representative sample of 784 U.S. immigrants, and found that individuals living with at least one child at home were less likely to attend religious services than those without children, but this difference was not significant. The average age of the children in both of these studies was not given, however, so it's unclear whether these findings are subject to the interaction noted above where parental attendance doesn't increase until the child is at least several years old.

Despite the findings associated with the latter two studies, the overwhelming majority of the research has found positive associations between parental status and various indicators of religious involvement. Thus, we can tentatively conclude that parents are more likely to be religiously involved and attend religious services than non-parents.

Summary of the Familial Status-Attendance Relationship. The findings summarized here generally support the family life-cycle model. Specifically, the age-attendance data suggest that attendance rates tend to decline from late adolescence through young adulthood before increasing again during the late-twenties to mid-thirties (i.e., the approximate ages at which most people get married and have children). Furthermore, there are clear associations between religious service attendance and marital status, with longitudinal studies suggesting that frequent attendance may increase the likelihood of marital stability and decrease the likelihood of marital discord, separation and divorce. Furthermore, there is evidence from longitudinal studies that marriage is associated with an increase in attendance. The presence of children also appears to be positively associated with attendance. Thus, each of the three areas of research touched on in this section (i.e., age, marital status and parental status) are consistent with the family life-cycle model.

Theoretical Explanations for the Family Life-Cycle-Attendance Trends

Although there is some overlap, the reasons underlying the associations between religious service attendance and both marriage and parenthood are largely different. Accordingly, the rationale underlying each set of associations is discussed separately.

Marital Status. There are likely a number of factors underlying the associations between religious service attendance and marital status. First, it has been suggested that religious persons may simply be more likely to marry than less religious persons. As Koenig et al. (1999) found, frequent attenders tend to have larger social networks (and more confidants within those networks) than less frequent attenders. Thus, the religiously involved may have more opportunities to meet potential marital partners. In addition, the institution of marriage is decidedly bound-up with religion. Roman Catholic doctrine, for instance, specifies that marriage is a sacrament,[55] while other religious groups also view marriage (at least unofficially) as a sacred covenant between marital partners and God (Petts, 2009; Thornton et al., 1992). Conversely, most religions frown upon unions taking place outside of marriage (i.e., cohabitation; Thornton et al., 1992). Accordingly, religious groups are likely to generate numerous messages, both from the pulpit and the congregation, that marriage is desirable, while cohabitation is not. Religiously involved youth, then, are likely to be swayed toward marriage and away from cohabitation, while non-religious youth may hear a more secular message and be swayed in the opposite direction (Petts, 2009; Thornton et al., 1992). While there is not much empirical evidence in this regard, Thornton et al. (1992) used longitudinal data from a random sample of Detroit-area youth, and found that baseline measures of religious importance and attendance were positively associated with marital status five years later. Thus, it seems that at

[55] A sacrament is defined by Merriam-Webster's Dictionary (2011) as "a Christian rite . . . that is believed to have been ordained by Christ and that is held to be a means of divine grace."

least part of the association between attendance and marital status might be due to selection factors, where religious persons are more likely to marry than less religious persons.

Second, the religious pathways taken by married and cohabitating couples must be considered. As mentioned earlier, numerous studies have found that measures of religious involvement and attendance increase following marriage, but decrease following cohabitation (Petts, 2009; Thornton et al., 1992; Williams & Lawler, 2001). Given the close relationship between marriage and religion, the reasons for these religious involvement trends are relatively straightforward. Specifically, those who marry, and have a history of religious involvement are likely to have developed cognitive schemas suggesting that married couples attend religious services (something along the lines of, “It’s just what married couples do”). Furthermore, the reinforcement a couple received prior to marriage should continue (or even increase) now that they are in a relationship that is valued by their congregation. And, as Petts (2009) and Stolzenberg et al. (1995) have noted, religious services can be a good place to meet other young couples who are able to provide emotional and social support for one another. All of these factors, then, likely play a role in the tendency for religious participation to increase after marriage.

On the other hand, young cohabitating couples with a history of religious participation may no longer feel welcomed or accepted in their congregations (Petts, 2009; Stolzenberg et al., 1995; Thornton et al., 1992). These feelings can originate within the self (via the experience of cognitive dissonance), or derive from explicit (e.g., direct confrontation or religious sermons) or implicit (e.g., avoidance) messages from religious leaders and co-religionists (Stolzenberg et al., 1992). Relatedly, Thornton et al. (1992) noted that going against the proscriptions of a religious organization can put couples at odds with their co-religionists, thereby deteriorating important relationships and social support networks. All of these factors, then, are likely to create attendance barriers, while reducing the overall rewards couples might otherwise receive from their involvement. Thus, it seems likely that married and cohabitating couples head down very different religious participation paths that only serve to strengthen the association between attendance and marital status.

Third, religious youth may be likely to seek out marital partners that are similarly religious. Finke (2003) suggests that youth will tend to preserve their religious capital by marrying someone of the same faith. Using a linguistic example, Finke notes that "if one is already proficient in French, one maximizes cultural capital by remaining within a French-speaking community rather than moving and having to invest in learning a new language" (p. 3). In the same way, youth who have developed a bank of religious capital (i.e., religious knowledge; familiarity with rituals, language and symbols characteristic of the religion and emotional attachments to features and people within the religion) are likely to protect (and even try to bolster) their stock of capital by marrying someone of the same ilk. Iannaccone (1990, 1995) supports this contention, stating that shared-faith couples can more efficiently produce "religious commodities" (i.e., religious satisfaction) than split-faith couples given that the resources of shared-faith couples tend to be used complementarily, while those of split-faith couples tend to be used competitively (or at best, neutrally).

Importantly, a number of studies have found that same-faith (or religiously homogamous) couples have better marital outcomes and are more religiously active than split-faith (or religiously heterogamous) couples. Curtis and Ellison (2002; cited in Mahoney et al., 2001), for example, found that religiously heterogamous couples reported more marital distress than homogamous couples, while Call and Heaton (1997), in a study of over 4,500 couples participating in the National Survey of Families and Households, found that heterogamous couples (in this case, wives attended weekly, while husbands never attended) had divorce rates that were higher than those for both religiously homogamous (both partners reported weekly attendance) and non-religiously homogamous (both partners reported never attending) couples. Turning to attendance, Williams and Lawler (2001) used nationally representative longitudinal data on over 1,200 married Christians, and found that couples who were religiously homogamous when they were engaged were more likely than heterogamous couples to attend religious services both at the time of their engagement and at the time of the interview (up to 20 years later). Additionally, Iannaccone (1995) analyzed the effects of religious homogamy and heterogamy on attendance rates using data from two national surveys of American Catholics (conducted in 1963 and 1974) and combined data from the GSS (1972 through 1991). He found that homogamous Catholics attended mass about 12 times more per year than heterogamous Catholics (i.e., where the spouse was a non-Catholic), while homogamous couples in the GSS (representing multiple faiths) attended about 9 times more per year than heterogamous couples. Importantly, Iannaccone's ancillary analyses suggested that religious homogamy may have synergistic effects on attendance. Specifically, he found that the attendance rates for homogamous couples were significantly higher than they were for similarly religious singles, who themselves had higher rates of attendance than religiously heterogamous couples. Given that religious youth are likely to select marital partners with similar religious backgrounds, and that religious homogamy tends to yield attendance rates that are even higher than would be expected given the individual religious participation histories of each partner, it follows that this self-selection factor is likely accounting for part of the attendance rate differences between married and non-married individuals.

Finally, Mahoney et al. (2001), in their meta-analytic review, outlined a number of ways that religious participation can help bring about longer lasting marital relationships. Given that a number of religious organizations treat marriage as a sacred covenant, it is not surprising that the process of participating in a religious marriage, where vows are taken before God and witnesses, can have a sanctifying effect on how individuals (especially those that value religion) view their relationship. Couples who view their marital relationship as sanctified are no longer “loosely” tied together through secular means; rather, they are bound together in a relationship that has spiritual meaning and significance. This perception that one’s marriage is somehow “set apart” or sacred can have far reaching implications for the quality of the relationship. For example, Mahoney et al. (1999, p. 323) analyzed qualitative data from 97 married couples, and found that couples described their marriages with words such as “blessed, holy, [and] heavenly,” while often perceiving God to be manifest in their relationship. Importantly, couples who held these views also reported better communication patterns, less conflict and aggression, more perceived benefits from the marital relationship and greater overall levels of marital satisfaction than those who did not hold these views. In a similar study, Dollahite and Goodman (2006, p. 143) interviewed 32 “highly religious” couples representing 6 religious groups within the Abrahamic faith traditions (Judaism, Islam and four distinct Christian denominational groups), and found that a majority of couples believed that marriage was created by God to fulfill important personal (e.g., marriage provides a means of bettering the self, as well as personal fulfillment and happiness), relational (e.g., marriage was viewed as a means of sharing God’s love for us with others) and spiritual (e.g., couples reported that by engaging in a loving relationship, they were able to draw closer to God) purposes. In addition, all 32 couples perceived that their relationship was better off because God was believed to be either directly (e.g., as a source of guidance, strength, accountability, grace and answered prayer) or indirectly (e.g., through religious texts and teachings) involved in their marriage. Although the results from these qualitative studies (Goodman & Dollahite, 2006; Mahoney et al., 1999) are far from representative, they do highlight some of the pathways by which the perception of a sanctified marriage can alter the way couples relate to one another, and in so doing, affect the quality of the marital relationship.

Besides the sacralizing effect, joint participation in private and organizational religious activities offers opportunities for couples to confront their weaknesses and limitations together, acknowledge their mistakes to one another and to both forgive and be forgiven. In addition, joint religious participation provides opportunities for sharing deeply held beliefs with one another, exploring questions of ultimate concern together and providing (religious) social support for one another. Importantly, all of these activities can help broaden and deepen a couple's level of intimacy and commitment to one another (Mahoney et al., 2001). Furthermore, religiously involved couples may be more likely than couples who are not religiously involved to "minimize or dismiss minor conflicts . . . engage in attributions and behaviors that resolve marital conflict effectively and make greater use of religious coping methods (e.g., prayer, spiritual support)" (Mahoney et al., 2001, p. 90). Scanzoni and Arnett (1987) also found that religious women were able to make more personal sacrifices while remaining satisfied in their marriage than less religious women. And, Brody, Stoneman, Flor & McCrary (1994) observed the interactions of African American couples, and found that self-reported levels of religiosity were positively related to communication quality between marital partners. Given the data reviewed here, then, it seems likely that part of the association between attendance and marital status is due to the relative marital longevity enjoyed by frequent attenders.

In sum, the data reviewed here suggest that: (a) religious youth are more likely to marry (and to marry someone that is similarly religious) than non-religious youth, while the latter are more likely to cohabitate; (b) marriage is associated with an increase in attendance, while cohabitation is associated with a decrease; (c) religious homogamy in marriage produces synergistic effects such that the couple are more religiously active than they would be if they were single; and, (d) religious couples tend to have more satisfying and longer lasting marriages than less religious couples. Given that each of these conditions exacerbate the difference in religious participation between married and non-married persons, we can formulate the following hypothesis:

Hypothesis 8. The percentage of married participants in a study will be positively related to religious service attendance estimates.

Parental Status. The available data on parental status and attendance suggests that parenthood is associated with elevated levels of religious service attendance, and that this is especially true for parents of children who are between the ages of five and ten. Qualitative data help explain both of these observations. Specifically, Nash and Berger (1962) interviewed new members of suburban churches about their reasons for joining, and found that the presence of school-age children was an important determinant in the decision to become a member. Furthermore, the researchers pointed out that many churches provide religious education for children, and a number of parents cited this as a reason for joining. Similarly, Albrecht et al. (1988) analyzed data collected from 1,874 Mormons, and found that over half of those who returned to the Mormon Church after experiencing a period of disengagement did so because they wanted to provide a religious education for their children. Quantitative studies also support these findings. Marty, Rosenberg and Greeley (1968), for instance, used nationally representative survey data and found that a healthy majority of adults believed that a religious education should be provided for their children, and data from the Gallup Poll and Princeton Religion Research Center (1978; cited in McGuire, 1981, p. 46) demonstrate that a religious education for children is a “major consideration” in parents’ decisions to begin attending religious services again. In addition, Becker and Hofmeister (2001) studied a representative sample of upstate New Yorkers, and found a direct effect of having children on the use congregational ministries (e.g., Sunday School). Follow-up, in-depth interviews revealed that the use of congregational ministries was often for the sake of their child’s religious education. Thus, it appears that the provision of a religious education is one of the driving forces behind parents’ decisions to re-engage in the life of a congregation.

Besides the desire to provide a religious education for their children, data from parents indicates that the formation of a family engages a new set of internal motivators that lead them back to the pew. Interestingly, these motivators seem to be different for women and men. Becker and Hofmeister (2001), for example, used representative data from the Religion and Family Project, which included over 1,000 respondents living in four communities in upstate New York. In addition, in-depth interviews were conducted with 70 men and women following the initial phone survey. In their analysis, Becker and Hofmeister found that there was no direct effect of having children on attendance for women, but there was an indirect effect. Specifically, having children seemed to increase the salience of religion for these women, which, in turn, led to higher rates of religious service attendance. In-depth interview data revealed the same; having children seemed to trigger something for these women, such that God and religion were important after having children. As a result, these women began participating again (or more frequently). For men, Becker and Hofmeister’s data indicate that family formation triggers a “provider” role. Religious service attendance was directly associated with parenthood, but in-depth interview data suggested that men just “automatically” began attending again, stating that “they had not really thought about it and could not really articulate why they had returned to church; that it simply seemed “appropriate” and “natural” once they had started a family” (Becker & Hofmeister, 2001, p. 717). Thus, it seems that men and women may have different reasons for increasing their religious involvement after having children. Yet, this is just one study, however, and further research will be needed to verify the validity of this finding. It is important to note, in any case, that the participants in Becker and Hofmeister’s study were able to identify an internal change that accounted for their more frequent religious involvement. This self-awareness of an internal change after having children lends further credibility to the parental status-attendance associations.

In sum, religious service attendance is relatively frequent among parents, especially those of school-age children. While moms and dads may experience different internal motivations for increasing their religious participation, there is a tendency for both parents to want to provide a religious education for their children. Accordingly, the following hypothesis can be formulated:

Hypothesis 9. The percentage of participants who are parents of school-aged children will be positively related to religious service attendance estimates.

Socioeconomic Status

Socioeconomic status (SES) is traditionally composed of three variables: Education, income and occupational status. In practice, however, the SES construct is represented in one of two ways: (a) one or more of the three indicators are measured, and each indicator is analyzed separately; or, (b) two or more indicators are combined to create a composite measure prior to analysis. The former method allows researchers to make specific statements about each indicator, while the latter method allows researchers to make global statements about the SES construct. Although both methods have been used in the attendance literature, the majority of the available research focuses on individual SES indicators. Accordingly, I briefly review the attendance literature as it pertains to composite measures of SES before delving more deeply into the individual SES indicators.[56]

Composite Measures of SES. The available evidence suggests that attendance is either positively related or unrelated to composite measures of SES. For example, Hoge and Roozen (1979) reported a positive relationship between SES and attendance among a sample of Catholics, while Campbell and Fukuyama (1970) found a positive relationship between SES and a measure of organizational religious involvement, which included attendance, among a sample of 8,000 members of the United Church of Christ. Beeghley et al. (1981) combined data from the 1972 to 1978 GSS, and found SES to be positively related to attendance among Methodists, Catholics and White Baptists, but unrelated to attendance among African American Baptists. In addition, Demerath (1965) collected data from members of five mainline denominations and found a positive relationship between SES and a global measure of organizational religious involvement, but no relationship with church attendance. Using data from a nationally representative survey, Stark (1972) reported that SES was positively related to attendance among Catholics, as well as liberal, moderate and conservative Protestants. And, Chatters et al. (1992) and Batson et al. (1993) found SES to be either positively related or unrelated to attendance in virtually every study in their respective reviews. Clearly, this area of research could benefit from additional (and current) work, but perhaps the focus is better placed on individual SES indicators

[56] It should be mentioned that employment status is discussed instead of occupational status because the attendance literature primarily focuses on the former indicator.

given that they provide more specific information about the nature of the SES-attendance relationship.

Education. Studies dating back to the fifties (e.g., Burchinal, 1959), as well as more recent studies (see Chatters et al., 1992; Ferraro & Kelley-Moore, 2001; see Gartner, Larson & Allen, 1991; Iannaccone, 1998; Iannaccone & Everton, 2002; Koenig et al., 1999; Loury, 2004; Mueller & Johnson, 1975; Powell, Shahabi & Thoresen, 2003; Stark et al., 1997; Thompson & Remmes, 2002) have reported positive associations between education and religious service attendance. For example, Burchinal (1959) found a positive relationship between education and attendance using data collected from 512 adults living in the rural Midwest, while Mueller and Johnson (1975) found a small, but positive relationship using data from a nationally representative sample of approximately 1,900 adults. In a more recent study, Ferraro and Kelley-Moore (2001) found a positive relationship using data from a nationally representative sample of nearly 3,000 adults (age 25 and over) who participated in the 1986 ACLS. Koenig et al. (1999) found the same using a sample of nearly 4,000 elderly persons (age 65 and over) living in North Carolina, while Thompson and Remmes (2002) found a positive relationship among a sample of 214 elderly men (average age = 75 years) living in Massachusetts. In addition, Iannaccone and Everton (2002) collected four years' worth of weekly attendance counts from four congregations of varying size on the west coast, and found a positive relationship between education and attendance that was even stronger for those with a history of religious schooling. Importantly, Gunnoe and Moore (2002) suggested that these positive education-attendance associations are found across most of the major religions.

A few studies have also identified longitudinal associations between education and attendance. For example, Loury (2004) reviewed two studies that used data collected from young women (ages 14 to 21) who participated in the 1979 National Longitudinal Study of Youth (NLSY). In the first study, Ribar (1994) found that young women who attended either infrequently or frequently were more likely to complete high school than those who attended at more moderate levels. In a reanalysis of these data, Evans, Oates and Schwab (1992) found that young women who attended at least weekly were less likely to drop out of high school than those who attended less often, regardless of ethnicity, family structure, mother's education and economic disadvantage. Loury (2004) pointed out, however, that these studies failed to rule out important alternative explanations, thereby leaving open the possibility that a "third variable," such as the family's propensity for educational pursuits, might be responsible for the association. In an attempt to rule out this potential confound, Loury reanalyzed the 1979 NLSY data, this time controlling for sibling similarities in educational attainment, as well as individual differences in schooling behavior and aspirations. Despite the introduction of these controls, Loury still found that adolescent attendance was predictive of educational attainment, with frequent attenders being more likely to complete high school and attend college than less frequent attenders. Interpreting the multivariate results, Loury reported that those who indicated never attending in 1979 (19% of the sample) averaged half-a-year less schooling over the 14-year follow-up period than those who reported weekly attendance in 1979 (37% of the sample).

In another study, Muller and Ellison (2001) used longitudinal data from the 1992-94 National Education Longitudinal Study (NELS), and found that a baseline measure of organizational religiosity, which included a measure of attendance, was able to predict higher educational expectations, higher levels of communication with parents about school, greater participation in advanced math courses, greater amounts of study time, lower levels of truancy and higher graduation rates. Additionally, Regnerus and Elder (2003) analyzed longitudinal data from over 9,000 adolescents (grades 7 through 12) who participated in the first two waves of the Add Health study (1994-95 and 1996), and found that baseline attendance was positively associated with remaining academically on-track, even after controlling for 16 demographic, risk and protective factors. In follow-up analyses, Regnerus and Elder found this relationship to be even stronger for children living in impoverished neighborhoods (i.e., high-poverty, frequently-attending adolescents were actually more likely to remain academically on-track than low-poverty, infrequently-attending adolescents), suggesting that religious service attendance may have a compensatory effect for those who have the greatest need. In an alternative approach to these data, Uecker et al. (2007) examined the association between educational achievement and reductions in attendance over the college years using data provided by over 15,000 adolescents and young adults who participated in Waves I (1994-95) and III (2001-02) of the Add Health study. Interestingly, Uecker et al. found an inverse relationship, where college graduates were the least likely to reduce their attendance over the seven-year follow-up period, followed by those who earned an associate's degree, those who were currently attending a four-year school, those who were currently attending a two-year school and, finally, by those who never attended college. This finding persisted even after controlling for a number of demographics, marital and parental status, religious affiliation, sexual behavior and alcohol and drug use.

Despite the consistently positive education-attendance associations noted above, there are a few notable exceptions. For example, Williams and Lawler (2001) used data provided by over 1,200 married adults who participated in a nationally representative sample, and found that while there was a positive relationship between education and attendance in the overall sample, closer inspection revealed that this held only for religiously homogamous couples. Similarly, Cotter and Song (2009) pooled data from the 2003 to 2005 ATUS (n = 5,071 married persons between the ages of 21 and 65), and found that education was positively related to attendance among men, but not among women. More generally, Chatters et al. (1992) used data from a nationally representative sample of 446 elderly African Americans (age 55 and over) who participated in the NSBA, and found no relationship between education and organizational religiosity, which included a measure of attendance. In this case, however, a relationship may have been obscured by combining attendance with three other indicators of organizational religiosity. In a more convincing study, Presser and Stinson (1998) examined the education-attendance association using nationally representative data collected from five GSS, Gallup and EPA surveys conducted between 1992 and 1994. Direct attendance items were used for each of the GSS and Gallup surveys, with the time-use item being used for the EPA survey. Surprisingly, Presser and Stinson found no relationship between education and attendance, regardless of which item was used.

Additionally, a few studies have even reported negative associations between education and attendance. For example, Musick et al. (2004) used data provided by over 3,600 adults aged 25 and over who participated in the ACLS, and found that those with less education were more likely to attend religious services. Other studies have also reported negative associations between education and attendance, but these findings seem to be limited to certain subsets of the American population. For example, Cadge and Ecklund (2006) collected data from a nationally representative sample of 784 immigrants to the U.S., and found that high school graduates attended less than non-graduates. Iannaccone (1998) also observed that while conservative Christian denominations (and religious sects) tend to enjoy relatively high rates of attendance, they also tend to attract less educated members into their congregations. It is important to point out, however, that this observation was made at the group-level, and not at the individual-level.

Thus, it is possible that the individual-level relationship between education and attendance *within* conservative denominations is still positive, but further research is needed to determine if this is the case.

In their initial analysis, Stark et al. (1997) also found a negative relationship between education and attendance. Specifically, they combined 1972 through 1990 GSS data on approximately 30,000 respondents, and compared the religiosity of professors and scientists (n ~ 300) with that of graduate students (i.e., those with two or more years of graduate training; n ~ 1,300) and the general public (i.e., everyone not included in the previous two categories; n ~ 28,400). Table 2 provides descriptive statistics on seven religious indicators, including attendance, for each of the three groups. The general population ranked highest on all seven religious indicators, with the graduate student sample falling somewhere in between the general population and the professor/scientist group. With regard to attendance, 38% of professors and scientists, 43% of graduate students and 45% of the general public reported attending religious services at least once a month. When controls for gender, ethnicity, marital status, region and religious affiliation during childhood were included, however, professors and scientists were just as likely, and graduate students were more likely than the general public to report monthly attendance, suggesting that less religious persons tend to enroll in graduate school and pursue careers as professors and scientists at a greater rate than more religious persons.

Interestingly, the findings reported by Stark et al. (1997) may be limited to certain fields of study. Citing results from a survey of over 60,000 college faculty initiated by the 1969 Carnegie Commission, Stark et al. indicated that faculty in the "hard" sciences were more likely to be religiously affiliated, attend regularly, describe themselves as "moderately" or "deeply" religious, and were less likely to be opposed to religion than faculty in the "soft" sciences, even after controlling for gender, ethnicity, age and religious upbringing. In particular, faculty in psychology and anthropology were singled out as being particularly irreligious, with these groups being nearly twice as likely as faculty in the hard sciences to never attend religious services or to be religiously *un*affiliated. Unfortunately, the Carnegie Commission data do not provide a ready comparison with the general public; hence, it is unknown how faculty in the hard sciences compare with the general public on measures of religiosity. Furthermore, the Carnegie data were collected during an atypical period in American history (i.e., during the late 1960s), and may not generalize to other time periods. Accordingly, additional work will be needed to verify and expound upon the findings. Until that happens, however, Stark et al. have identified a potentially important caveat to their initial GSS findings, suggesting that perhaps only select fields of academics and scientists are less religious than the public.

Income. In an interesting assessment of the relationship between religious service attendance and income, Iannaccone and Everton (2002, p. 198) state the following:

> The conventional wisdom, stretching back to ancient times, strongly suggests that rich people are less religious than the poor ('It is easier for a camel to go through the eye of a needle than for a rich man to enter the kingdom of God' [Matthew 19:23, NIV]). At the level of personal belief this may be true, but at the level of external involvement, nearly all survey data demonstrate a weak, but generally *positive*, relationship between income and religious observance.

A quick scan of the literature confirms this conclusion. For example, Mueller and Johnson (1975) used data provided by approximately 1,900 adults who participated in a nationally representative survey conducted by the Institute for Survey Research at Temple University in 1970, and found that income and attendance shared a small, but positive relationship (r = 0.04). Interestingly, Mueller and Johnson also found an interaction with gender, such that the income-attendance association was stronger for men (r = 0.10), and essentially zero (r = -0.01) for women. In another study, Cotter and Song (2009) pooled data from over 5,000 married persons between the ages of 21 and 65 who participated in the 2003 to 2005 ATUS, and found that families with incomes between $40,000 and $75,000 spent significantly more time on religious activities than those with lower incomes. In contrast to the previous study, however, Cotter and Song found that this relationship held only for females. Hays et al. (1998) studied the relationship between income and attendance using data from a random sample of nearly 3,000 elderly persons living in North Carolina who were first interviewed in 1986, and then again three years later. When examining the longitudinal aspect of these data, Hays et al. found that men with relatively high incomes at baseline were more likely to increase their attendance over time than men with lower incomes. The same did not hold true for women, however, as baseline income was unrelated to attendance three years later.

Taking another angle, DiIulio (2002) reviewed studies that looked at the influence of attendance on income among disadvantaged youth. He concluded that attendance at religious services tends to help African American youth from disadvantaged neighborhoods escape poverty and avoid delinquency, crime and other social ills. DiIulio indicated that there were over two dozen studies on delinquency and crime that found as much. Fan (2008, p. 304) came to a similar conclusion after another review of the literature, stating that "there is substantial evidence showing that religion has a significant positive impact on children's educational attainment and future earnings." Thus, it appears that income and attendance may share reciprocal effects.

As was the case for education and attendance, however, a few studies have found null and even negative relationships between income and attendance. For example, Iannaccone (1998, 2003) used data from the GSS, and found no relationship between income and attendance. Additionally, Taylor (1986, 1988b) used nationally representative data from the NSBA, and found that income was unrelated to attendance in multivariate models that included gender, marital status, urbanicity, region of residence, age and education. When Taylor (1988a) dichotomized attendance and compared attenders to non-attenders (i.e., those who had not attended a religious service other than a funeral or wedding since the age of 18), however, he found that the former had higher incomes than the latter. Finally, Williams and Lawler (2001) used data provided by over 1,200 married adults who participated in a nationally representative telephone survey, and actually found a negative relationship between income and a measure of religious behavior, which included attendance. After including controls for other religious indicators, however, this inverse relationship disappeared. Despite the mixed findings reported here, it is important to remember Iannaccone and Everton's (2002, p. 198) contention that income and attendance share "a weak, but generally positive" relationship. If this contention is true and the relationship between income and attendance is positive, but close to zero, then it follows that a relatively large proportion of the income-attendance associations will be non-significant simply due to sampling error.

Employment Status.[57] Unlike the findings for education and income, the few studies that have investigated the relationship between employment status and religious service attendance have generally found a negative relationship. For example, Cotter and Song (2009) pooled data from over 5,000 married persons who participated in the ATUS between 2003 and 2005, and found that attendance rates on a given Sunday were lowest in situations where both partners were employed full-time. In another study, Cadge and Ecklund (2006) analyzed data collected from a representative sample of 784 immigrants to the U.S., and found that unemployed respondents were more likely to attend than employed respondents, regardless of the number of hours worked per week. Ulbrich and Wallace (1984) also found a negative relationship between employment

[57] Recall that employment status is used here as a proxy for occupational status given that the majority of the attendance literature focuses on the former variable.

status and attendance among women, with employed and unemployed women attending an average of 34 and 43 times per year, respectively. In a cross-national study, Campbell and Curtis (1994) combined representative survey data from 22 countries, including the U.S., and found employment status to be negatively related to both religious beliefs and attendance. And, in the lone longitudinal study found for this review, Ferraro and Kelley-Moore (2001) used nationally representative data provided by nearly 3,000 adults (age 25 and over) who participated in two waves of the ACLS (i.e., in 1986 and again two-and-a-half years later), and found that employment status was negatively related to attendance in both cross-sectional and longitudinal analyses, even after controlling for demographic and religious characteristics. The longitudinal association, however, held only for those who were affiliated with a religion.

Despite these findings, there is some evidence that the relationship between employment status and attendance may be qualified by gender. For example, de Vaus (1984) combined GSS data from 1972 through 1980, and found a negative association between employment status and attendance among women, with the unemployed (54%) and part-time employed (56%) being more likely (but not significantly so) to be regular attenders than those who were employed full-time (51%). The association for men, however, was positive with the unemployed (30%) and part-time employed (33%) attending significantly less than those who were employed full-time (41%). Additionally, de Vaus examined attendance levels for men and women as a function of occupational status, and found that women's' attendance was unaffected by their occupational status, while men's attendance tended to be more frequent if they were in high status jobs than if they were in low or medium status jobs. In another study, Hertel (1995) combined data from the 1972 through 1990 GSS, and found that married women who were unemployed were more likely to be religiously involved than those who were employed full-time. For single women and all men, however, full-time employment was positively associated with religious involvement. Thus, these data suggest that the relationship between employment status and attendance may be dependent upon gender.

There are also a couple of studies suggesting that employment status may be unrelated to attendance. For example, Becker and Hofmeister (2001) used longitudinal data from 1,000 adults living in four communities in upstate New York who participated in the Religion and Family Project, and found that there was no relationship between employment status and attendance among men or women. Yet, there was an interaction among full-time employed men, such that those with highly individualistic views of religious authority (i.e., the view that one does not need organized religion to understand and live-out one's religion) were more likely to attend religious services than those with more communal views. In addition, studies that have indirectly tested the relationship between employment status and attendance have failed to find an association. Specifically, Carroll and Roozen (1975), de Vaus (1984) and Lazerwitz (1961) hypothesized that if participation in the labor force is related to lower levels of attendance, and males both attend less and are more likely to be employed than women, then the gender gap in attendance should diminish when employment status is controlled. The results from each of the three studies, however, failed to support this hypothesis, indicating that either employment status is unrelated to attendance or that its role in explaining the gender gap is trivial. Given that a majority of the studies reviewed here have found a relationship between employment status and attendance (either negative or qualified by gender), however, the latter interpretation seems more appropriate.

Theoretical Explanations for the SES-Attendance Trends

There are three notable trends in the literature on socioeconomic status and religious service attendance that deserve further attention. First, a majority of the findings indicate that positive relationships exist between attendance and composite SES, education and income. Second, and despite the first observation, a sizeable portion of the SES-attendance findings suggest that no relationship exists. Finally, one indicator of SES, employment status, seems to be negatively related to attendance. Below, I summarize the findings and review a number of explanations for these three very different trends.

Positive SES-Attendance Relationships. As noted above, positive relationships have been consistently found between attendance and composite SES, education and income. Importantly, these relationships have held up across both convenience and nationally representative samples, as well as across most of the major religious affiliations represented in the United States. There is also evidence that these variables share reciprocal effects over time, with attendance predicting (and being predicted by) educational achievements and financial earnings. And, longitudinal associations between these variables have been found to be stronger for those coming from economically disadvantaged neighborhoods.

With the exception of the latter finding, these results are surprising given that one of the most popular theories in the history of the scientific study of religion states that it is the poor, the oppressed and the marginalized who are most likely to seek the comforts and compensatorial promises of religion (see Dittes, 1971). Deprivation theory, as it has been called, predicts that those who are deprived of status, opportunity, material wealth, relationships and the other "cares of this world" will be more likely to participate in religion where they are likely to: (a) hear messages that are both welcoming and that devalue the pursuit of worldly possessions; (b) be surrounded by like-minded others; and, (c) gain opportunities for both leadership and fellowship. Yet, while lower SES persons may, in fact, benefit more from their religious participation (e.g., see DiIulio, 2002; Koenig et al., 2001; Pargament, 1997), the literature reviewed here indicates that they are less likely to be involved with religion in the first place. As Stark (1972, p. 500) put it: "In order for economic deprivation to result in . . . religious commitment, it is necessary first that a religious perspective is a *plausible option* . . . But the fact remains that the economically deprived are those for whom religious options are least likely to be relevant." Thus, despite deprivation theory's long history in the social scientific study of religion, new and alternative explanations are needed to explain the relationship between SES and attendance in the general population.

In light of deprivation theory's failure to explain the general population findings, a few researchers have suggested that perhaps religious participation merely reflects a tendency of high SES persons to become more involved with social and civic organizations of all types (e.g., see Demerath, 1965; Goode, 1966; Stark, 1964; Lenski, 1953, 1961, 1963). If this is true, then the relationship between SES and religious participation should disappear when participation in non-religious organizations is controlled. The evidence, however, provides only partial support for this argument. Goode (1966), for example, found that while religious participation *is* positively related to non-religious participation, the relationship between SES and attendance is only partially explained by this latter variable.[58] In addition, Mueller and Johnson (1975) used nationally representative data from over 2,400 adults, and found that the relationship between

[58] Goode (1966, p. 108) actually states that the relationship is "greatly attenuated," but an examination of his data indicates that this language is too strong. For example, in an uncontrolled analysis of data from the "Appalachian sample," Goode reported that 77% of White Collar workers attended at least weekly, while 49% of Blue Collar workers did so, for a difference of 28%. When non-Church activity was controlled, however, the respective differences between these groups were 25% for the low non-Church activity group and 20% for the high non-Church activity group. Thus, even after controlling for non-religious participation, there was still a difference in weekly attendance of at least 20%. Similar findings were reported for the relationship between education and attendance.

SES and attendance persisted after controlling for participation in non-religious organizations. And, Hoge and Roozen (1979) reported in their review that several researchers (e.g., Blaikie, 1972; Estus & Overington, 1970; Lazerwitz, 1964) found the relationship between SES and religious participation to weaken, but by less than half after controlling for participation in non-religious organizations. Thus, while high SES persons may, in fact, be more likely to participate in organizations of all kinds, there are clearly additional factors responsible for the positive associations between religious involvement and SES.

One of the additional factors may stem from the typical composition of religious groups in America. Specifically, religious groups tend to be composed of a single social class or ethnic group (e.g., see Winter, 1962; cited in Hoge & Roozen, 1979). And, according to a number of researchers, those religious groups that consist primarily of middle- and upper-class congregants tend to approximate status groups (e.g., see Cotter & Song, 2009; Demerath, 1961; Hertel, 1988; both cited in Cotter & Song, 2009; Hoge & Roozen, 1979). Status groups—originally described by Max Weber—have been defined as "associations in which participation by members bestows status, identity and honor" (see Hoge & Roozen, 1979, p. 55). Given that religious participation is generally valued in America, it makes sense that attending one of these middle- or upper-class religious groups could be a means of conferring and confirming status. It should also be pointed out that business professionals who rely on social networking may find additional incentive to participate in religion where they are likely to find numerous opportunities to meet others and build relationships. Thus, at least part of the positive relationship between SES and attendance might be explained by a need for status, identity, honor and social networking.

Despite the fact that status group theory makes intuitive sense, it remains difficult to test. Yet, a couple of studies have attempted to do so. Specifically, Hoge and Polk (1976; cited in

Hoge & Roozen, 1979, p. 56) used data obtained from an interdenominational sample of Protestants, and found that attendance was positively correlated (r = .21) with responses to the following statement: "Church membership has helped me to meet the right kind of people." In another study of Presbyterian and Methodist Church members, however, Hoge and Carroll (1978) failed to find the same association. Thus, the available evidence regarding status group theory is both limited and inconclusive, and further research will be needed to determine if it is able to explain a portion of the relationship between SES and attendance.

Another theory that has been offered states that those who participate in religious services are privy to messages that both encourage pro-social behavior and discourage anti-social behavior; and, in turn, these behaviors are thought to lead to greater educational opportunities and higher earning potentials. For example, religious participation has been tied to the development of values, character, honesty, discipline, self-control, responsibility, commitment, persistence, social skills, a sense of structure in one's life and future-oriented thinking (DiIulio, 2002; Fan, 2008; Koenig et al., 2001; Muller & Ellison, 2001). In addition, religious participation has been linked with lower levels of crime and delinquency, drug use, alcohol abuse, premarital sex and pregnancy (DiIulio, 2002; Koenig et al., 2001; Lipford & Tollison, 2003). Thus, religious participation seems to lead youth into behaviors that allow them to maximize their educational and financial earning potential, while simultaneously discouraging them from engaging in behaviors that could disrupt or derail these pursuits. Importantly, reviews by DiIulio (2002) and Fan (2008) support this contention. Moreover, Fan (2008) suggests that middle- and upper-class parents tend to be more concerned with their children's human capital formation—including the development of positive peer relationships and a broad educational background—and consequently are more likely to get their children involved in religion precisely because of the pro-social benefits noted above. Thus, SES and religious participation may form a cycle, where high SES parents get their children involved in religion so that they can maximize their education and earning potential, which encourages them to do the same when they have children of their own. While not all of the links in this theoretical chain have been adequately tested, it does appear that youthful religious participation is linked to greater educational achievements and higher incomes later in life.

Null SES-Attendance Relationships. While a majority of the SES-attendance findings are positive, a sizeable portion of the findings (21% for education, 25% for income and 18% for composite SES) indicate that no relationship exists. Accordingly, explanations need to be considered for these findings. First, it is possible that the true relationship between attendance and SES is zero, and the null relationships reported above merely reflect this reality. The most convincing argument for this possibility comes from the Presser and Stinson (1998) study. Recall that they examined the education-attendance association using five nationally representative datasets and three attendance items (i.e., the direct GSS and Gallup items, and the EPA time-use item). In each case, Presser and Stinson found that education was unrelated to attendance. While the findings from this study are convincing, a number of other studies have used nationally representative datasets and found positive associations (e.g., see Ferraro & Kelley-Moore, 2001; Loury, 2004; Muller & Ellison, 2001; Regnerus & Elder, 2003; Uecker et al., 2007). Thus, if the true relationship is zero, one would still have to explain why a majority of the findings suggest a significant and positive relationship. Accordingly, it seems very unlikely that there is no relationship between attendance and SES.

A more likely explanation is that the mix of positive and null SES-attendance findings are a result of two factors. The first is that the true relationship between attendance and SES is small, but positive as Iannaccone and Everton (2002) suggested. Recall, for example, that Mueller and Johnson (1975) used data from a nationally representative sample of approximately 1,900 adults, and found a significant positive correlation between income and attendance that was equal to just 0.04. If the true relationship between SES and attendance is this small (or close to it), then random sampling theory would explain at least a portion of the null findings given that the true relationship is so close to zero.

The second factor that might be partly responsible for the mix of positive and null SES-attendance findings is that the true relationship may not hold for all population sub-groups. Recall that notable interactions were found for the education- and income-attendance associations by religious homogamy (education and attendance were unrelated for religiously heterogamous couples), immigration status (education and attendance were negatively related among immigrants), religious affiliation (conservative religious denominations and sects tend to be lower in SES than mainline religious affiliations; and, there may be no relationship between SES and attendance among the largest African American religious affiliation [Baptist] in America), scientific discipline (those in the social sciences attend less than both the general public and those in other scientific fields) and gender (income and attendance were unrelated among women, but positively related among men). Thus, it may be that while the relationship between SES and attendance is small and positive in the general population, the introduction of relatively large concentrations of these interaction groups is attenuating the overall relationship enough that a sizeable portion of the findings are turning out non-significant. Before accepting this conclusion, however, additional research is needed to: (a) establish the strength of the SES-attendance relationship; (b) verify whether the interactions noted above do, in fact, exist, and (c) determine the extent to which these interactions are attenuating the overall relationship.

Negative SES-Attendance Relationships. It is perplexing that education and income would share a positive relationship with religious service attendance, while employment status shares a negative relationship. Yet, this clearly seems to be the case given that the negative association has held up in cross-sectional and longitudinal studies that use data obtained from convenience and nationally representative samples. Accordingly, a number of explanations have been offered for the negative relationship between attendance and employment status. For example, several researchers have suggested that unemployed persons attend more frequently because they have more time and energy for religious pursuits than employed persons (e.g., see Glock, Ringer & Babbie, 1967; cited in de Vaus, 1984). Other researchers have cited examples of deprivation theory, suggesting that unemployed persons are more likely to attend than employed persons in order to fulfill social interaction needs and leadership roles that might otherwise be satisfied in the workplace (e.g., see Glock, Ringer & Babbie, 1967; Moberg, 1962; Yinger, 1970; all cited in de Vaus, 1984). Finally, some researchers have suggested that workforce climates may encourage competitive behavior and secular views that contrast with religious perspectives (e.g., see Lenski, 1953; cited in de Vaus, 1984). While each of these explanations seem conceptually plausible, they have yet to be empirically tested, and therefore, add little to what is actually known about the relationship between employment status and attendance.

One explanation for which there is evidence, however, comes directly from the empirical literature. Specifically, seven studies that were reviewed in the sections on income and employment status identified a gender interaction. In six of the studies, the interactions indicate that, for women, a null or negative relationship exists between attendance and both income and employment status, while, for men, a positive relationship exists between these variables (see Becker & Hofmeister, 2001; de Vaus, 1984; Hays et al., 1998; Hertel, 1995; Mueller & Johnson, 1975; Ulbrich & Wallace, 1984).[59] In explaining this interaction, both de Vaus (1984) and Becker and Hofmeister (2001) suggested that attendance may take on different meanings for men and women. Specifically, they suggested that, for men (especially those with families), religious service attendance is part of a larger orienting system that is consistent with the provider role. Under this conceptualization, attendance, along with education and employment are mutually supportive activities that can provide a sense of "maturity and social establishment" (Becker & Hofmeister, 2001, p. 719). For women, on the other hand, attendance is likely part of the nurturing or caretaking role. Thus, attendance and employment are likely antagonistic activities given that they cross over traditional gender roles for women. Regardless of whether this explanation is accurate (it has not been tested), the fact that the interaction has been consistently found indicates that it is the most likely culprit behind the negative employment status-attendance associations.

In sum, the literature seems to suggest that attendance is positively related to education, income and composite measures of SES, while being negatively related to employment status. While a number of explanations have been offered for these associations, very few have been adequately tested. Therefore, much can still be learned about the relationships between the SES indicators and attendance. For now, however, this investigation focuses on determining the influence that SES has on estimated rates of attendance in America by testing the following hypotheses:

Hypothesis 10. Studies that utilize samples of relatively well educated participants will yield relatively high attendance estimates.

---

[59] The seventh study (Cotter & Song, 2009) actually found the opposite pattern for income and attendance, with the relationship being null for men and positive for women.

Hypothesis 11. Studies that utilize samples of relatively wealthy participants will yield relatively high attendance estimates.

Hypothesis 12. Studies that utilize samples of primarily employed persons will yield relatively low attendance rates.

Hypothesis 12a. Studies that utilize samples of primarily employed *males* will yield relatively high attendance rates.

Hypothesis 12b. Studies that utilize samples of primarily employed *females* will yield relatively low attendance rates.

General Summary

It was long thought that just over 40% of the American population attended religious services on a regular basis. Yet, as we have seen in this review, recent research has suggested that these estimates are inflated due to problems with ambiguous item wording and socially desirable responding. Efforts to minimize these problems have yielded a great deal of variability in the resultant attendance estimates, with some being half as much as the previous gold standard estimates. A large portion of this variability stems from the introduction of a distinctly different type of question. Specifically, researchers began asking respondents if they attended a religious service in a given week, whereas previous questions asked respondents about their attendance habits over an extended (and often undefined) time period. Naturally, the estimates from the former types of questions were much lower than those from the latter. There is still a good deal of variability within each of these two types of questions, however. Thus, there is no current gold standard estimate that can tell us how many Americans attend on a regular basis or on any given Sunday.

Given that religious service attendance seems to be a robust predictor of a large number of important biopsychosocial outcomes, the attendance construct deserves further attention so that we can better understand the “active ingredients” of attendance that are responsible for these important associations. Before delving more deeply into the attendance construct, however, it is important to first establish basic facts about the construct, namely to determine how many Americans attend religious services both on a regular and weekly basis. Importantly, any attempt to address the issue of attendance frequency must account for factors that are known to alter or influence attendance estimates. As previously mentioned, item wording and mode of data collection are two such factors that need to be controlled. In addition, a number of sociodemographic variables have also been found to systematically influence attendance rates. Specifically, the literature reviewed here suggests that samples with relatively high proportions of females, African Americans, middle- and late-age adults, married adults, parents of school-age children, educated and wealthy adults and unemployed females tend to yield relatively high rates of attendance. Thus, any attempt to estimate either regular or weekly attendance must account for these psychometric and sociodemographic variables.

Because no single study can definitively address the issue of attendance frequency while simultaneously controlling for all of the psychometric and sociodemographic variables mentioned above, meta-analytic techniques were employed to synthesize data from the available literature on attendance frequency. Specifically, meta-analytic techniques allow for: (a) the systematic identification of a large number of attendance estimates that have been reported in the published and unpublished literature; (b) the estimation of mean attendance rates among Americans across all available studies; and, (c) the control of a number of variables—including item wording, mode of data collection and the sociodemographic variables covered in this review—that have been shown to systematically influence attendance rates. By doing this, I hope to help establish new gold standard estimates for religious service attendance in America. These new estimates can then be used to gauge future efforts that attempt to improve upon existing item wording or data collection modes, as well as to gauge the representativeness of samples when exploring associations between attendance and the numerous biopsychosocial outcomes that have been linked to religious service attendance.

CHAPTER III

METHOD

This chapter presents a discussion of the history, purpose and procedures associated with research synthesis and meta-analysis, beginning with a discussion of the social and scientific milieu from which meta-analysis emerged. This is followed by an overview of the developmental history of the meta-analytic method and a discussion on how the method is employed in this study. The latter discussion includes the following elements: (a) Operationally defining *regular* religious service attendance; (b) literature search procedures; (c) study inclusion and exclusion criteria; (d) data extraction and coding procedures; (e) assessing and improving coding quality; (f) preparation of the meta-analytic database; (g) cumulation of study findings; (h) assessment of data heterogeneity and precision; (i) visually presenting the collection of study outcomes via forest plots; and, (j) meta-regression analysis.

Problems with Traditional Methods of Reviewing the Literature

The need for systematic and empirical methods of reviewing a literature arose in the 1960s and 1970s after researchers and policymakers became increasingly frustrated by the inconsistent (and often contradictory) findings that were reported within a number of research literatures (see Hunter & Schmidt, 2004). Senator Walter Mondale voiced his frustration about this problem in his address to the American Psychological Association in 1970:

> I had hoped to find research to support or to conclusively oppose my belief that quality integrated education is the most promising approach. But I have found very little conclusive evidence. For every study, statistical or theoretical, that contains a proposed solution or recommendation, there is always another, equally well documented, challenging the assumptions or conclusions of the first. No one seems to agree with anyone else's approach. But more distressing I must confess, I stand with my colleagues confused and often disheartened (cited in Hunter & Schmidt, 2004, p. 18 – 19).

Narrative Reviews

One of the contributing factors to this state of affairs was the predominant use of narrative reviews (Hunter & Schmidt, 2004). In a narrative review, synthesists attempt to summarize the findings from a particular literature, and develop theories based on what the preponderance of evidence seems to be saying about the phenomenon under study. A problem with this approach (and the one mentioned by Senator Mondale) is that bodies of literature often fail to provide clear-cut answers. Studies differ with respect to the research designs, samples, measures and analyses they employ. Each of these methodological factors can influence a study's findings, thereby creating "artifactual" variability in the literature. As Hunter and Schmidt (2004) state, the task of sifting through this variability and coming to a correct conclusion about the phenomenon under study may be doable if there are only a few studies, but if the literature is large, the task can quickly become too complex.

In the face of such complexity, Hunter and Schmidt note that synthesists typically undertake one or more of the following activities: (1) summarize the findings in a bibliographic-style report; (2) summarize the findings from a small subset of studies that meet certain criteria for "methodological rigor;" and/or, (3) attempt to review and integrate all of the findings. Each approach presents problems. Synthesists who use the bibliographic approach may only give marginal consideration to deciphering the underlying message about the phenomenon under study. Instead, the primary focus is on summarizing the findings from each study separately and independently, leaving the task of deciphering and clarifying the underlying message to the reader. Synthesists who restrict their samples to those of a certain methodological rigor not only discard valuable study data (i.e., that which is discarded because it is lack of "rigor"), but also rely on methodological criteria that are subjectively chosen and unreliably implemented (Lipsey

& Wilson, 2001). Last, synthesists attempting to review and integrate an entire body of literature (e.g., see Koenig et al., 2001) can easily arrive at erroneous conclusions given the complexity of the task.

Vote-Counting

In response to the problems cited above, some synthesists have adopted a simple vote-counting procedure to make sense of a literature's findings. Vote-counting involves adding up the number of significant findings in a literature, comparing them to the number of non-significant (or opposite direction) findings and declaring the most prevalent finding the "winner." Theories are then proffered to help explain the majority finding. The narrative review presented here in Chapter II, for example, uses a vote-count. For each socio-demographic variable that was reviewed, the most prevalent association found in the literature (e.g., women attend more than men) was used to form a hypothesis about the underlying relationship between attendance and the respective demographic. Non-significant and alternative findings were also presented, but the respective conclusions were derived from the predominant finding in the literature. In this way, vote-counting allows synthesists to retain all of the findings from a literature while providing a simple means of reducing their complexity. Despite the inclusive and simplistic nature of vote-counting, however, a number of problems have been identified with this approach.

Perhaps the most obvious problem with vote-counting is that it gives equal weight to every study in a review regardless of a number of factors that are capable of influencing a synthesist's conclusions (Hunter & Schmidt, 2004; Shadish, Cook & Campbell, 2002). For example, a study that uses a small convenience sample, an unreliable measure and a less than ideal research design will contribute the same amount of information in a vote-count as a study that uses a large probability sample, a reliable measure and an appropriate research design. While both of these studies contain relevant information, the latter study should be given greater weight when arriving at a conclusion given that its methodological characteristics make it more likely than the former to produce an accurate estimate of the population parameter. The vote-count procedure, however, fails to make this adjustment, thereby giving equal weight to both studies and making it more difficult for the synthesist to arrive at the correct conclusion.

A second problem with vote-counting is that narrative reviews, while broad and inclusive in their representation of the published literature, are typically not exhaustive in their search procedures, nor are they necessarily representative of all available data on a topic. For example, narrative reviews primarily focus on studies that have been published in peer-reviewed journals and books, while often neglecting (or underrepresenting) other important sources of information, such as internal organizational research reports, unpublished theses and dissertations, unpublished conference papers and other unpublished manuscripts that tend to be more difficult to obtain. In addition, it is commonly assumed (although there is mixed evidence; see Hunter & Schmidt, 2004; Bax et al., 2008; Buschman & Wang, 2009; Smith, McCullough & Poll, 2003) that published sources are more likely to contain statistically significant findings than unpublished sources.[60] If this assumption holds true, then any conclusion derived from vote-counting will be biased against the null hypothesis. Together, this set of conditions is referred to

[60] Hunter and Schmidt (2004) note that this is particularly likely for studies carried out prior to the emergence of PC-based statistical software, when significance tests required much more time and effort to compute. With present computing capabilities, however, significance tests are much easier to run. Consequently, modern-day researchers are more likely to test multiple hypotheses, and to generate at least one significant finding. Thus, the difference between modern published and unpublished studies is determined less by the significance of one's findings than of other factors (e.g., a lack of time or resources) that have little or no bearing on a study's findings. In addition, Hunter and Schmidt note that many reviews are conducted on variables that are of secondary interest in primary studies. In such situations, the distribution of significant and non-significant findings in a literature is not tied to the decision to publish, thereby reducing the possibility of publication bias.

as *publication bias*.[61] Thus, the lack of exhaustive search procedures opens the door for bias to affect conclusions drawn from vote-counting narrative reviews.

A third problem, and one that is often overlooked by synthesists, is that vote-counting narrative reviews fail to take into account the effects of study artifacts. Study artifacts are defined by Hunter and Schmidt (2004, p. 33) as "study imperfections" that have the ability to "impact . . . the results of a [research synthesis]." Typically, the most problematic and influential of these artifacts is sampling error. Sampling error refers to the difference between a statistic and a corresponding population parameter that results from the collection of data from a sample rather than the entire population (Vogt, 1999). All studies failing to gather data from an entire population (i.e., virtually all studies) are assumed to suffer from sampling error because, in all likelihood, the characteristics of a sample will not perfectly match those of a population even when large probability samples are drawn. Consequently, studies that are otherwise identical will yield different results simply because they are tapping into slightly different segments of the population. As studies accumulate on a topic, then, so does the variability among the findings. In fact, the central limit theorem predicts that research outcomes will approximate a normal distribution when a sufficient number (i.e., typically 30 or more) have been collected (see Hinkle, Wiersma & Jurs, 1998, p. 172). Thus, synthesists are faced with the reality that research findings are actually *expected* to vary—and on occasion, vary considerably—across studies. This raises the possibility that a collection of findings that might typically be interpreted as inconsistent or contradictory may actually be nothing more than the visible effects of sampling error. Hunter and Schmidt (2004) even state that it is not uncommon for sampling error to account for most, if not all, of the variability in a particular research literature!

[61] Hunter and Schmidt (2004) correctly note that this is just one instance of a larger problem of locating and finding all available studies on a topic; hence, they prefer the term *availability bias*.

A second study artifact that can pose problems for synthesists is measurement error. Measurement error is broadly defined as any "inaccuracy resulting from flaws in a measuring instrument" (Vogt, 1999, p. 173). These inaccuracies can be both random and systematic (Hunter & Schmidt, 2004). Random measurement error can be caused by a variety of factors (e.g., misreading questions, memory lapses, temporary changes in attitude or mood, or mismarking response options) that vary from one administration to another such that the net effect over an infinite number of measurements is expected to be zero. For any single measurement, however, random error has the effect of adding variability (or "noise") to a collection of findings (Trochim, 2001). This characteristic of random measurement error has two important consequences for the synthesis of attendance data.

First, attendance rates estimated across studies, or over time, may look like they are varying, when, in fact, the observed variation is nothing more than the result of random measurement error. When comparing attendance rates from two items, this could lead a researcher to falsely conclude that real differences exist between the attendance rates when, in fact, they do not (i.e., a false positive or Type-I error). Conversely, small, but systematic differences in attendance rates may be obscured by the effects of random measurement error, thereby leading researchers to falsely conclude that no differences exist between items when, in fact, they do (i.e., a false negative or Type-II error).

Second, the added variability associated with random measurement error has the effect of attenuating bivariate and multivariate relationships (Hunter & Schmidt, 2004). In a primary study, this poses problems for researchers attempting to investigate the relationship between attendance and some other individual-level characteristic (e.g., attendance and income). In meta-analyses, this poses problems for synthesists attempting to investigate the relationship between study-level attendance (i.e., the average attendance rate for a study sample) and other study-level characteristics, such as study-level income (or the average income of the study sample). Thus, random measurement error has the effect of adding variability to a set of data, which may lead to incorrect conclusions about attendance rates, and their relationship to other study characteristics.

Systematic measurement error, on the other hand, has the effect of biasing instrument scores in one direction or the other (i.e., either up or down), thereby introducing inaccurate point estimates into the literature. For example, attendance items that allow socially desirable responding are expected to produce artificially high attendance rates. Similarly, items that use ambiguous wording are also expected to produce rates that are artificially high given that respondents are likely to include religious events other than worship services in their answers (e.g., watching or listening to religious programming on TV or the radio). Conversely, estimates gleaned from count data are likely to systematically underestimate the attendance rate given the difficulties of both accounting for all attendees at a given service and accurately assessing the active membership rolls (a full membership list would be expected to overestimate the number of congregants that could be expected to be present on any given Sunday; Iannaccone & Everton, 2002). Importantly, however, one of the primary goals of this study is to estimate precisely how much systematic error exists for each of the attendance measures. Because this source of systematic error is measured and, in effect, partialled out of the analysis, its consequences for this study should be minimal.

An additional source of systematic measurement error, and one that was alluded to in Chapter II, stems from the disparate use of response scales. The most obvious example of this is when we compare the original Gallup item, which asks respondents if they have attended "Church or Synagogue" in the last seven days and provides just two response options ("Yes" and

"No") with the original GSS item, which does not specify a time frame and provides *eight* response options ("Never," "Once a year or less," "Several times a year," "Once a month," "2 to 3 times a month," "Nearly every week," "Every week" and "Several times a week," with the latter four being counted as *regular* attendance). Given the different time-frames and response options, it is obvious that results from the two items are not directly comparable. This congruence problem exists even among similarly worded items that provide multiple response options. For example, the wording of the Monitoring the Future (MTF) survey item is identical to the original GSS item, but the response options are curtailed to five: "Never," "Rarely," "About once or twice a month" and "About once a week or more," with the last two constituting *regular* attendance. Here, the lowest GSS and MTF response options that are considered as "regular attendance" are "2 to 3 times a month" and "About once or twice a month," respectively. Given that the latter response option is more inclusive (i.e., the MTF item includes those who attend once a month, whereas the GSS item does not), the MTF item will, *ceteris paribus*, be expected to yield higher rates of attendance. This measurement problem is replete in the attendance literature, and narrative synthesists who fail to account for these response option differences will arrive at findings that are biased by this form of measurement error. In this study, a procedure referred to as "harmonizing" (e.g., see Rossi and Scappini, 2014) is used to adjust the response scales so that they are maximally similar.

Besides sampling and measurement error, narrative-vote counting reviews can also be adversely affected by an artifact known as, "restriction of range" (Hunter & Schmidt, 2004). Restriction of range results from samples that are drawn from relatively small segments of the population, and have the effect of biasing attendance estimates up or down depending on which tail of the distribution the sample is drawn from. For example, if a sample of elderly women is drawn, then one could expect an attendance estimate that is biased upward given that older adults and females seem to attend more frequently than their respective counterparts. Restriction of range can also have undue effects on bivariate and multivariate relationships. As alluded to earlier, parametric statistics are attenuated when variability is lost. Thus, studies that sample a homogenous group of participants are likely to produce attenuated bivariate and multivariate parameter estimates. Drawing participants from opposite ends of a population distribution, however, can have the opposite effect. For example, if a sample of young adults and elderly adults is drawn, the variance in attendance would be inflated given that these two groups represent the extreme ends of the attendance distribution. Any correlations estimated between attendance and age, then, would likely come out (falsely) as linear and highly positive given that the valleys and peaks in attendance across the lifespan are "skipped over." Thus, both range restriction and range exaggeration have the ability to introduce bias into the findings. In this study, it should be possible to estimate, and partial out the effects of range restriction by incorporating study sample characteristics (e.g., percent female, average age) into the meta-analytic models as potential moderators.

Summary

In sum, the early methods of reviewing the literature were fraught with problems. There were no systematic procedures for searching the literature and identifying both published and unpublished sources; thus, narrative reviews were generally composed of study samples that were unrepresentative and biased against the null hypothesis. In addition, synthesists tended to give equal weight to studies despite sample size and important methodological differences that were capable of influencing a study's findings. Important study artifacts such as sampling error, measurement error and restriction of range were largely ignored, thereby leading synthesists to inaccurately interpret error variance for real variance. It is not surprising, then, that researchers and policymakers were losing faith in the social sciences' ability to answer important social questions. Beginning in the mid-1970's, however, researchers began developing new methods for more accurately synthesizing a research literature.

The Meta-Analytic Method

Like other social scientists of the 1960s and 1970s, Gene Glass recognized the need for a new method of synthesizing a body of literature. To this end, Glass (1976) proposed an empirical approach that utilized tools already familiar to scientists across most disciplines. In rationalizing this approach, he stated that,

> Most of us were trained to analyze complex relationships among variables in the primary analysis of research data. But at the higher level, where variance, nonuniformity and uncertainty are no less evident, we too often substitute literary exposition for quantitative rigor. The proper integration of research requires the same statistical methods that are applied in primary data analysis (p. 6).

To distinguish this "higher level" analysis from both primary and secondary analysis, Glass (1976, p. 3) coined the term, *meta-analysis*, and defined it as "the analysis of analyses;" or, more specifically, a set of statistical procedures that allow for the "analysis of a large collection of analy[tical] results [obtained] from individual studies for the purpose of integrating the findings." Cooper, Hedges and Valentine (2009, p. 6) later adopted this definition, but went on to distinguish between meta-analysis, which refers specifically to the set of *statistical procedures* that can be used to integrate research findings, and *research synthesis*, which is a broader, more comprehensive term defined as "a set of literature review characteristics . . . [that] attempt to integrate empirical research for the purpose of creating generalizations." According to this distinction, research synthesis refers to the various methods and procedures that comprise the enterprise of collecting, reviewing, integrating and summarizing the findings from a literature, whereas meta-analysis refers specifically to the statistical procedures involved with integrating the findings. Given the popularity of the latter term, however, the remainder of this report uses the terms “meta-analysis” and “research synthesis” interchangeably.

Development of the Meta-Analytic Method

Like other new methodological and statistical developments, the meta-analytic method has undergone several revisions since the concept was formally introduced by Glass in 1976. Because these revisions are important for understanding the capabilities and advantages of meta-analysis, each of the major developmental stages will be discussed briefly.

Glassian Meta-Analysis

In moving from primary to meta-analysis, Glass (1976) argued that to make sense of the data, synthesists would need to change the unit of analysis from the level of a participant (i.e., an individual-level outcome) to that of a study (i.e., a study-level outcome, or as it is more commonly referred to, an effect size). This change in the unit of analysis made it possible for synthesists to discern the underlying message in a body of literature by applying the same statistical techniques they were used to using when making sense of participant-level data. In a demonstrative analysis, Smith and Glass (1977) gathered 375 studies that used both a treatment and a control group to examine the effectiveness of a variety of psychotherapeutic interventions on a number of psycho-social outcomes. Smith and Glass then extracted the necessary information from each study to calculate effect sizes representing the effectiveness of the psychotherapeutic interventions. The weighted mean of the 800 effect sizes that were extracted from the 375 studies was then calculated to provide an estimate of the average overall effect of psychotherapy. Likewise, the variance of the effect sizes was calculated so that confidence intervals could be constructed around the estimated mean. Together, these statistics allowed

Smith and Glass to quickly and accurately summarize the center and width of the distribution of psychotherapeutic intervention effects reported in the literature. Interpreting these data, Smith and Glass noted that the average psychotherapeutic intervention had the effect of moving a participant at the 50th percentile of the control group to the 75th percentile of the treatment group. The simplicity of this conclusion—made possible by the application of Glassian meta-analytic techniques—allows readers to quickly discern the efficacy of a typical psychotherapeutic intervention, and represented a giant step forward in the synthesis of literature.

While this new approach to research synthesis provided the type of answer that Senator Mondale had called for six years earlier, the method still had limitations. The most prominent limitation was that Glass (1976) ignored the effects of sampling error, believing that the observed variance among the effect sizes was due entirely to a combination of theoretical and methodological moderators (e.g., the duration and type of therapy, or the quality of the research design). Yet, Hunter and Schmidt (2004) found that sampling error is by far the most prominent source of variance in a body of literature, often leaving little (or no) variance to be explained by moderators. Accordingly, synthesists who mistake the effects of sampling error variance for real variance are likely to capitalize on chance when testing potential moderator variables. Thus, the approach advocated by Glass (1976) not only ignores a prominent source of error variance, but also increases the chances of obtaining false positives when testing moderators.

Another flaw in the Glassian meta-analytic approach is that Glass treated the effect size, and not the individual study, as the unit of analysis. Recall that Smith and Glass (1977) extracted 800 effect sizes from 375 studies, meaning that studies often contributed multiple effect sizes to the analysis. Because the effect sizes obtained from a single study can be expected to be more closely related to each other than they would to the effect sizes obtained from other studies,[62] the process of lumping them together in the analysis violates the assumption of statistical independence. According to Hunter and Schmidt (2004), violations of the statistical independence assumption typically have a conservative effect on meta-analytic results by overestimating the observed variance between the effect sizes. Thus, the Glassian approach would make it more difficult to find significant relationships between moderator variables and study outcomes, a problem that would be exacerbated in situations characterized by small sample sizes.[63] This was not a problem for Smith and Glass (1977) as they were primarily interested in *describing* the distribution of effect sizes (i.e., the mean effect of psychotherapy and the variance around the mean), and not necessarily in making inferences about the population. When significance tests are used and inferences are desired, however, this problem becomes relevant. Thus, the Glassian approach is useful for describing a literature, but not for making inferences.

A third problem with the Glassian approach was that Smith and Glass (1977) combined effect sizes representing a number of different therapeutic modalities and outcomes. This approach has been criticized as combining "apples and oranges," thereby making the interpretation of the average effect size difficult. As Hunter and Schmidt (2004) pointed out, however, Smith and Glass's (1977) initial intent was to *generally* summarize the effects of psychotherapeutic interventions reported in the literature. Furthermore, Smith and Glass followed-up their initial analysis by conducting a series of meta-analyses for each type of therapeutic modality; and, there is nothing in Glass's approach that precludes an analyst from doing the same for each dependent variable. Thus, Hunter and Schmidt (2004) concluded that

[62] Effect sizes obtained from a single study share a number of common influences—such as the same participants, intervention type, location, time period and even the researchers conducting the study—that are not equally shared by other studies.
[63] In Glassian meta-analysis, sample size is determined by the number of effect sizes available.

the “apples and oranges” criticism is not specific to the Glassian approach per se, but to the way in which the method was carried out.

The last major criticism of the Glassian approach is one that applies to the larger meta-analytic enterprise. Specifically, Smith and Glass (1977) were criticized for meta-analyzing the results from all available studies, regardless of their methodological quality. Some have argued for only including the results from the most rigorous studies to obtain a synthesis of the “best evidence” available (e.g., see Slavin, 1986). Proponents of this method state that by selecting only the most rigorous studies, they can exclude studies that are likely to yield biased findings. As discussed previously, however, there are very few universally agreed upon criteria for determining methodological rigor. Consequently, attempts to implement such criteria across studies are often done so unreliably (Lipsey & Wilson, 2001). Furthermore, Glass (1976, p. 4) contended that “it is an empirical question whether poorly designed studies give results significantly at variance with those of the best designed studies.” Accordingly, Glass encouraged synthesists to code the methodological characteristics of their studies, and then to empirically test their effects in the meta-analytic model. If study outcomes are found to vary as a function of certain methodological characteristics, then a determination can be made to either exclude the lower quality studies, or to report those results separately. If study outcomes do not vary as a function of methodology, however, then all the study outcomes can be combined in the final analysis.[64] Thus, the Glassian meta-analytic approach seems to not only overcome the problem of unreliability when selecting criteria for methodological rigor, but it also provides an

[64] Slavin (1986) correctly pointed out that Glass’ proposal for empirically identifying methodological moderators means that synthesists are charged with rejecting the null hypothesis that there is no systematic methodological variance. In cases where sample sizes are small, however, meta-analyses are more likely to make Type-II errors, incorrectly concluding that methodological criteria have no bearing on study outcomes. Synthesists need to be aware of this limitation, and should consider approaches that will minimize the problem (e.g., using a liberal alpha-level).

efficient means for determining the presence of methodological characteristics that may systematically influence study results. Accordingly, this method of controlling for methodological rigor is widely accepted (e.g., see Hunter & Schmidt, 2004), and is used in this study.

In sum, the methods proposed by Glass (1976) propelled the research synthesis field forward by leaps and bounds, providing a more accurate and efficient means of summarizing a body of literature. As with any new methodology, however, there were some notable limitations. Primary among these limitations was the assumption that all the observed variance in the literature was real or substantive (and not a result of sampling error or other artifacts). While ignoring the effects of sampling error is not likely to change the substantive conclusions about an average effect size (given that sampling error is random), it does make it difficult, or even impossible, to arrive at accurate conclusions about the presence of moderators. Ignoring the effects of other artifacts (e.g., measurement error or restriction of range), however, can influence conclusions made not only about the presence of moderators, but *also* about the average effect size. In addition, Glass mistakenly used the effect size as the unit of analysis instead of the study, thereby introducing statistical dependency into the analysis. While this mistake is notable, the effect on the statistical model is conservative, making the identification of moderators more difficult due to the loss of power. Finally, the Glassian meta-analytic method has been criticized for lumping all studies together regardless of methodological quality, and for comparing “apples and oranges” by combining multiple independent and dependent variables together. These criticisms do not apply to the Glassian meta-analytic method per se, however, but to the way in which the method has been applied. In fact, synthesists can overcome these latter limitations by empirically testing potential method moderators (to determine if study quality makes a difference), and by running separate meta-analyses for different groupings of independent and dependent variables so that the results can be easily interpreted. Overall, then, Glass helped move the field of research synthesis forward, but left the perfection of this new approach to future synthesists.

Study Effects Meta-Analysis

In an early attempt to overcome the weaknesses of the Glassian meta-analytic method, synthesists began implementing changes that would later be referred to as “study effects meta-analysis” (see Bangert-Drowns, 1986, p. 391). These changes consisted of addressing some of the limitations associated with the Glassian approach. For example, synthesists began using criteria to evaluate the methodological rigor of studies being considered for inclusion in the meta-analysis (e.g., whether or not a treatment meets criteria established by experts in the field). Applying these criteria allowed synthesists to eliminate studies with “deficiencies judged serious enough to distort study outcomes” (Hunter & Schmidt, 2004, p. 458). Second, synthesists began focusing their meta-analyses on the relationship between *specific* independent and dependent variables instead of lumping variables together. Narrowing the focus allowed synthesists to answer specific questions about a particular treatment or relationship instead of painting a broad picture as Smith and Glass (1977) had done in their meta-analysis. Finally, the unit of analysis was changed from an effect size to a study. In cases where multiple effects were available from a study, the effects were averaged together to provide the synthesist with one outcome per study. In this way, synthesists were able to avoid violating the statistical independence assumption, and thereby minimize the probability of making Type-II errors when evaluating moderator variables (Hunter & Schmidt, 2004). Together, these changes allowed the field of research synthesis to take another step forward; yet, the most glaring weakness (i.e., sampling error) had yet to be addressed.

Homogeneity Test-Based Meta-Analysis

In the next evolution of the meta-analytic method, Hedges (1982) and Rosenthal and Rubin (1982a,b; Rosenthal, 1991) attempted to address the problem of sampling error. Specifically, they introduced a chi-square statistic, *Cochran's Q* (Cochran, 1954), to determine if the observed variance between study outcomes was more than could be expected from sampling error alone. A non-significant *Q*-test indicates that the observed variability is nothing more than sampling error. It also means that the synthesist can forgo moderator analyses, thereby avoiding potential Type-I errors that might otherwise have been made under the assumption that the observed variability was real. A significant *Q*-test, on the other hand, indicates that real variance exists, and the synthesist can commence with moderator analyses in an attempt to identify the source of the real variation between studies. Thus, the introduction of the *Q*-test represented a marked improvement over the Glassian approach.

Despite the conceptual advantages of a statistical test of heterogeneity, the use of Cochran's *Q* represented a return to one of the practices that made it difficult to review a literature in the first place: A reliance on significance tests. Because *Q* is distributed as chi-square, the synthesist must be wary when the sample of studies is either too small or too large. If the sample is too small, *Q* may lack the statistical power to detect real variance between studies when it exists. If the sample is too large, *Q* will be biased against the null hypothesis, and can lead the synthesist to conclude that real variability exists when, in fact, it does not. Thus, a reliance on the *Q* statistic to determine the presence of real variance between studies can lead to both Type-I and Type-II errors.

In addition, the homogeneity test-based meta-analytic method ignores the potential influence of other artifacts, such as measurement error and restriction of range that can cause artifactual variability between study outcomes. As Hunter and Schmidt (2004) point out, ignoring these artifacts can lead synthesists to conclude that study outcomes are heterogeneous when, in fact, they are not. Thus, while the method proposed by Hedges (1982) and Rosenthal and Rubin (1982a,b) helped synthesists take another big step forward, and perhaps their biggest since the introduction of the meta-analytic method, the remaining problems with this method were still numerous and potentially problematic.

Psychometric Meta-Analysis (Hunter-Schmidt Method)

The most recent development in the meta-analytic method has been referred to alternately as validity generalization (e.g., see Schmidt, Hunter, Pearlman, Rothstein, Sackett, Schmitt et al., 1985) psychometric meta-analysis (e.g., see Hunter & Schmidt, 2004) and the Hunter-Schmidt method (e.g., see Borenstein et al., 2009). The latter name stems from the notable contributions that Hunter, Schmidt and colleagues (Hunter & Schmidt, 1990, 2004; Hunter, Schmidt & Jackson, 1982; Schmidt & Hunter, 1977) have made in the advancement of this method. The distinguishing feature of the psychometric method is that it provides a mechanism for minimizing or removing the effects of several artifacts, including sampling error, measurement error, restriction of range, artificial dichotomization of continuous variables and others (e.g., see Hunter & Schmidt, 2004, p. 35). This is the *sine qua non* for synthesists because it allows them to get away from simply describing the results of imperfect studies, and onto describing what would have happened if the studies had been conducted perfectly (Hunter & Schmidt, 2004). As Rubin (1990, p. 157) stated,

> We really do not care *scientifically* about summarizing this finite population [of observed studies]. We really care about the underlying scientific process—the underlying process that is generating these outcomes that we happen to see—that we, as fallible researchers, are trying to glimpse through the opaque window of imperfect empirical studies.

The removal of artifactual variance allows synthesists to “glimpse through the opaque window,” and view the underlying processes at work. It is this ability, Hunter and Schmidt (2004) argue, that provides the basis for the development of cumulative knowledge, sound theory and informed policy. In short, the ability to remove known sources of error from a collection of studies allows synthesists to provide the types of answers for which researchers and policymakers have long been searching.

Despite the state-of-the-art nature of psychometric meta-analysis, synthesists have pointed out that these methods are limited in their scope of application (e.g., see Borenstein et al., 2009; Johnson, Mullen & Salas, 1995). For example, Borenstein et al. (2009) noted that the psychometric meta-analytic method was developed primarily for synthesists who are investigating continuous outcomes with standardized effect sizes, such as correlation coefficients (*r*) and standardized mean differences (*d* or *g*). And, although this method has been extended for use with proportions (e.g., see Viswesveran & Schmidt, 1992), Borenstein et al. (2009, p. 344) note that it can produce “very different results” when applied to binary data and proportions given the reliance on sample size to weight studies as opposed to the inverse-variance weighting scheme proposed by Borenstein et al. (2009, p. 344). Furthermore, Johnson et al. (1995) noted that it is rare for information on artifacts (e.g., reliability or restriction of range) to be reported within studies. This essentially creates a missing data problem given that the ideal psychometric meta-analysis involves removing the study-level artifact from the study-level effect.

Consequently, the psychometric method is often not a practical choice, especially for binary data and proportions.

Importantly, both practical limitations present problems for the study of religious service attendance. First, attendance is measured either on a binary scale (e.g., attended in last 7 days or not) or as a proportion (e.g., the proportion of a sample reporting regular attendance). Second, there is little, if any, available artifactual information related to the measurement of attendance.[65] Thus, the psychometric method advocated by Hunter and Schmidt (2004) is neither an optimal nor a practical choice for cumulating attendance data. Fortunately, Borenstein and colleagues have developed a similar method that is capable of not only cumulating a wide variety of study outcomes (e.g., standardized effect size measures, binary outcomes and proportions), but also of identifying, estimating and partialling out some of the measurement-related artifactual variance, such as restriction of range, through the use of study-level characteristics as control variables or moderators.[66] This method of cumulating findings and controlling for artifacts is referred to as "Conventional Meta-Analysis" with "Artifact Correction."

Conventional Meta-Analysis with Artifact Correction

Conventional meta-analysis distinguishes itself from the Hunter-Schmidt method in several important ways. The first distinction lies in the way in which the two methods go about weighting and cumulating study effects. The Hunter-Schmidt method weights each study by its sample size. Large sample studies are given more weight and contribute more to the overall summary effect(s) than small sample studies, with the result being a sample-weighted mean. The

[65] Reliability is typically the most readily available artifact indicator, but because internal consistency reliability cannot be estimated from the single attendance item, and because most studies that use the attendance indicator are not concerned with the test-retest reliability, this information is largely, if not completely, unavailable.
[66] The Hunter-Schmidt method of removing artifacts can also be applied to effect sizes prior to cumulating them using the Borenstein et al. (2009) method, but the availability of such information precludes the use of these techniques in this study.

conventional method, on the other hand, uses the inverse of a study's variance to weight its contribution to the overall summary effect(s). Using the variance to weight each study's contribution is important because the variance not only includes information about a study's sample size, but also on two potential sources of sampling error: (a) The sampling error that results from sampling people into studies (i.e., the within-studies sampling error) and, in random-effects models, (b) the sampling error that results from sampling studies into a meta-analysis (i.e., the between-studies sampling error). Under this weighting scheme, studies with small variance estimates (i.e., those with larger sample sizes and less sampling error) are given more weight, while studies with large variance estimates (i.e., those with smaller sample sizes and more sampling error) are given less weight. When the study effects are cumulated, each study contributes proportionally to the cumulative effect based not only on study sample size, but also on the precision with which the study effect was estimated. As mentioned above, this difference in weighting methods can lead to "very different results" when cumulating binary data and proportions (see Borenstein et al., 2009, p. 344). Thus, it is preferable to use the inverse-variance weighting method for cumulating religious service attendance data.

The second distinction between the two methods lies in the decision-making process leading to moderator analyses. With the Hunter-Schmidt method, the synthesist compares the amount of variance expected from sampling error to the observed variance. If the sampling error variance is equal to, or greater than 75% of the observed variance, Hunter and Schmidt (2004) recommend forgoing moderator analyses based on the assumption that *all* of the remaining variance comes from artifacts that could not be identified or measured. Should the amount of variance expected from sampling error account for less than 75% of the observed variance, however, moderator analyses can be carried out as planned. In this way, the Hunter-Schmidt method minimizes the potential for capitalizing on chance. The conventional method can also make use of the 75% guideline, but Borenstein et al. (2009) push synthesists to go further by discerning the underlying message communicated by three variance indicators.

The first variance indicator is Cochran's *Q*. As mentioned in the discussion on homogeneity-test based meta-analysis, the *Q* statistic provides a test of the null hypothesis that the observed variation is nothing more than what can be expected by chance. Recognizing that *Q* is sensitive to the number of studies and that it only provides an indication of *whether*, but not *how much* real variability exists, a second variance indicator was introduced by Higgins, Thompson, Deeks and Altman (2003). Higgins et al. introduced a proportion of variance statistic, I^2, to determine the proportion of observed variance that is attributable to systematic differences between study outcomes. This statistic provides a nice complement to *Q* in that it provides an indication of the proportion of total variability that is available to be explained by moderators. In fact, if the analyst subtracts I^2 from one (i.e., $1 - I^2$), the difference is the same variance indicator used by Hunter and Schmidt (2004), which allows for the use of the 75% guideline. As alluded to above, however, the I^2 statistic does not provide an indication of the *absolute* variability between study outcomes. As a result, Borenstein et al. (2009) use two variance components, tau (τ) and tau-squared (τ^2), to assess the total amount of variance between study outcomes. The former (τ) represents the standard deviation of the distribution of study effects while the latter (τ^2) represents the between-studies variance. Because these values are on the original metric of the effect size, the analyst can gain a sense of the absolute variability that exists between study effects. Putting all three variance indicators together allows the synthesist to determine: (a) if the between-studies variance is significantly greater than what could be expected by chance; (b) the proportion of total variance that is real or systematic; and,

(c) the total amount of variation that exists between study effects. Thus, the synthesist can gain a much better understanding of whether moderator analyses are warranted, or if they are likely to result in Type-I or Type II errors.

The fourth distinction between the two methods lies in how they deal with artifacts other than sampling error. In the Hunter-Schmidt method, synthesists are encouraged to remove artifactual variance from each study effect prior to cumulating. The conventional method, on the other hand, approaches this problem by attempting to statistically model the effects of artifacts by using sub-groups analysis (for artifacts represented by discrete variables) and meta-regression (for artifacts represented by either discrete or continuous variables). A synthesist could use sub-groups, for example, to separately analyze: (a) studies that used a continuous outcome and (b) those that artificially dichotomized a continuous outcome. If the strength of a moderator relationship is significantly weaker (due to attenuation) in the latter sub-group, then the synthesist could point to artificial dichotomization of outcomes as the culprit; focus solely on the results obtained from the continuous outcome sub-group; and, as a result, effectively eliminate the error variance associated with artificial dichotomization. Continuing with this example, consider that some of the remaining studies have restricted participation to certain age groups (e.g., adolescents or the elderly), thereby creating a potential restriction of range problem. If enough age-restricted samples are available to represent the life-span (e.g., some studies restrict their sample to youth and adolescents, while others restrict their samples to young adults, middle-aged adults and the elderly), then the average age of a sample can be included as a covariate in a meta-regression model. If restriction of range is a problem, then its coefficient will be statistically significant, and the amount of variation attributable to its artifactual effects will be effectively partitioned, leaving it unavailable to be explained by other moderators.

In sum, Borenstein et al. (2009) state that while there is little difference in the results produced by the Hunter-Schmidt method and the conventional method when using standardized effect sizes (e.g., *r* and *d*), there are real differences between the two methods when using binary measures or proportions as outcomes. Because attendance is measured using both binary measures and proportions, the conventional method of meta-analysis is more appropriate for this outcome. In addition, the conventional method's use of inverse-variance weighting allows for the inclusion of more study-level information into the weighting of a study's contribution to the cumulative effect(s). The conventional method also uses a more sophisticated decision-making process for determining whether moderator analyses are appropriate. Finally, the conventional method uses a practical approach to artifact correction that most researchers are familiar with (i.e., group blocking comparisons and statistical modeling via regression). Given the appropriateness and key advantages of the conventional method for studying religious service attendance, this study will follow the procedures outlined by Borenstein et al. (2009).

Meta-Analysis of Religious Service Attendance

As mentioned above, Borenstein and colleagues (2009) have made key contributions to the development and advancement of the conventional meta-analytic method. Of particular importance to the present study is their extension of the method to handle a wide variety of study outcomes, including binary outcomes and proportions. In addition, Borenstein, Hedges, Higgins and Rothstein (2005, 2015) developed a highly flexible and user-friendly software program, *Comprehensive Meta-Analysis, Version 3.3* (CMA), to meta-analyze study effects data using the conventional method. Accordingly, the procedures outlined by Borenstein et al. (2005, 2009, 2015) were used to guide the present study. In addition, other useful sources (e.g., Cooper et al., 2009; Hedges & Olkin, 1985; Hunter & Schmidt, 2004; Lipsey & Wilson, 2001; Schulze, 2004) were used to supplement specific aspects of the process. For example, Cooper et al. (2009) provide detailed guidelines for conducting a thorough literature search; and, Lipsey and Wilson (2001) address unique issues associated with database preparation and management. Thus, the procedures used in this meta-analysis were mainly guided by the work of Borenstein et al. (2005, 2009, 2015), but were also informed by other notable sources. Before going into the details of the procedure, however, it is important to operationally define religious service attendance.

Operational Definition of Religious Service Attendance Frequency

In this study, measures of religious service attendance were limited to those that allow for an estimation of the frequency with which people attend religious services, where religious services are inclusive of "Church services," "worship services," "Mass," "Synagogue," "Chapel," "Temple," "Mosque" or some other regularly held service involving religious rituals that focus on a higher power or the sacred. Estimating attendance frequency implies that response scales provide individuals an opportunity to indicate "how often" they attend religious services. For example, ordinal response scales, such as those ranging from "Less than once per year" to "2 to 3 times per month" to "More than once a week" allow the respondent to specify how often they attend per year, per month or per week, respectively. Dichotomous response scales that ask respondents to indicate whether they attended in the past week (e.g., "Did you attend in the past seven days or not?" "Yes" or "No") or yesterday (*cf*. ATUS time use item) are also acceptable because they provide a measure of attendance frequency within the past week. Scales that do not provide this level of behavioral frequency, however, cannot be included. For example, Van Wagoner (2016) asked respondents if they are "Currently attending" or "Not currently attending." While this measure provides an indication of whether respondents are attenders, it does not tell us how often they attend, and, therefore, cannot be included in this study.

Moreover, measures that *explicitly* incorporate activities other than religious services, such as prayer meetings, scripture studies or choir practices, either as a single item or a composite scale, cannot be included in this study because they are measuring the frequency of several religious activities, only one of which is religious service attendance.

Literature Search

The literature search is one of the most important aspects of a meta-analysis for some of the same reasons that good sampling is important in primary research. For instance, the thoroughness of the literature search (like the quality of a sampling frame in primary research) has direct implications for the representativeness and generalizability of the meta-analytic findings. If the literature search is less than thorough, essential data sources may be omitted from the analysis, thereby altering (or biasing) the findings and limiting their generalizability (Cooper et al., 2009; Rothstein, Sutton & Borenstein, 2005). In addition, the ability of the synthesist to detect moderating effects is directly related to the number of studies located during the literature search. As more studies are located and retrieved, statistical power increases, thereby providing the synthesist with a greater chance of detecting moderators. This latter point is important given that statistical power is often poor in meta-analytic studies, where sample sizes (i.e., the number of study effects) tend to be small (Hunter & Schmidt, 2004). Thus, a quality literature search is paramount to achieving a representative sample; detecting moderating effects, if they exist; and, generalizing the results beyond the sample of studies. To this end, a series of steps were undertaken to acquire data from three general sources: (1) published literature; (2) grey literature; and, (3) primary data sources.

Published Data Sources

Although the term, *published literature* is frequently used, its defining characteristics can be fuzzy. For example, Cooper et al. (2009) state that researchers often assume that published literature refers only to works that have undergone a peer-review process. Yet, Strong (1990), writing from the perspective of copyright law, states that published literature can include "any document for which copies will be supplied to any requester" (cited in Cooper et al., 2009, p. 61). Similarly, Cooper et al. (2009) go on to state that published literature can include any document produced by a commercial entity whose primary objective is to publish. While these latter definitions may be technically accurate, they are not entirely functional. These definitions assume that published documents include those that are, or can be made, *open* to the public. Yet, not all documents that are open to the public are equally accessible. For example, a newsletter or fact sheet produced for a local audience will remain obscure and inaccessible to a researcher who is not part of the local audience. Similarly, documents produced for a conference may only be accessible to those in attendance if the organizing body does not archive and make the conference documents (or conference program) publicly available. And, a book that only briefly touches on the subject of interest will likely not be discovered and accessed by a synthesist. In each of these cases, the sources are publicly available, but are not as accessible as other sources. Thus, both *availability* and *accessibility* are important determinants of what we refer to as the "published literature."

Using these guidelines as a starting point, this study generally treats the following sources as published literature: Electronic and print research journals; prominent books in the religious research field; electronically available conference proceedings; theses and dissertations that can be accessed either electronically, in print or by lender request; and, fact sheets, reports and newsletters that are circulated, or made available, to a national audience (e.g., members of a research society). Locating these sources will involve the following search procedures: (a) Electronic bibliographic database searches; (b) manual scans of religious research journals and special issues on religion; (c) manual scans of prominent books that focus on the scientific study of religion; (d) forward citation index searches; and, (e) manual scans of reference sections in studies obtained from other search procedures. Together, these procedures should yield a relatively large number of studies with religious service attendance data.

Grey Literature

Synthesists have begun referring to unpublished data sources as *grey literature* (Cooper et al., 2009). While several definitions of grey literature have been offered (e.g., see Auger, 1998; McKimmie & Szurmak, 2002; Rothstein & Hopewell, 2009), Weintraub (2012, p. 1) defines it as including "newsletters, reports, working papers, theses, government documents, bulletins, fact sheets, conference proceedings and other publications distributed free, available by subscription or for sale." Rothstein and Hopewell (2009, p. 105) add that grey literature can also include books and book chapters "because . . . studies located in them are often difficult to identify through typical search procedures." In general, then, grey literature is inclusive of most data sources that are not commercially available, or that cannot be easily located or accessed via electronic search engines or indexes (Cooper et al., 2009). In order to locate as many of these sources as possible, the following search procedures were employed: (a) manual searches of books and book chapters focusing on religious behavior or involvement; (b) electronic and manual searches of theses and dissertations; (c) electronic and manual searches of conference proceedings; (d) manual scans of religious research newsletters and fact sheets; (e) electronic searches for religious research websites; and, (f) data source requests of research groups and colleagues that have conducted religious research.

Successfully navigating these procedures requires a fair amount of ingenuity and tenacity, as well as an open, flexible and dynamic approach to the search process. For example, search terms often need to be modified to yield a greater number of hits during electronic searches and multiple collaborators frequently need to be contacted if responses to initial data requests go unanswered. Successful implementation of these procedures, however, increases the generalizability of the findings.

Primary Data Sources

Because religious service attendance data are collected by national research groups on a consistent basis (e.g., see the Gallup Poll; Gallup Organization, 1985), and because primary data are preferable to secondary data, it is important to contact, and request data from as many groups and organizations that collect religious data as possible. Where data sharing is not possible, reports providing summary statistics should be requested. Importantly, these reports may be more complete than what is published in books or research articles given that they are requested for the specific purposes of this study. Locating research groups and organizations that can provide access to primary data (or private reports) involves: (a) conducting electronic searches for probability surveys conducted by national polling organizations; (b) scanning the published and grey literature to identify research organizations that collect data on attendance using probability surveys; (c) scouring social scientific research clearinghouses for survey data on attendance; and, (d) asking colleagues for suggestions on, or access to, additional primary data sources.

Each of these data sources will be described in more detail, but first it is necessary to discuss the key-word search terms to be used with the electronic search procedures.

Key-Word Search Terms

The selection of appropriate key-word search terms is an iterative process that develops as the synthesist learns to negotiate the nuances of different search engines and databases. To start, it is recommended to use broad, over-inclusive terminology so that the synthesist will not only be able to retrieve studies from anticipated areas of study, but from new, unanticipated areas as well (Cooper, 2009). For instance, attendance data are expected to come from the bio-psycho-social sciences, religious research organizations and from national surveys that offer a religious focus, but it is also possible that attendance data will be found in studies of civic participation, economics, history, philosophy and the like. As the search process progresses, however, the synthesist can narrow the focus of the search terms to those generating the most hits.

Table 3 contains the key-word search terms to be used during the electronic searches. During the initial phase, two broad religious and spiritual prefixes (i.e., "Relig" and "Spirit") can be coupled with wildcards (e.g., symbols, such as "?" or "*", that stand for a fill-in-the-blank character)[67] to locate all literature containing the respective prefixes. For example, the search term, "Relig*" (or a variant of this) returns results for all words containing "Relig," such as "Religion," "Religiosity," "Religious," "Religiously" and so on. If this type of search yields too many potential hits, however, secondary prefixes, root words and other terms (see Table 3) can be used to narrow the search results to those focusing on religious behavior. A variety of other search terms can also be used to represent categories that are likely tied to religious behavior. For example, Table 3 provides several categories (e.g., *Names of Religious Services*) and

[67] Each search engine uses a unique set of wildcards so it is necessary to identify the appropriate wildcards prior to each search.

associated search terms (e.g., “Church,” “Mass,” “Service,” “Temple,” “Worship”) that are likely tied to studies focusing on religious behavior. Categories also exist for *Religious/Spiritual Instruments* because multi-item religious and spiritual instruments often include attendance items.[68] Because religious extremes are often underrepresented in the literature, efforts should also be made to include the irreligious (via search terms such as, “Nones,” “Unchurched,” “Atheist” and the like) and the hyper-religious (via search terms such as, “Born Again,” “Evangelical,” “Fundamentalist” and the like; see Table 3). Finally, very specific primary search terms, such as “Religious Service Attendance” and “Church Attendance,” can be used to verify that no sources have been left undiscovered (see Table 3).

Locating Published Literature

As stated above, the search procedures used for locating and retrieving published literature include the following: electronic bibliographic database searches; manual scans of religious research journals and special issues on religion; manual scans of prominent books that focus on the scientific study of religion; forward citation index searches; and, manual scans of reference sections in studies obtained from other search procedures.

Electronic Bibliographic Database Searches. The key-word search terms described above (see also Table 3) can be used in conjunction with the electronic bibliographic databases (see Table 4) available to this author through the online libraries at Judson University (2016) and Southern Illinois University at Carbondale (2016). Although these databases can be searched individually, the emergence of search engines allow researchers to search the contents of multiple databases simultaneously. For example, *FirstSearch* allows users to simultaneously search the contents of the following bibliographic databases: *ArticleFirst, Ebooks, ECO, ERIC,*

[68] It was previously mentioned that attendance items that are part of a larger scale are to be omitted. Before omitting these potential data sources, however, attempts were made to request the attendance-specific data from the authors who used or developed the respective instruments.

GPO, IllinoisCatalog, MEDLINE, OAIster, PapersFirst, Proceedings, WorldCat and WorldCatDissertations (see Table 4). *EBSCOhost* allows users to simultaneously search the contents of over 100 bibliographic databases, some of which overlap with *FirstSearch* (e.g., *Ebooks* and *ERIC*) and other databases. Because search engines overlap in their bibliographic database coverage, only those that provide at least some unique content will be used.[69]

Conversely, not all bibliographic databases are included in the available search engines; thus, searches of excluded databases in relevant fields of study should also be conducted (see Table 4). Importantly, the content indexed by the search engines and individual databases listed in Table 4 cover a wide variety of published and grey literature sources, such as blogs, books, conference papers, government documents, law reviews, legal cases, magazines, newspapers, consumer information reports, journal articles, theses and dissertations, wire services and more. The search engines and databases also cover the gamut of disciplines from which attendance data are likely to be found. Thus, the use of the key-word search terms identified in Table 3 in the search engines and individual databases described in Table 4 should allow for the identification of many potentially relevant studies.

Manual Scans of Key Literature Sources. Religious service attendance is a preeminent indicator of religious behavior and religiosity. Accordingly, attendance is included in studies as both a primary and secondary variable. When it is included as the latter, it is often reported as a demographic to describe the general religiosity of a sample (Larson, Pattison, Blazer, Omran & Kaplan, 1986). This reporting tendency has two implications. First, attendance data are likely to be in good supply, which is encouraging considering the number of potential moderators to be tested. Second, *locating* studies that include attendance as a secondary variable is relatively difficult because secondary variables are typically not indexed as key terms in electronic

[69] *Ovid*, for example, is not included here because it overlaps with the content of other databases.

bibliographic databases, thereby making them invisible to the search. To minimize this problem, manual scans of "high probability" literature sources (i.e., those most likely to contain attendance data) were conducted.

Four high probability published literature sources were identified for this study. The first includes religious research journals and special issues on religion. The religious research journals to be scanned include the following: *International Journal for the Psychology of Religion, Interdisciplinary Journal of Research on Religion*, *Journal for the Social Scientific Study of Religion*, *Journal of Religion and Health*, *Journal of Religion, Spirituality and Aging*, *Psychology of Religion and Spirituality*, *Religious Research Review* and *Sociology of Religion*. Second, manual scans are also to be carried out on special issues of non-religious journals that focus, in-part or in-whole, on the social scientific study of religion or spirituality. For example, an issue in *American Psychologist* (volume 58, issue 1) devoted four articles to the study of Spirituality, Religion and Health, while *Psychological Inquiry* (volume 13, issue 3) devoted an entire issue (7 articles) to the study of Religion and Psychology. Third, manual scans are to be conducted on journals that produced the greatest number of "hits" during the electronic bibliographic search process. While it seems likely that the religious research journals listed above would yield the most hits, it was hoped that this process would yield a few additional non-religious journals that also contain a sizeable number of hits. Fourth, the literature obtained for this study's introduction (Chapter I) and literature review (Chapter II) was scanned given the high density of attendance data in these studies.

A two-step process was used to manually scan potential data sources. The first step involved examining the method and results sections of each study or literature source for attendance data. If attendance data were fully reported, then these data were extracted from the source directly. If the method or results sections indicated that attendance data were collected, but are not fully reported, then the author(s) of the study was contacted with a request to provide the necessary data. The second step involves manually scanning the reference sections of each study so that additional potential hits could be identified, acquired and included in the meta-analysis. While time consuming, manual reviews are nevertheless a useful and important method of supplementing the data obtained from the bibliographic database searches.

Forward Citation Search Indexes. Forward citation search indexes provide a means of identifying additional studies that have cited a *known* work or author (White, 2009). Specifically, prominent studies and notable researchers in the field of religious studies[70] are entered (either by title or author) into a forward citation search index. The index then returns a list of publications that have cited the original source, thereby allowing synthesists to identify additional studies that offer a similar focus as those already identified. As White (2009) points out, this process allows synthesists to retrieve data sources from areas of study that lie outside the typical realm of interest. Thus, forward citation search indexes are a good method for increasing the cross-disciplinary representativeness of a literature search.

There are now several forward citation search indexes available to synthesists. Prominent among these indexes are the *Science Citation Index*, the *Social Sciences Citation Index*, the *Arts and Humanities Citation Index*, the *Conference Proceedings Citation Index* and the *Book Citation Index*. Importantly, the *Web of Science* (described in Table 4) allows synthesists to simultaneously search all five indexes. *Web of Science* also incorporates *Mann's Related Records Search*, which allows users to locate additional articles that share common citations with

[70] Prominent studies can include those that are central to the formation of hypotheses in this study, or those that are cited frequently by others in the field. Notable authors can include those who have made particularly important contributions to the field of religious studies, or those who have repeatedly published in the area.

the target article. Thus, the forward citation search tool should help complement the other search procedures in locating attendance data from both anticipated and unanticipated fields of study.

Locating Grey Literature

The following methods can be used for locating and retrieving grey literature: manual scans of books and book chapters, searches and manual scans of conference proceedings, primary data acquisition and colleague requests.

Manual Scans of Books and Book Chapters. Because individual study data can be difficult to locate in a book (hence the classification of books and book chapters as grey literature), it is a good idea to manually scan books and book chapters that have a high probability of containing religious service attendance data. Specifically, books and book chapters that focus on religious behavior or that limit themselves to a review of the literature pertaining to the biopsychosocial study of religion should be manually scanned. The reference sections of these sources can also be scanned so that as many potential data sources can be identified as possible.

Searches and Manual Scans of Conference Proceedings. The proceedings from major religious research conferences can be searched. Specifically, these conferences include the *Association for the Sociology of Religion*, the *Religious Research Association*, the *Society for the Psychology of Religion and Spirituality* and the *Society for the Scientific Study of Religion*. The proceedings from these conferences dating back as far as possible can be obtained (either in electronic or paper format), and either electronically searched (where possible) or manually scanned to identify studies focusing on religious behavior. In cases where the original papers and presentation materials are not available, attempts were made to contact the original authors with a request for both the attendance and moderator data. It should be noted, however, that some conference proceedings are included in electronic search databases and indexes (see Table

4). For example, the database, *Proceedings,* in *FirstSearch* offers users an opportunity to electronically search conference proceedings, and *Web of Science* allows users to perform forward citation searches on conference proceedings (see Table 4). Thus, it should be possible to locate a number of grey literature sources by conducting both electronic and manual searches of religiously focused conference proceedings.

Colleague Requests. Finally, a reference list of all studies and data sources acquired through each of the previous search procedures should be compiled and sent to several audiences (e.g., authors who have published numerous times in the field, colleagues who have undertaken religious studies and to the listserv's or other communication outlets of professional religious research associations) with a request to suggest additional studies or data sources that might contain attendance data. This procedure is expected to yield a small number of unpublished (or unlocated) studies, and can serve to make the literature search as thorough and representative as possible.

Locating Primary Data

National polling agencies, religious research groups, religious granting foundations and national clearinghouses are all potential sources of primary religious service attendance data. Examples of these data sources include the following: *Association of Religion Data Archives (ARDA)*, *Barna Group, Center for Spirituality*, *Theology and Health* at Duke University, *Gallup Organization*, *General Social Survey (GSS), Religious Congregations and Membership Study*, *Pew Research Center*, the *Templeton Foundation* and the University of Michigan's *Inter-University Consortium for Political and Social Research* (ICPSR). Additional primary data sources were identified during the respective searches of the published and grey literature, and by conducting a separate search for similar entities that are likely to collect attendance data.

Taken together, the search procedures outlined above for locating published literature, grey literature and primary data should yield a relatively large number of studies and data sources that are representative of the religious service attendance data that have been collected by researchers over the majority of the past century. In turn, these data sources should allow for a precise estimate of the prevalence of attendance in America, and to a more thorough understanding of how methodological and socio-demographic factors influence the prevalence estimate.

Inclusion and Exclusion Criteria

Studies to be included in the meta-analytic sample must provide enough data so that a frequency of attendance estimate can be computed. This includes the total size of the sample reporting their attendance (N), and the number (n) or proportion (p) of people in the sample who reported attending at a specific rate, frequency or point in time, such that a probability of attending on any given Sunday can be approximated.[71] Data that *explicitly* represent other outcomes (e.g., frequency of attendance at *other* religious events or activities, such as prayer meetings, scripture studies or choir practices), those that combine attendance with other measures of religiosity (e.g., a four-item measure of religious commitment) and those that ask respondents to retrospectively indicate their attendance (e.g., the 2008 International Social Survey Programme asks, “And what about when you were around 11 or 12, how often did you attend religious services then?) were excluded.[72] Where possible, however, the authors of studies that report either too little information to compute an attendance estimate or that use a

[71] Some response scales are too vague or general to be of much use. For example, Mason et al. (2013) asked respondents a dichotomous question: “Do you attend services about once a week or more, or less often than once a week?” While the former response option provides an indication of the proportion attending weekly or more, the latter is too vague to produce a comparable estimate of the respondents’ likelihood of attending on any given Sunday. Thus, this item, and others like it, were omitted from the study sample.

[72] As noted in the literature review, the ambiguous wording of some attendance items makes it impossible to exclude all non-religious service activities. Thus, items that *leave open the possibility* of incorporating non-service-related activities were coded so that the variability due to the wording of these items can be partitioned and assessed. Items that *explicitly include* non-service related activities, however, were excluded.

convoluted measure of attendance were contacted and asked for more specific data that would allow for an estimation of attendance frequency.

Besides measurement issues, sample data must have been collected in the U.S., and be composed of American participants[73] to be included in this study given that the focus is on estimating the prevalence of attendance in America. Furthermore, only attendance data collected between 1939 (i.e., the first year in which Gallup data were collected on attendance) and 2017 were included in this study because the earliest known "gold standard" attendance estimate dates back to 1939 (see Gallup & Jones, 1989). Finally, non-published attendance data provided by religious denominations and congregations were excluded given that they are: (a) likely to use a wide variety of data collection methods that may not be comparable to those used by social scientists; and, (b) unlikely to provide an accurate estimate of the population from which their counts are based.[74]

Data Extraction and Coding

After gathering all available sources of religious service attendance data, the synthesist is faced with the task of extracting the relevant information from each study. According to Lipsey and Wilson (2001), this information generally revolves around four types of data: (1) study effects; (2) theoretical moderators; (3) methodological moderators; and, (4) contextual moderators. Study effects data refer to the outcomes of interest for the meta-analysis (i.e., the information needed to calculate a prevalence of attendance estimate), while the latter three types refer to those factors that could plausibly influence (or moderate) the study effects data.

[73] For a study to be included in this meta-analysis, it is not necessary that all sampled persons be U.S. Citizens, but that the sample includes only foreigners who are currently living in America (either temporarily for work or school, or permanently as they work toward citizenship) at the time of data collection.
[74] Membership lists would be the most obvious indicator of the respective population sizes, but these may be too infrequently updated to provide a useful reference population from which to calculate the prevalence of attendance.

All else equal, it is important to identify and code as many of the theoretical, methodological and contextual moderators as possible so that: (a) their influence can be empirically examined during the analysis; and, (b) they can serve as control variables in multivariate meta-analytic models. Accordingly, Lipsey and Wilson (2001) recommend using an over-inclusive approach to coding and extracting moderator data.[75] This approach requires the synthesist to go through an iterative process where they: (a) start with a list of theoretically identified moderators to code and extract; (b) code and extract moderator data from a sample of studies; (c) assess the availability and quality of the extracted moderator data; (d) use the assessment to identify moderators to add (if they are commonly reported and of sufficient quality) and eliminate (if the data are of poor quality or are largely missing) provided they are not of central theoretical importance to the study; (e) continue coding and extracting using the new list of moderators; and, (f) repeat this cycle as study data continue to be coded and extracted. The initial list of moderators to be coded (i.e., the starting point), along with the study effects data, are described below using the framework provided by Lipsey and Wilson (2001).

Study Effects Data

As stated above, study effects data refer to the data elements associated with the outcome of interest. This includes the following data elements, which are necessary to compute a frequency of attendance estimate: (a) the total size of the sample reporting their attendance (N), and either (b) the number (*n*) or (c) the proportion (*p*) of people in the sample who report attending at a specific rate, frequency or point in time. The study effects data to be extracted also include: (a)

[75] Hunter and Schmidt (2004) warn against extensive moderator testing given that, in their experience, most between studies variance is due to sampling error and not to systematic differences between studies. Hunter and Schmidt also point out that most meta-analyses lack adequate statistical power to detect moderating effects even when they exist, and consequently, attempts to identify these moderators often capitalize on chance. While this caveat should be heeded, it remains important to code and extract as much of the potential moderator data as possible so that the synthesist has the flexibility during the analysis to select an appropriate number of moderators to test based on both theoretical grounds and the final sample size; hence, the over-inclusive approach to coding and extracting is appropriate here.

the *source* of the measurement instrument, item or procedure (e.g., American Time-Use Study, Baylor Religion Survey, Gallup Poll, GSS, local study, etc.); (b) the exact wording of the question (or data collection protocol in the case of time-use or count-based studies); and, (c) the exact wording of the response options (or format) used in each study. Collecting information on these elements allows the synthesist to combine data from similar sources and similarly worded questions and response options during the analysis.

Theoretical Moderator Data

Theoretical moderators are variables that have been identified by previous research or theory as factors that could plausibly influence study outcomes. For example, in their meta-analysis on the prevalence of sexual harassment in the workplace, Ilies et al. (2007) theorized that measurement type would influence the reported prevalence of sexual harassment across studies. The findings confirmed this effect, with respondents reporting greater rates of sexual harassment when presented with behavioral checklists than when asked directly if they had ever been sexually harassed. Similarly, the literature reviewed here in Chapter II suggests that estimates of the prevalence of attendance vary as a function of measurement type, data collection mode and several socio-demographic characteristics. The coding and extraction of these data elements are described in more detail below.

Measurement Type and Data Collection Mode. The literature on religious service attendance suggests that attendance estimates vary as a function of the measurement instrument or procedure used to collect the data. Recall that the traditional Gallup and GSS items tend to yield attendance estimates of slightly greater than 40%. Items designed to minimize or eliminate problems with ambiguous wording, the lack of a temporal reference period and social desirability bias each seemed to reduce these baseline estimates by a couple of percentage points. Items that limited the time frame to within the past week seemed to further reduce attendance estimates by as much as 10% to 14%, and, most dramatically, when attendance counts were taken (instead of self-reported), the estimates were nearly halved. Thus, the way in which attendance is measured seems to have a systematic effect upon the prevalence estimates reported in the literature.

Accordingly, the method of measurement employed in each study was coded and extracted as two variables. The first variable to be coded and extracted attempts to ascertain whether the item minimizes the noted problems with the original GSS and Gallup items. That is, does the item or procedure minimize in some way (dummy coded as “0”), or does it suffer from (dummy coded as “1”) each of the following: (a) Ambiguous wording (i.e., where no attempt has been made to limit responses to regularly held service involving religious ritual that likely focus on a higher power or the sacred); (b) ambiguous time-frame (i.e., where no time-frame reference has been specifically stated in the question or data collection protocol); or, (c) social desirability bias (i.e., does the item or method of data collection minimize in some way the tendency to respond in a socially desirable manner, where person-to-person interviewing [face-to-face or phone] is considered the most likely method to invoke socially desirable responding). The second variable to be coded and extracted looks at data collection mode, providing a more specific and nuanced indicator of the methods that are linked with different levels of social desirability bias. These methods include the following: Face-to-Face, Phone, Self-Administered Paper-and-Pencil, Self-Administered Online, Time-Use[76] and Count-Based). Together, these indicators allow for an estimation of the effect that ambiguous wording, ambiguous time frames, social desirability and data collection mode have on the prevalence of attendance estimates found in the literature.

[76] While the time-use method of data collection is typically administered either over the phone or via paper-and-pencil (as a diary), this method is so distinct from other methods of data collection that a separate category seems appropriate.

Gender. The data to be extracted and coded for gender include the total sample size reporting their gender (N), and the number (n) or proportion of respondents (p) indicating each gender classification. In cases where only an n or p is reported for one gender classification, the common N (e.g., the number of respondents in the sample) was used to derive the n and p for the other binary gender classification (e.g., if there are 300 females and a total sample size of 500, then an assumption was made that there are 200 males). This approach assumes that the entire sample provided their gender classification, that the gender classifications were binary and that there were no missing data. While these assumptions are unlikely to be met, it is expected that there will only be a few missing and non-binary cases in each study; consequently, the statistical error associated with this approach is expected to be trivial.

Race and Ethnicity. The data to be extracted and coded for race and ethnicity include the total sample size reporting their race and ethnicity (N), and the number (n) or proportion of respondents (p) indicating each race and ethnicity classification. Because most researchers report only the most common categories, however, it is expected that most studies will offer data on the number or proportion of respondents who are White, African American or "Other," with a smaller, but still sizeable number of studies reporting the proportion of respondents who are Hispanic or Asian. Researchers who collected race and ethnicity data, but did not fully report the n-sizes or proportions were contacted with a request to provide the missing data.

Age. Age is a slightly more complex variable to code and extract than gender or race and ethnicity given that it can be reported in a number of different formats. Because age is a ratio-level variable, a mean and standard deviation are often used to describe the center and spread of a sample's age distribution. Age frequently manifests itself as a skewed distribution (e.g., academic studies focusing mostly on traditional-age students may also include older non-traditional students, producing a positively skewed distribution), however, so researchers also use the median (and range), as it is a more accurate representation of center than the mean under these circumstances. A third method of reporting age data is to provide frequencies for one or more age categories. For example, some researchers attempt to minimize social desirability bias by providing respondents with broad age categories that are not equally spaced (e.g., 18-24, 25-29, 30-39, 40-49, 50-64, 65 and over). Unfortunately, researchers who use this method may only report summary statistics for the modal category, thereby leaving out important information on the shape of the distribution. In such cases, the authors were contacted with a request to provide the full range of age data.

Familial Status. The coding of both marital and parental status will, perhaps, be the most difficult of the sociodemographics discussed thus far. For starters, these data are rarely reported unless familial status is a focus of the study. Thus, large amounts of missing data are expected. When marital and parental status are reported, the data are sometimes referred to only vaguely. For example, researchers may describe their sample (or a portion of it) as "cohabitating or married couples" without providing specific proportions for each group. Because cohabitation and marriage are often associated with very different religious outcomes, however, the omission of specific *n*-sizes or proportions for each group can have adverse consequences for the moderator analyses. Similarly, researchers may report that their sample consists of parents, but in most cases will omit data on the age of the children, especially when it is not a focus of the study. Here again, this omission can affect the analysis given that attendance estimates have been found to vary with the age of the child.

The coding of marital status can also be problematic given that researchers tend to provide respondents with different response options across studies. Even when response options are consistent, however, researchers may collapse their marital status categories differently. For example, it is not uncommon for researchers to intentionally or unintentionally combine those who are single, cohabitating, divorced or widowed into a single "not married" category, thereby providing a dichotomous outcome (married, not married) for analysis. Still others may compare those who are married with those who are divorced, and omit other categories from the analysis. Thus, not only are missing data likely to be a problem for familial status, but harmonizing the different response scales and reporting methods across studies can be a challenge. Accordingly, it was important to use an open and flexible approach to the data coding process. This involved recording all available marital and parental status data provided in the original studies so that the best and most consistent representation of these data could be used during the analysis.

Socioeconomic Status. The coding and extraction of socioeconomic status (SES) is also expected to be complex and problematic. Recall that there are four SES indicators commonly used to investigate the relationship with attendance: (a) composite measures of SES, (b) income, (c) education and (d) employment status. Composite measures of SES are typically composed of income, education and occupational status, but this is not always the case. Some studies use only two of these indicators, while others incorporate another indicator altogether (e.g., employment status or familial SES for college students). In addition, the ways in which the individual SES indicators are measured and combined vary from study-to-study. Thus, the composite SES data obtained across studies are likely to reflect a "hodge-podge" of SES information that could either obscure or facilitate a relationship with attendance. Therefore, composite measures of SES will not be included in this study.

With regard to the specific indicators of SES, virtually all studies measure income, education and employment status as categorical. Yet, the categories used and reported often differ from study-to-study, especially for income and education. In addition, researchers frequently report only the modal income and education categories when describing their samples. Thus, missing data on the entire set of income and education categories may be a problem. Here again, then, it is important to take an open and flexible approach to the coding process. As was the case with the other demographic variables, the open coding approach involves recording all available data provided in the study reports (i.e., *N*-sizes, mean and standard deviation, median and range, and *n-sizes* and proportions for all available categories) and then harmonizing those data in a way that best represents the study samples' SES characteristics. Where SES data are collected but not reported, the study authors will be contacted and asked to provide the missing data.

Methodological Moderators

Very little is known about how attendance estimates are influenced by the methods used to study them. For instance, it has been suggested here and elsewhere (e.g., see Hadaway et al., 1993, 1998; Tourangeau & Smith, 1996) that in-person interviews lead to higher attendance estimates than self-report methods because of socially desirable responding, but there have been no empirical investigations to specifically validate this claim (or to estimate the magnitude of the effect). It has also been assumed that probability samples lead to more accurate estimates of attendance than non-probability samples, but little is known about the magnitude of any difference that might exist between these two types of samples. By collecting methodological data from a large pool of studies, one can investigate these and other methodological questions, thereby shedding light on any effects that result from the selection of certain methods over others. In turn, this knowledge can help researchers avoid introducing methodological bias into their studies (Lipsey, 2009).

In determining which specific methodological factors to extract, code and empirically examine, Hunter and Schmidt (2004) suggest an over-inclusive approach. Cooper (2009, p. 28) refers to this approach as forming a "web of evidence." Therefore, a number of potential moderators that fall under five categories were coded and extracted: (1) sampling methodology; (2) sample characteristics; (3) participant motivation and participation rates; (4) data collection mode; and, (5) research design. Specific factors that were coded under each of these broad methodological categories will be discussed in detail, but it should be mentioned that only those with sufficient data were included in the meta-analysis.

Sampling Methodology. Although there are few characteristics that are universally agreed upon as defining a methodologically rigorous study, one of those characteristics is the use of probability (vs. non-probability or convenience) sampling methods. As Ilies et al. (2003, p. 612) stated, probability sampling methods give each member of a population a "nonzero, known probability" of being included in the sampling frame. Because of this, researchers are able to generalize their study findings to other people, places and times. Convenience samples, on the other hand, typically represent a small segment of the population. As a consequence, parameter estimates derived from convenience samples are likely to be biased in ways that reflect the idiosyncrasies of a local (e.g., a community) or limited (e.g., college students, volunteers) group of participants. Given this important distinction, it is important to code whether the study data came from probability or convenience samples, and to empirically determine whether these types of sampling methods yield different attendance estimates.

Sample Characteristics. Regardless of how the sample was selected, several sample characteristics are worth coding and exploring as potential moderators. Specifically, the region (East, Midwest, South, West; see Table 5) from which the data were collected should be coded and empirically tested given that important differences in religious outcomes have been consistently found between regions (e.g., see Gallup & Castelli, 1989; Pew Forum on Religion & Public Life, 2012). In addition, the religious affiliation of the sample participants should be documented as this variable is associated with differences in religious behavior (e.g., see Gallup & Castelli, 1989; Iannaccone, 1994; Lazerwitz, 1961). Specifically, the total sample reporting their religious affiliation (N), along with the number (*n*) and proportion (*p*) of respondents who selected a particular religious affiliation was extracted and coded. These religious affiliation data can then be recoded prior to analysis so that they reflect theoretically meaningful religious groups (e.g., see Smith, 1990; also Iannaccone, 1994) that are numerically well-represented.

In addition to the characteristics noted above, it is also important to record whether a sample represents a special population. For example, some samples may include only the religious or just high schoolers or college students. Other samples may be taken from those who are hospitalized, those who are seeking mental health treatment or those who are in the criminal justice system. Conversely, some samples may exclude certain segments of the population. For example, it may not be feasible to collect data from those in poor functional health; or, perhaps a researcher wants to focus on a single religious affiliation, thereby excluding all others. Importantly, these special populations may show very different religious behaviors than the general population at large. Thus, it was important to code and empirically determine whether these sampling criteria were able to influence estimates of attendance.

Participant Motivation and Participation Rates. A great deal of research has demonstrated that study participation rates vary as a function of participants' motivation for, and interest in the topic being studied, and that these sources of motivation may be related to study outcomes (e.g., see Dillman, 2000; and also Briggs et al., 2009; Woosley, 2005). For example, those who volunteer to participate in a study on religion may be more interested in religion than non-participants. Thus, attendance rates may be artificially inflated in studies with lower response rates. Conversely, this bias should be lessened in studies with high participation rates given that those who might otherwise have been non-respondents were somehow enticed to participate and share their behavior and attitudes toward religion. Thus, it is important to record both the methods used to entice individuals to participate and the resulting participation rates. This will allow for an exploration of the relationships, if any, between attendance and both participant motives and participation rates.

Data Collection Mode. As discussed in Chapter II, the mode of data collection is expected to influence estimates of attendance frequency. For example, the original Gallup Poll and GSS items were administered via face-to-face and phone interviews, respectively. Both of these administration modes are suspected of eliciting socially desirable responses (i.e., higher attendance estimates). Other attendance items, however, are administered via self-administered paper-and-pencil or web surveys, both of which allow the respondent to feel more anonymous and to be less inclined to acquiesce to social desirability forces. Still other researchers rely on observational methods (i.e., counts)—the least intrusive of all—to obtain their attendance estimates. Thus, although the mode of data collection is suspected of influencing attendance estimates, there is little evidence to confirm or refute this assumption. Accordingly, mode of data collection was extracted and coded from studies so that a more precise understanding of its influence can be ascertained.

Given the large amount of overlap in the attendance literature between data collection mode and measurement type, it may not be possible to disentangle the respective effects that each of these methodological characteristics exert. For example, recall that the original Gallup and GSS items were criticized for their ambiguous wording and lack of a specific time frame, as well as for eliciting socially desirable responses. The EPA time-use item, on the other hand, reduced the amount of pressure to respond in a socially desirable manner by removing all mention of religious services. Yet, this item also eliminated the use of ambiguous terminology and an ambiguous time frame. Thus, if there are any differences in the attendance estimates yielded by these two item types, it will be difficult, if not impossible, to determine which item element is responsible for the differences. The combination of other item types, however, may allow for the elimination of some of these confounds. For example, the item wording used in the Monitoring the Future Survey was identical to the original GSS item, but the former was self-administered via paper-and-pencil survey, thereby eliminating some of the "need" to provide a socially desirable response that may be present in the GSS phone interview. Thus, any differences between the attendance estimates yielded by these two items would, all else equal, be due to socially desirable responding. Taken together, then, data collection mode may be more useful for some comparisons than others.

Research Design. The last set of methodological features that were extracted and coded as potential moderators were related to research design. Using a classification framework presented by Trochim (2001) for understanding research design differences, the first characteristic to code is whether random assignment to groups was used. Those studies using random assignment were classified as experimental designs. Along with this classification, the number and nature of the experimental groups employed in the study will also be coded. The next question to ask of those studies that are not using random assignment is whether a comparison group was used. Studies using comparison groups (or control constructs; see Shadish et al., 2002) were coded as quasi-experimental studies, while those that did not were coded as non-experimental studies. It is expected that non-experimental studies will produce one attendance estimate for the overall sample, whereas experimental and quasi-experimental studies may produce an attendance estimate for each group. Thus, it was also important to code the socio-demographic characteristics of each group.

In addition to a study's experimental status, it is also important to note whether a cross-sectional or longitudinal design is used. That is, are the data collected at just one point in time, or are they collected at multiple times points? For studies that collect data at multiple time points, it was important to record exactly when those data were collected in case the timing of a particular data point should influence the associated attendance estimate. In addition, any available data on the attrition rates from one observation to another should be recorded. These data, if reported in enough studies, would allow for a determination of whether a general association exists between attrition and the attendance rates obtained at follow-up intervals.

Contextual Moderators

Besides coding information on potential theoretical and methodological moderators, it is also important to code information on the contextual factors associated with a study. For example, it may be useful to code information on the researcher (e.g., gender, field of study or academic discipline, institutional affiliation type, religious affiliation of institution), the circumstances under which the study was carried out (e.g., funding source for the study; the year in which the study data were collected or the year in which the study was published, completed or made available) and the outlet in which the study data were reported (e.g., dissertation, book, academic journal, internal research report, newsletter, unpublished manuscript, etc.). As Lipsey (2009) stated, these factors are unlikely to directly influence the study effects data, but may be indirectly related to factors that are capable of influencing study results. For instance, while a researcher's field of study is unlikely to directly impact a study's findings, it may shape their choice of methodology, which, in turn, could influence the study outcomes. Additionally, the outlet used for reporting is not likely to directly influence study results, but it has long been assumed that published studies yield different results than unpublished studies (Hunter & Schmidt, 2004). Thus, it is important to code and examine each of these contextual factors.

In addition, it is important to code information that will allow the synthesist to quickly relocate information from the original study reports, and to identify and track specific studies during the analysis. Accordingly, the page numbers where the original study data are pulled should be recorded, and a unique identification (ID) number should be given to each unique study. The study ID number is important because some study results are reported in more than one location (e.g., in a dissertation, conference presentation and publication), and identifying studies with a unique code allows the synthesist to track these occurrences. In turn, the ability to track multiple reports of the same study allows for the reduction of missing data (data missing from one report may be available in another), while allowing the synthesist to avoid violating the assumption of statistical independence when using data from sub-groups of the full sample (e.g., if attendance estimates are provided for the full sample, and for men and women separately, then the inclusion of all three estimates in a single analysis would violate the assumption of statistical independence). A modification of the original study ID will allow for the coding of sub-sample data (e.g., sub-sample data for study 1 can be coded as 1a, 1b, etc.). Together, these contextual factors, along with the theoretical and methodological moderators, allow for the formation of a web of evidence from which a more complete picture can be formed about the factors that influence estimates of attendance in America.

Assessing and Improving Coding Quality

The general procedure for extracting and coding study data is described below, along with the procedures for assessing and improving the quality of the coding and extraction process.

Development and Refinement of the Coding Protocol

The development and use of a coding protocol is important given the sheer number of data elements to be extracted and coded from each study. Wilson (2009) noted that a coding protocol includes both a coding form and a coding manual. Coding forms provide a means for recording the study data, while the coding manual provides specific instructions and guidelines on how to code the data. The manual covers all possible scenarios and can be particularly helpful to the coder when coding decisions are ambiguous. To use Hunter and Schmidt's (2004) analogy, the coding forms are like surveys, the data elements to be coded represent the questions and the coding manual defines the response options.

Given that it is impossible to anticipate all possible data coding scenarios, it is wise to begin by drafting a coding protocol, and then revise it iteratively by applying the protocol to a sample of studies. The protocol should begin with the researcher's thoughts on how to code the study effects and moderator data under a variety of circumstances. Once the initial protocol has been developed, the synthesist can then use a sample of studies to assess how well the protocol captures the data variations that present themselves. Based on this assessment, the synthesist should revise the protocol by refining, adding and even eliminating items. In order to achieve the highest quality data, this iterative process should carry on throughout the entire data collection period.

Strategies for Reducing Coding Error

Horowitz and Yu (1984) reported that the most common coding errors occur because the coder simply misses the data in the original study. Another common error, according to Orwin and Vevea (2009) involves the misapplication of codes. Data are located and extracted correctly, but the coders incorrectly apply the coding scheme. The use of low inference codes (i.e., those that simply require the transcription of study data to the coding form) helps reduce the frequency with which this latter type of error is made, but even the most seemingly straightforward variables can sometimes be difficult to code correctly (Wilson, 2009). Coder bias may also introduce error. For example, a coder who expects (or hopes) to find a certain outcome may pay more attention to the information that is consistent with their expectation than to information that contradicts their schema. Other mistakes are likely to result from coder fatigue given that meta-analyses require sifting through a large number of studies (Wilson, 2009).

Some suggestions have been given to help reduce the frequency with which errors are made during the coding process. First, writing the coding items so that they are closed-ended and as specific as possible helps reduce the amount of inference needed on the part of the coder. For example, instead of writing a coding item that asks, "What are the age characteristics of the sample?" one could ask a series of questions, such as, "Which of the following measures of central tendency were reported: mean, median, mode or frequencies by age category?" This can then be followed by specific questions regarding the value of the central tendency and variance indicators (as well age categorical data). Writing items in this way not only helps reduce ambiguity for the coder, but also minimizes opportunities for errors of omission and coder bias.

Coder training and practice can also help reduce error. Training should involve going over the protocol in detail, and providing feedback as each coder covers a sample of studies. This process not only allows coders to understand which data elements the synthesist is looking for, but also to understand why they are important, thereby creating an opportunity for the coders to "buy-in" and become invested in the process. The process also allows the coders to identity initial problems and questions, and for the synthesist to identify ways in which the protocol can be improved (perhaps creating more specific, less ambiguous items).

Confidence ratings can also be used for items that are identified as potentially ambiguous or difficult to code during pretesting. Confidence ratings can be configured several different ways (e.g., see Janda, 1970; Orwin & Corday, 1985; both cited in Orwin & Vevea, 2009), but generally are used as a numeric indicator of the confidence that the coder has in the accuracy of their codes for a particular item or variable. Several different rating scales have been used, but Orwin and Corday (1985; cited in Orwin & Vevea, 2009) found that a three-point scale worked best. These ratings can then be used to assess the adequacy of the coding for a particular variable. If the confidence ratings tend to be low, then perhaps the variable should be dropped, or at least included as a covariate if results are found to differ according to the confidence ratings.

Last, it is wise to use two or more raters for at least a sample of studies. The use of multiple raters allows for an assessment and correction of numerous problems, including the identification of coder bias and systematic coding misapplications. Importantly, the use of multiple raters also helps identify problems and ambiguities that might exist with the coding protocol. Thus, by using multiple raters for a sample of studies, numerous problems can be identified and corrected. Furthermore, one can also provide a quantitative measure of the degree to which the raters are seeing, interpreting and recording the same information from the studies.

Assessment of Coding Quality

In most cases, coding quality can be assessed by calculating a measure of rating consistency or agreement. These measures of rater consistency and agreement include statistical indicators of both *intra*rater and *inter*rater reliability. Intrarater reliability can be assessed by having a single coder rate the same documents on two or more occasions, and comparing the similarity between the pair of ratings. This can be particularly useful for assessing coder drift (Wilson, 2009). Specifically, the ratings given to a sample of studies that are coded early on in the process can be compared to the ratings given to the same set of studies after a number of additional studies have been coded. This allows the coder to determine whether (and how much) coder drift has occurred, and to correct for it before completing the coding process, thereby potentially saving time and resources.

Interrater reliability, on the other hand, can be assessed by comparing the coding discrepancies between multiple raters. This form of interrater reliability is useful for diagnosing a number of problems, including coder error and bias. In addition, it can be useful for identifying variables that are difficult to code, and that should either be modified in the coding protocol or dropped from the analysis.

The statistical formulas to be used for assessing both types of reliability depend on several factors. One important factor is the level of measurement used in the coding scheme. For categorical variables, there are two formulas. The first is proportion agreement, which is simply the number of agreements divided by the total number of agreement chances. The proportion agreement statistic is easy to calculate, and according to Jones, Johnson, Butler and Main (1983), is useful when there is little variability among the ratings (e.g., when one outcome is predominantly recorded over others). Yet, the percent agreement statistic fails to control for chance agreement. For example, the expected rate of agreement between two raters for a dichotomous variable is 50%, but due to random error, the level of agreement will be greater than 50% half of the time simply by chance. Thus, chance agreement rates that look favorable can mask rater uncertainty and disagreement. For this reason, it is recommended that the minimum rate of acceptable agreement should be 90% or greater. Yet, even this criterion may not be adequate when one of the two outcomes naturally occurs more than 50% of the time. For example, suppose that 80% of attendance studies use convenience (as opposed to probability) samples. Because the ratio of convenience to probability samples is greater than one-to-one, the likelihood of coding the former sampling type over the latter surpasses 50%, creating a situation where 90% agreement may still be reached via random error. Such a scenario would lead the synthesist to falsely conclude that the raters were coding similarly, when, in fact, they are adding error variance to the dataset.

In order to control for chance agreement, Cohen (1960) introduced the Kappa (*k*) statistic. Cohen's Kappa takes on a value of "0" when agreement is equal to chance; a value of "1" if agreement is perfect; and, a negative value if agreement is less than chance, thereby also providing a measure of disagreement. Importantly, sampling distributions have been generated for Cohen's Kappa that allow for the use of significance tests. Thus, it is possible to quickly determine whether the observed rate of agreement is significantly greater, less than or no different than chance agreement. One weakness of Cohen's Kappa, however, is that it loses statistical power when there is little variability among the ratings (Jones et al., 1983). Percent agreement does not have this limitation. Therefore, Cohen's Kappa should be used when there is ample variability among the ratings, but under conditions of homogeneity, percent agreement should be considered as an alternative reliability indicator.

For variables measured on an ordinal, interval or ratio scale, two additional options for estimating interrater reliability should be considered. Pearson's product-moment correlation coefficient (*r*) is one option, and measures the degree of association between two sets of ratings. As with Cohen's Kappa, Pearson's *r* can reflect disagreement (with negative values), no agreement (with values close to zero) and agreement (with positive values), and includes the use of a significance test to determine whether the association between two sets of ratings is greater than what can be expected by chance. Pearson's *r* can also be averaged across pairs of raters if there are more than two (Orwin & Vevea, 2009). One limitation associated with Pearson's *r*, however, is that it only measures the extent to which two sets of ratings covary, but does not assess the extent to which the two sets of ratings agree in terms of their magnitude. For example, Pearson's *r* would be nearly perfect ($r = 0.99$) for the following set of ratings even though the raters are clearly disagreeing about magnitude:

Rater 1: 5, 7, 9, 5, 10

Rater 2: 50, 55, 65, 50, 70

Thus, Pearson's *r* is useful for determining whether a pair of ratings covary at greater than chance levels, but it fails to take discrepancies of magnitude into account.

To overcome this limitation, researchers have turned to the use of the Intraclass Correlation Coefficient (ICC or r_i). The ICC not only accounts for the degree of association (or covariability) between two sets of ratings, but also takes into account discrepancies of magnitude, thereby providing a measure of association and agreement between pairs of ratings. Like Pearson's *r*, the ICC can take on values ranging from -1.0 (perfect disagreement) to 0.0 (chance agreement) to 1.0 (perfect agreement), and is accompanied by a significance test to determine departures from chance agreement. In addition, the ICC can be generalized to situations where ratings are obtained from more than two raters. Like Cohen's Kappa, however, the ICC requires a fair amount of variability among the ratings in order to have enough statistical power to detect significant levels of agreement when they exist. As long as there is an adequate amount of heterogeneity, then, the ICC is the preferred index of interrater reliability for ordinal, interval and ratio data.

Thus, Orwin and Vevea (2009) recommend using Cohen's Kappa for categorical variables and the ICC for continuous variables in cases where heterogeneity is sufficient. When homogeneity exists between ratings, however, Percent Agreement and Pearson's *r* are the preferred measures. In order to decrease the likelihood of homogenous ratings, Lipsey and Wilson (2001) recommend using large samples of 50 or more studies to assess interrater reliability. Accordingly, a single rater can code and extract data until 50 studies containing attendance and moderator data have been identified. At that point, an additional rater can be asked to extract and code data from the same 50 studies. If agreement between the two raters is not at an acceptable level for any of the coded variables, the two raters can review and discuss their differences until the rate of agreement is acceptable. In addition, the primary coder can recode the first 50 studies after a majority of the coding has been completed for all studies. If intrarater reliability is sufficiently high, then the coding can continue until the process has been completed. If intrarater reliability is too low, however, then the discrepancies between the codes should be investigated, and changes to the protocol considered. In this way, coder drift can be identified and minimized.

Meta-Analytic Database Preparation

Once the study effects and moderator data have been extracted, the synthesist must then prepare the database for analysis. This includes determining the database structure,[77] deciding how to deal with missing data, cleaning the data and harmonizing data from different scales so that meaningful analyses can be undertaken.

Database Structure

Meta-analytic databases typically take one of two forms. The first is a flat database structure in which one row of data represents one study (this is, perhaps, the most familiar database structure as it is the most obvious choice for novice data analysts). If there are multiple study effects within a single study (either from longitudinal data collection or from multiple groups), they must first be cumulated so that each study provides only one set of data for the analysis. This is perhaps the most straightforward method of setting up a meta-analytic database as it helps the analyst avoid violating the assumption of statistical independence and it is relatively easy to work with and comprehend. It does suffer from a loss of information, however, given that when multiple study effects exist within studies, they typically contain unique information pertaining to a point in time or particular segment of the population (e.g., mental health patients) that is lost when the effects are cumulated.

The second meta-analytic database structure is referred to as hierarchical. A hierarchical database can either take the form of a relational database or a flat file where each row represents a unique *study effect* (as opposed to a unique study in the traditional flat file). *Comprehensive Meta-Analysis* (CMA), the software developed by Borenstein et al. (2005) and used in this study, uses the flat file approach, but prior to analysis this approach is inefficient. For example, if there

[77] The database structure actually needs to be determined prior to data extraction (i.e., so that data can be directly entered into the database during the extraction and coding process), but it is discussed here along with the other database preparation activities.

are two effects in a given study, two rows of data must be designated for the respective study effects. Yet, the associated methodological and contextual moderators are likely to be the same for both study effects given that they come from the same study. Thus, the moderator data must be entered multiple times. A more efficient approach is to set-up a hierarchical relational database. Here, data are entered into two databases. The first is used to contain the study effects data, along with any characteristics that are uniquely associated with each study effect (e.g., participant demographics). The second database contains study-level characteristics that apply to all of the study effects generated by a single study (e.g., methodological and contextual variables). This data entry approach eliminates the redundancy associated with entering the latter information more than once. When preparing to analyze, the two datasets can be joined via the use of a common identification number that links the study effects data with the study-level characteristics. In this way, the hierarchical database structure provides a more efficient means of entering the meta-analytic data. Accordingly, the hierarchical database structure is the one that was used in this study.

Missing Study Effects and Moderator Data

Missing data are a common problem for synthesists, affecting approximately 25% of studies in psychology (Bushman & Wang, 2009). This problem is particularly pernicious when the data are systematically missing. As discussed previously, the significance of a finding in a primary study may determine whether it is reported, which could, in turn, bias the results of a meta-analysis against the null hypothesis. In addition, missing data can adversely affect a synthesist's ability to achieve a level of statistical power needed to identify moderators. Yet, the former problem is not expected to be an issue here given that the purpose of this study is not to determine *whether* the prevalence of attendance in America is different from zero, but by *how* *much*. Furthermore, the amount of attendance data expected from the literature, and the way in which attendance data are typically reported should mitigate against the latter problem. Specifically, most studies reporting attendance either focus on the prevalence of this outcome, or they report it as a demographic characteristic of their study sample. Studies that compare attendance rates across groups are also likely to provide the data needed (N-sizes and *n* or *p*) given that it is customary to provide descriptive data on the groups involved in multi-group comparisons. Only in studies that correlate attendance with other outcomes is the problem of missing data expected to be a potential issue given that researchers do not uniformly provide detailed descriptive data on the variables included in a correlational analysis. Yet, the vastness of the literature on attendance should ensure that an adequate number of study effects can be extracted and included in the meta-analysis.

Missing data may be more of a problem for the moderator variables, however. For example, it is unlikely that all (or even many) studies with an attendance estimate will report data for the familial or socioeconomic status variables. This is largely due to the fact that data on these variables are collected less frequently than for other demographics, such as gender, race and ethnicity and age. Thus, a determination will have to be made about whether there is enough data available for these variables to warrant inclusion in the analysis. Given that the number of attendance estimates in the literature is expected to be large, however, there may be enough data to at least provide some information on how the composition of these variables in study samples influences the observed attendance estimates. And, although, it is not expected that moderator data will be missing systematically, a safeguard can be used to ensure that this is the case. Specifically, Lipsey and Wilson (2001) and Hunter and Schmidt (2004) recommend empirically examining whether effect sizes from studies reporting moderator data are different than those that do not. For example, if there is a sizeable portion of missing data for the sampling variable (indicating whether a convenience or random sample was drawn), then the synthesist should determine whether the attendance estimates differ between studies that describe their sampling procedure and those that do not. If a difference is found, then a variable that incorporates the reporting information can be created and included as a covariate during the analyses.

Additionally, there are other steps that can be taken to minimize potential missing data problems. First, it should be recognized that it is not uncommon for study results to be presented in multiple documents or sources. For example, a doctoral student may not only publish their doctoral dissertation through their university, but may also choose to present their findings at a conference, and, later, to publish them in a journal. The synthesist can use these sources to fill in the data gaps. If this approach fails (or if multiple sources are unavailable), then the original authors should be contacted with a request for the missing data. This latter approach can be difficult, however, given that it is not only time-consuming, but it is likely that, in numerous instances, the original researchers' contact information has either not been provided or has changed since the study data were made available. Thus, it is likely that missing data will remain a problem after the data collection phase of this study. Still, these approaches should be used to minimize the number of gaps in the meta-analytic database.

Cumulation of Findings

Although the use of proportions as meta-analytic study effects is rare, Borenstein and colleagues (2005, 2009, 2015) have developed procedures and software for cumulating these outcomes. Thus, the methodology used in this study will rely heavily on the procedures outlined by Borenstein et al., which are discussed below.

Calculation of Study Proportions

The first step in cumulating findings involves calculating a proportion of regular attenders and “any given Sunday” attenders for each study in the meta-analytic sample. Most studies will have already calculated attendance rates, but for those that have not, proportions must be calculated from the available study data. This involves dividing the number of attendees by the total sample size for a particular study as indicated by the following equation:

$$p_i = \frac{n_i}{N_i}$$

where N_i is the total number of participants in study *i* available to report attendance (i.e., the sample size), and n_i and p_i represent the number of attendees and the proportion of the sample reporting attendance in study *i*, respectively. Once the attendance estimates for each study have been calculated and recorded in the meta-analytic database, the synthesist must then turn their attention to the selection of the most appropriate mathematical model. That is, the synthesist must choose to employ either a Fixed-Effect Model or a Random-Effects Model. Both are described in more detail below.

Fixed-Effect Models

The fixed-effect model assumes that there is one true effect size in the population; that all sampled studies are functionally identical in terms of the people, places and times represented; and, that all sample statistics are estimators of the one true population parameter (Borenstein et al., 2009). Thus, the study effect is assumed to be the same (or fixed) across all studies, and any observed variation between studies is the result of sampling individuals into studies (i.e., random sampling error). Given these assumptions, the mathematical model can be expressed as:

$$p_i = \theta + \Sigma_i \,,$$

where θ is the parameter representing the proportion of attendees in the population and Σ_i represents the random error associated with the attendance estimate in study *i*. Importantly, because Σ_i is expected to be random, the sampling distribution of Σ_i can be estimated, partitioned and removed from the meta-analytic model, which allows for an unbiased estimate of θ.

Estimating the Fixed Effect. One problem in the cumulation of findings is that studies contain different amounts of information about the population parameter, with large sample size studies containing more information than small sample size studies. If study proportions are simply averaged together, this characteristic would be lost, and all studies, regardless of their sample size, would contribute equally to the grand mean proportion. To correct this, a weighting scheme must be employed so that each study's contribution to the grand mean is proportional to the amount of information contributed by each study. To accomplish this, Borenstein et al. (2009) recommend the inverse-variance method of weighting. This approach involves multiplying the parameter estimate for study *i* by the inverse of its variance, where the inverse-variance can be calculated as:

$$W_i = \frac{1}{V_{P_i}}.$$

Here, W_i is the weight for study *i*, and V_{P_i} is the variance for the proportion in study *i*, which is estimated as:

$$V_{P_i} = \frac{P_i * Q_i}{N_i - 1},$$

where

$$Q_i = 1 - P_i.$$

Put differently, Q_i represents the proportion of *non*-attendees in study *i*.

Once the weights for each study have been calculated, the weighted mean proportion can be estimated by summing the product of the weights and proportions from each study and dividing by the sum of the weights. This can be expressed as:

$$\overline{P} = \frac{\Sigma(W_i * Pi)}{\Sigma W_i},$$

where $\overline{P}$ is the weighted mean fixed-effect proportion.

Estimating the Fixed-Effect Variance and Standard Error. Even though $\overline{P}$ represents the weighted grand mean proportion, it is still just an estimate of the population parameter. Thus, confidence intervals should be constructed around the weighted grand mean to assess the range within which the true population parameter is likely to be found. To do this, one must first calculate the variance of the observed study proportions by taking the inverse of the sum of the study weights:

$$V_{\overline{P}} = \frac{1}{\Sigma W_i},$$

where $V_{\overline{P}}$ is the variance of the observed proportions.

The standard error of the weighted grand mean proportion can then be estimated as the square root of the variance:

$$SE_{\overline{P}} = \sqrt{V_{\overline{P}}}\ .$$

Constructing Confidence Intervals for the Fixed-Effect. Once the standard error of the weighted grand mean proportion has been estimated, lower and upper confidence limits can be constructed. Specifically, the standard error is multiplied by the respective *Z*-critical value corresponding to the 95% confidence limits in the normal distribution.[78] This product is then subtracted from, and added to the weighted grand mean proportion to identify the respective lower and upper 95% confidence limits. This can be formally expressed as:

$$95\%\ \text{CI} = \bar{P} \pm (1.96 * SE_{\bar{P}}).$$

Borenstein et al. (2009) also note that a *Z*-test[79] can be used to determine if the weighted grand mean proportion is significantly different from zero. For this study, however, the purpose of calculating the weighted grand mean proportion is not to determine *if* it is different from zero, but to determine *how much* different than zero it is, and whether this value changes in the presence of moderators. Thus, *Z*-tests will not be conducted on the weighted grand mean proportion.

Influence of Sample Size on the Fixed-Effect Confidence Limits and Statistical Power. Borenstein et al. (2009) point out that the standard error of the weighted grand mean proportion can be alternately expressed as:

$$SE_{\bar{P}} = \sqrt{\frac{\sigma^2}{k * n}},$$

where σ^2 represents the population variance, *k* represents the number of studies and *n* represents the number of participants represented across all studies in the meta-analysis. Looking at the denominator, it becomes evident that the $SE_{\bar{P}}$ will approach zero as either the number of participants (*n*) *or* studies (*k*) increase. It is common for meta-analyses to be characterized by a large *n*; thus, even if *k* is relatively small, most fixed-effect meta-analyses will yield relatively

[78] The normal Z-distribution can be used here given that the only source of error is expected to be random (Borenstein et al., 2009).

[79] $Z = \dfrac{\bar{P}}{SE_{\bar{P}}}$

small confidence intervals, which, in turn, increases statistical power and decreases the likelihood of Type-II errors. A large *n* and large *k*, however, will accomplish this much more efficiently.

Random-Effects Models

The random-effects model assumes that: (a) there are many populations (each with their own true effect size) represented in a meta-analytic sample of studies; (b) each sampled study is unique in terms of the people, places and times represented; and, (c) each study outcome is an estimator of its own true population parameter (Borenstein et al., 2009). Thus, each study effect is assumed to be different (or random) across all studies, and the observed variation between studies is the result of both sampling individuals into studies and sampling studies into the meta-analysis. Given these assumptions, the mathematical model can be expressed as:

$$P_i = \mu + \xi_i + \Sigma_i,$$

where μ represents the weighted grand mean proportion, ξ_i represents the deviation between the grand mean (μ) and the true effect size for study *i* (θ_i) and Σ_i represents the deviation between the true effect size for study *i* (θ_i) and the observed effect size in study *i* (Y_i). Substituting for the latter terms, the equation above can be re-expressed as:

$$P_i = \mu + (\mu - \theta_i) + (\theta_i - Y_i).$$

The latter two deviation terms represent two sources of random error. The second deviation term (i.e., $\theta_i - Y_i$) is from the fixed-effect model and represents the error that results from sampling people into studies. The first deviation term ($\mu - \theta_i$) is unique to the random-effects model and represents the error that results from sampling studies into a meta-analysis. The expected value for both deviation terms is zero because they are expected to be randomly distributed. Therefore, the null hypothesis for the random-effects model can be expressed as:

$$H_0 : \mu = 0 .$$

Simply stated, the null hypothesis for a random-effects model states that the grand mean of the population of true effects is zero.

Estimating the Random Effects. To test the plausibility of the null hypothesis, we must first estimate the grand mean of the effects. As with the fixed-effect model, we must first weight each study by the inverse of its variance. Because the random-effects model assumes two sources of error variance, however, we must now incorporate both the within study variance and the between study variance into the weighting scheme. There are several methods for estimating the between-studies variance, but Borenstein et al. (2009) recommend the Method of Moments (or the DerSimonian & Laird Method). The Method of Moments can be calculated as follows:

$$T^2 = \frac{Q - df}{C} ,$$

where T^2 is the estimator of the true variance between population means (τ^2); Q is a ratio of true variance to error variance (defined below); and, C is calculated as:

$$C = \Sigma w_i - \frac{\Sigma w_i^{\,2}}{\Sigma w_i} .$$

The individual study weights can then be calculated as:

$$w_i^{\,*} - \frac{1}{V_{y_i}^{\;*}} ,$$

where the * denotes a random-effects model, and

$$V_{y_i}^{\;*} = V_{y_i} + T^2 .$$

In words, the weight assigned to each study is now the inverse of the within-study and between-study variance.

The weighted grand mean of the randomly sampled effects can then be estimated as follows:

$$ES_M^{\ *} = \frac{\Sigma Y_i w_i^{\ *}}{\Sigma w_i^{\ *}},$$

where $ES_M^{\ *}$ is the weighted grand mean average of the population means.

Estimating the Random-Effects Variance and Standard Error. The variance of the observed effects around the grand mean can then be estimated as:

$$V_M^{\ *} = \frac{1}{\Sigma w_i^{\ *}}$$

with the standard deviation of the population effects (i.e., the standard error) estimated as:

$$SE_M^{\ *} = \sqrt{V_M^{\ *}}.$$

Constructing Confidence Intervals for the Random-Effects Weighted Mean. Once the weighted grand mean and standard error have been estimated, confidence intervals can be constructed around the grand mean. Specifically, 95% confidence intervals can be constructed as follows:

$$95\%\ \text{CI} = ES_M^{\ *} \pm (1.96 * SE_M^{\ *})$$ [80]

Implications of the Random-Effects Weighting and Variance Calculations. Here, it is important to highlight the practical implications of the weighting scheme used in random-effects models. Recall that for fixed-effect model, the goal is to estimate a mean population effect that is common to (or fixed across) all studies. Because there is only one mean to estimate in the

[80] If the synthesist is interested in knowing if the grand mean effect size is different than zero, a *Z*-test can also be performed, where $Z = \frac{ES_M^{\ *}}{SE_M^{\ *}}$.

population, every study in a fixed-effect meta-analysis is an estimator of that mean. Yet, some studies (i.e., those with larger sample sizes) provide more information about the population mean than other studies (i.e., those with smaller sample sizes). Accordingly, the weighting scheme reflects this disparity by assigning relatively large weights to studies with more information, and relatively small weights to studies with less information.

Conversely, the goal in random-effects models is to estimate multiple population means. Under this assumption, each study in a random-effects meta-analysis is an estimator of a unique population mean. Therefore, each study, regardless of its sample size, contains unique and important information. Accordingly, the weighting scheme in random-effects models incorporates a constant that helps curb the effect of sample size. Specifically, the between-studies variance term (i.e., the constant) is included in the calculation of study weights, which has the effect of pulling weights that were extremely large or small under the fixed-effect model toward the middle under the random-effects model. Thus, studies with small sample sizes make more of a contribution, and studies with large sample sizes make less of a contribution in a random-effects model than in a fixed-effect model (see Borenstein et al., 2009, p. 78).

Another implication of the random-effects weighting scheme is that the addition of the between-studies variance to the within-studies variance means that the total variance around the grand mean will be larger. Because of this larger variance, the standard error of the mean will also be larger. In turn, the confidence intervals around the grand mean will be wider (or less precise), thereby creating a more conservative test of the null hypothesis. This conservative effect has been touted as one of the advantages of using random-effects models. Specifically, while fixed-effect models have more statistical power, the assumptions underlying this model are rarely met. Accordingly, significance tests of the mean fixed-effect may lead to Type-I errors.

Conversely, random-effects models not only have better underlying assumptions in most cases, but the more conservative significance test also prevents against Type-I errors. In addition, if the between studies variance is close to zero, then this variance component will essentially drop out of the analysis, thereby reducing the mathematical model to a fixed-effect model. Thus, when meta-analysts lack a guiding theory about the study data, a random-effects model should be selected given its conservative nature.

A third implication of the random-effects weighting scheme relates to statistical power. To demonstrate, recall that the standard error of the mean in a fixed-effect model is represented as:

$$SE_M = \sqrt{\frac{\sigma^2}{k * n}}.$$

Given even a minimal number of studies (*k*), the addition of participants (*n*) into studies will decrease the standard error until it approximates zero, thereby shrinking the confidence intervals and increasing statistical precision and power until it is close to unity (1.0).

On the other hand, the random-effects model requires both large within-study sample sizes and a large number of studies to achieve the same level of statistical power. Specifically, the standard error of the mean for the random-effects model is represented as:

$$SE_M{}^* = \sqrt{\frac{\sigma^2}{k * n} + \frac{\tau^2}{k}}.$$

Here, we have the same within-study variance term as in the fixed-effect model (i.e., $\frac{\sigma^2}{k * n}$), where, again, with a minimal number of studies and a large number of participants, the standard error of the mean would approach zero. But, the addition of the second variance term in the random-effects model (i.e., $\frac{\tau^2}{k}$) means that no matter how large the combined sample size becomes, the standard error of the mean will still be restricted by the size of k. In order for statistical power to approach unity, then, both *n and k* must be large. This is often a problem in meta-analyses where samples of 50 studies might be considered relatively large. Thus, statistical power in meta-analyses is often a problem even though a large number of participants are usually represented.

Selecting Fixed-Effect or Random-Effects Models

The Common Practice. Hunter and Schmidt (2004) and Borenstein et al. (2009) note that meta-analyses in the published literature often base their model-selection decision on the significance of Cochran's Q statistic. Recall that Cochran's Q is a chi-square statistic that evaluates the ratio of true variance (between studies) to error variance (within studies). As the common practice goes, if Q is significant, then the meta-analyst is compelled to conclude that real variation exists between studies (i.e., multiple population means are represented in the study sample), and so a random-effects model is selected. Conversely, if the Q-statistic is non-significant, then the meta-analyst concludes that the observed variance is nothing more than random variation distributed around a single population mean. In such cases, a fixed-effect model is selected.

Problematic Criteria. The problem with this approach is that the statistical power of the Q-test relies heavily on the number of studies (k) sampled. Because most meta-analyses are not characterized by large k, the Q-statistic often suffers from low statistical power (Hunter & Schmidt, 2004; Borenstein et al., 2009). As a result, synthesists using this approach are likely to incorrectly select the fixed-effect model. Because the fixed-effect model assumes that all of the observed variance is due to the sampling of individuals into studies, any variation that exists between studies is omitted. If real variation does exist between studies, the fixed-effect approach will yield variance estimates that are too small and confidence intervals that are too narrow. Consequently, the probability of committing a Type-I error is increased.

The Proper Approach: Using Theory to Guide Model Selection. Rather than use an empirical approach that is often hampered by low statistical power, Borenstein et al. (2009) argue for a theoretical approach. Here, the meta-analyst uses knowledge about the nature of their data along with the underlying assumptions of fixed- and random-effects models to select the appropriate statistical framework. Specifically, if the studies in a meta-analysis are functionally identical or if the goal is to simply describe or summarize the sampled studies (as opposed to generalizing beyond the sample of studies), then a fixed-effect model is appropriate. For example,

> Suppose that a pharmaceutical company . . . use[s] a thousand patients to compare a drug versus placebo. Because the staff can work with only 100 patients at a time, the company will run a series of ten trials with 100 patients in each. The studies are identical in the sense that any variables [that] can have an impact on the outcome are the same across the ten studies. . . . [In addition], the goal of the analysis is to see if the drug works in the population from which the patients were drawn (and not to extrapolate to other populations) (Borenstein et al., 2009, p. 83).

In this example, both conditions for a fixed-effect model have been met. That is, it is reasonable to assume that the ten samples will each estimate a single, common effect size; and, the researchers are primarily interested in describing the results for these ten samples. This is not the typical case, however, and so the theoretical criteria for selecting a fixed-effect model are rarely met.

Random-effects models, on the other hand, are appropriate when the goal is to generalize beyond the sample of studies, and when these studies are *not* functionally identical. As an example of the latter criteria, Borenstein et al. (2009, p. 69) state that

> Effect sizes might be higher (or lower) when the participants are older, or more educated, or healthier than others, or when a more intense variant of an intervention is used, and so on. Because studies will differ in the mixes of participants [and other factors] . . . there may be different effect sizes underlying different studies.

In such cases, the sampled studies are no longer functionally identical. Because this is the typical case, and because the goal of most meta-analyses is to generalize beyond the sample of studies, random-effects models should be the predominant choice when selecting a meta-analytic statistical model.

Importantly, random-effects models are also the more conservative choice. First, by including a between-group variance component, random-effects models yield wider confidence intervals than fixed-effect models, thereby providing protection against Type-I errors. Second, if there is, in fact, no between studies variance, then this variance component drops out of the equation, which reduces the model to a fixed-effect analysis. Thus, even when expected functional differences between studies do not manifest themselves, it still makes sense to begin with the random-effects model because: (a) it fits the theoretical expectation underlying the analysis; and, (b) it reduces to a fixed-effect model when systematic variation between studies does not exist. Ultimately, however, the synthesist must select the model that best represents the theoretical expectations underlying the analysis.

Selecting a Model for the Present Study. For several reasons, this study will employ random-effects models to cumulate the proportion of study participants reporting regular attendance. First, the very nature of the methodological and socio-demographic moderators to be investigated in this study suggests that the sampled studies will not be functionally identical. Second, the goal of this meta-analysis is to not only describe or summarize the attendance data that has been reported to date, but to generalize beyond this study so that future researchers are able to derive an expected attendance estimate for their particular measurement method and population. Third, some of the moderator relationships to be tested are exploratory in nature (e.g., minority ethnic group comparisons), and a more conservative significant test is appropriate for these analyses. Finally, in the unlikely situation that systematic between-study differences do not exist, the random-effects models will reduce to fixed-effect models. Thus, the random-effects statistical model is clearly the appropriate choice for this study.

Assessing Heterogeneity

Even though measures of dispersion and heterogeneity[81] should not be used as the principle means for determining whether a fixed- or random-effects model should be selected, they can still yield valuable information about the distribution of effects across studies. Specifically, measures of dispersion and heterogeneity allow for an assessment of: (a) *whether* a portion of the observed variability across studies is real; and, if so (b) the *proportion* of observed variability that is real; and, (c) the *amount* of real variability that exists between studies. To assess each of these areas of heterogeneity, the following five indicators have been developed: Q, p, τ τ^2 and I^2, and are discussed in more detail below.

Q and p

As stated previously, Cochran's Q (1954) is used to help determine whether the observed variability across studies reflects real variability or nothing more than random sampling error. The Q statistic accomplishes this by evaluating the ratio of true variance to error variance (i.e., the ratio of the observed variance to the amount of variance expected by chance) against a chi-square distribution with k-1 degrees of freedom. Specifically, Q can be calculated from the following equation:

[81] Following Borenstein et al. (2009), the terms dispersion and variation are used to refer to the observed variance across studies, which includes both true variation and random error. Heterogeneity, on the other hand, was used to refer to the variation among the true effect sizes.

$$Q = \Sigma w_i Y_i^2 - \frac{(\Sigma w_i Y_i)^2}{\Sigma w_i}$$

Borenstein et al. (2009) point out, however, that because Q is a sum, the value of Q depends heavily on the number of studies (k). With too many studies, Q will almost always be significant even if no real variation exists. On the other hand, if there are too few studies, Q will fail to attain significance even when real variability exists. In meta-analyses, k is often small and so the power of the Q-test is correspondingly limited (Borenstein et al., 2009; Hunter & Schmidt, 2004). Taking this caveat into account, Borenstein et al. (2009) developed recommendations for interpreting the significance of Q. Specifically, if the p-value for Q is less than 0.05, the synthesist can conclude that true heterogeneity exists. If p is greater than 0.05, however, Borenstein et al. recommend refraining from the conclusion that no true variability exists given that low statistical power is often a problem with the Q statistic.

Hunter and Schmidt (2004) and Borenstein et al. (2009) also point out that while Q has some value for determining whether true variability exists, it does not tell the synthesist *how much* true variability exists. As a consequence, synthesists who use Q exclusively to determine whether explanatory analyses are warranted may misinterpret their results. For instance, if Q is significant, but the amount of true variability is small, then moderator analyses may lack the statistical power needed to detect underlying relationships if they exist. Thus, while Cochran's Q can be a useful indicator of whether true heterogeneity exists, it often suffers from low statistical power and provides no information about the amount of variability available to be explained.

τ and τ^2 (T and T^2)

To counter the problems associated with Q, synthesists sometimes turn to the variance components of τ and τ^2. These components represent the variance and standard deviation among the *true* study effects, and can be used to describe the distribution of effects around the grand mean (Borenstein et al., 2009). Just like the variance and standard deviation of a sample, these parameters provide the synthesist with insight regarding the absolute variability among the true effects. Accordingly, τ and τ^2 allow the synthesist to more critically evaluate the information communicated by Q and p. For instance, if the synthesist suspects that low statistical power is the reason for a non-significant Q, τ and τ^2 can be used to support or question their suspicion. If τ and τ^2 are large, then the low statistical power explanation gains credibility; but, if τ and τ^2 are small, then more confidence can be placed in a non-significant Q.

Because τ and τ^2 are parameters, they need to be estimated from the sample study data using T and T^2 (see Borenstein et al., 2009). Specifically, τ^2 can be estimated as:

$$T^2 = \frac{Q - df}{C},$$

where

$$C = \Sigma w_i - \frac{\Sigma w_i^{\,2}}{\Sigma w_i}.$$

Because the degrees of freedom can sometimes exceed Q, negative values for T^2 can be obtained. In such cases, T^2 is set to zero. Thus, values of T^2 can range from zero to infinity. Once T^2 has been calculated, τ can be estimated as:

$$T = \sqrt{T^2}.$$

Borenstein et al. (2009) point out that the process of dividing T and T^2 by C has two advantages. First, dividing by C places the variance components back on their original metric (i.e., the metric of the effect size), thereby making them interpretable. Second, the use of C reduces the influence of k on our assessment of the amount of variability available to be explained because T and T^2 are averages instead of sums like the Q-statistic. Thus, T and T^2

provide an indication of the absolute amount of true variability, while using a metric that is both interpretable and relatively unaffected by the number of studies sampled. On the other hand, Borenstein et al. (2009) point out that because T and T^2 are placed on the original metric, these statistics cannot be compared across meta-analyses that use different outcomes. Thus, T and T^2 do not allow synthesists to compare the absolute amount of true variability that is available to be explained across different types of outcomes (e.g., attendance frequency vs. prayer frequency).

I^2

To address the lack of comparability of T and T^2 across meta-analytic studies, Higgins, Thompson, Deeks and Altman (2003) introduced the I^2 statistic. The I^2 statistic represents the proportion of observed variability that is considered real, and can be estimated as follows:

$$I^2 = \frac{Q - df}{Q}.$$

Because I^2 is a proportion, its values range from 0 to 1.[82] Values close to zero indicate that there is very little true variability relative to the observed variability. Values close to one indicate that all of the observed variability is real. Importantly, multiplying I^2 by 100% makes it comparable to other proportion of variance statistics, such as R^2 and η^2. And, just as guidelines have been introduced to evaluate the magnitude of these other proportion of variance statistics, Higgins et al. suggested the following criteria to evaluate the magnitude of the I^2 statistic:

25% = Low
50% = Moderate
75% = High

Besides the intuitive nature of the I^2 statistic, it also has the advantage of being unaffected by either the metric of the effect size or the number of studies included in the meta-analysis (k). For these reasons, I^2 statistics can be compared across meta-analyses. Despite this advantage,

[82] Negative values are possible if Q is less than the degrees of freedom, but in such cases, I^2 is set to zero.

the I^2 statistic does not indicate whether a significant amount of real variability exists (as Q and p do); nor does it measure the absolute amount of real variability (like T and T^2). Thus, like the other indicators of heterogeneity discussed here, I^2 is a useful statistic, but by itself, provides only a limited picture of the meta-analytic variability.

Considering the Measures of Heterogeneity Together

As indicated above, each measure of heterogeneity offers a unique perspective on the meta-analytic variability. Cochran's Q and p provide an indication of whether true heterogeneity exists above that expected by chance. Yet, Q often suffers from low statistical power, and says nothing about the more important issue regarding the amount of true variability available to be explained. The τ and τ^2 statistics provide this information, but they are affected by the metric of the effect size and are not comparable across studies. Finally, I^2 provides an intuitive measure of the ratio of true variability to total variability, and can be compared across studies. Given that it is a ratio, however, the I^2 statistic cannot be used to determine the substantive implications of the amount of heterogeneity that exists across studies, nor can it be used to determine if the observed variability exceeds that expected by chance. Thus, the three types of indices need to be used together to paint a complete picture of the distribution of effect sizes.

To paint this picture, Borenstein et al. (2009) recommend the following approach: (1) Use Q and p to help determine whether substantive variance exists; (2) estimate τ and τ^2 (regardless of whether Q is significant) and evaluate the range of effects and the amount of heterogeneity available with regard to theory and "practical significance" (e.g., if k is small, Q is not significant, but τ is moderate-to-large, then there is likely to be some real variability available to analyze and explain); and, (3) calculate the I^2 statistic to determine the proportion of the observed variance that is real. If each of the three variance indicators suggest that true heterogeneity exists, then moderator analyses are clearly warranted. If the variance and proportion of variance statistics are sizeable, but Q is not significant, then moderator analyses are probably still warranted, especially if k is small. On the other hand, if Q is not significant, and the variance and proportion of variance statistics are relatively small, then moderator analyses should be abandoned in order to avoid Type-I errors.

Confidence Intervals for τ *,* τ^2 *and* I^2

When evaluating the variance components, τ , τ^2 *and* I^2, it may be useful to consider them within the context of their respective confidence intervals. Borenstein et al. (2009) provide these formulas in their text, but state that an adjustment is first needed for the standard error before the confidence intervals can be calculated. This adjustment varies slightly depending on the value of both Q and the degrees of freedom (df). Specifically, if Q is greater than df + 1, then the adjustment (B) is as follows:

$$B = (0.5) * \frac{\ln(Q) - \ln(df)}{\sqrt{2*Q} - \sqrt{2*df - 1}}.$$

If Q is less than df + 1, however, then the following adjustment is needed:

$$B = \sqrt{\frac{1}{2*(df-1)*(1-(\frac{1}{3*(df-1)^2}))}}.$$

Once the adjustment has been estimated, intermediate values must also be obtained:

$$L = Exp(0.5*\ln(\frac{Q}{df}) - 1.96*B),$$

$$U = Exp(0.5*\ln(\frac{Q}{df}) + 1.96*B).$$

Then, the 95% confidence limits for τ^2 can be calculated as:

$$LL_{T^2} = \frac{df * (L^2 - 1)}{C},$$

$$UL_{T^2} = \frac{df * (V^2 - 1)}{C}.$$

The 95% confidence intervals for τ can then be calculated as:

$$LL_T = \sqrt{LL_{T^2}},$$

$$UL_T = \sqrt{UL_{T^2}}.$$

And, finally, the 95% confidence intervals for I^2 can be obtained by:

$$LL_{I^2} = (\frac{L^2 - 1}{L^2}) * 100\%,$$

$$UL_{I^2} = (\frac{U^2 - 1}{U^2}) * 100\%.$$

Prediction Intervals

Prediction intervals (known elsewhere as credibility intervals; see Hunter & Schmidt, 2004) are also a useful tool for helping synthesists evaluate heterogeneity. Whereas confidence intervals describe the range within which the true grand mean might be found, prediction intervals describe the range within which a specified percentage of true effects are likely to be found. This definition assumes that there are multiple true effects, and so prediction intervals are only appropriate for random-effects models. Because the true effects can only be estimated, however, a normal distribution of effects cannot be guaranteed unless k is large. Thus, the t-distribution is used instead of the z-distribution when calculating prediction intervals. Specifically, the 95% prediction intervals can be calculated as:

$$LL_{Pred} = M^* - (t_{df} * \sqrt{T^2 + V_{M^*}})\ ,$$

$$UL_{Pred} = M^* + (t_{df} * \sqrt{T^2 + V_{M^*}})\ ,$$

where $df = k - 2$.

Just as we saw with the fixed- and random-effects models, confidence intervals include only the within study variance (V_m), whereas prediction intervals include both the within and between studies variance (V_m, T^2). Thus, as sample size increases, confidence intervals will continue to shrink, whereas prediction intervals will shrink only to a point; thereafter, a comparable increase in the number of studies (k) is needed to continue shrinking the interval.

Forest Plots

A forest plot is a visually enlightening graph that allows one to quickly discern both the magnitude of each study effect and the precision with which it was estimated. Forest plots accomplish this by vertically plotting the magnitude of each effect size (demarcated as a square) on separate lines with whiskers extending out on either side to represent the width of the respective confidence intervals (wider whiskers equal less precision). Forest plots also allow the analyst to discern whether each study effect is significantly different from zero, and whether two study effects are significantly different from each other. If the whiskers for a particular study effect cross zero, then the study effect is not significantly different from zero. On the other hand, if the whiskers do not cross zero, then the effect is significantly different from zero. Similarly, if the whiskers for two study effects overlap each other, then there is no difference between the two, whereas if the whiskers do not overlap, then a significant difference exists.

Figure 7 provides an example forest plot using data from seven fictitious studies reporting proportions as outcomes. The study proportions represented in the first seven rows range from approximately 0.20 to just over 0.50, with the third and sixth proportions being significantly different from each other (because their whiskers do not overlap). The final, and perhaps most important, element of the forest plot is the cumulative effect. Located at the bottom of Figure 7, the cumulative grand mean proportion is represented with a diamond. Here, the center of the diamond represents the point estimate for the cumulative effect, and the length of the diamond represents the confidence interval. From Figure 7, we can see that the cumulative proportion is approximately just over 0.40, with a confidence interval ranging from approximately 0.35 to just under 0.50. Thus, forest plots contain several invaluable elements as they allow the synthesist and reader to gain a quick visual summary of an entire literature.

Meta-Regression

In the same way that regression can be a more efficient statistical tool in primary research, meta-regression can also be more efficient in meta-analytic research. When multiple covariates or moderators need to be controlled, for instance, the traditional method of meta-analysis used a hierarchical approach where separate meta-analyses were conducted for each combination of moderator levels. For example, if a synthesist wanted to control for gender and race, then separate meta-analyses would need to be conducted for each combination of gender and race (e.g., African American males, White males, African American females and White females). Accordingly, continuous moderators had to be transformed into categorical moderators prior to the analysis, which incurs a loss of statistical power. Meta-regression, on the other hand, allows for the simultaneous control of both categorical and continuous moderators (Borenstein et al., 2015). Thus, an analyst may include a number of moderators (as many as are allowed by the degrees of freedom available for the analysis) to estimate the expected outcome across all studies after holding the values of the moderators constant. Given the efficiency of this analytic tool, it will be used to carry out the analyses associated with each of this study's hypotheses, and to ultimately estimate the prevalence of religious service attendance in America while holding both methodological and socio-demographic sources of variance constant.

Importantly, meta-regression not only relies on the indicators of heterogeneity discussed above (i.e., Q, p, τ τ^2 and I^2), but it also introduces the R^2 statistic that is commonly associated with regression analyses in primary studies. The meaning of each of the heterogeneity indicators, however, is slightly different in meta-regression. Whereas Q and p provide an indication of whether real variability exists between all studies in a standard meta-analysis, these indicators are used to determine whether real variability exists between all studies *at a given value of the moderator variable(s)* in meta-regression. A significant Q in meta-regression, then, indicates that real variability around the regression line remains to be explained (Borenstein et al., 2015).

In standard meta-analysis, τ and τ^2 represent the amount of real variability that exists between all studies. In meta-regression, however, these two variance estimates indicate the amount of real variability that exists between studies *at any given value of the moderator variable(s)*. Thus, τ represents the standard deviation among the real study effects or outcomes at any given value of the moderator(s) in the regression model.

Similarly, I^2 in standard meta-analysis provides an indication of the proportion of all observed study outcome variability that is real. In meta-regression, I^2 indicates the proportion of all observed variability *about the regression line* that is real. Building off of the I^2 statistic, meta-regression introduces the R^2 statistic. The R^2 statistic provides an indication of the proportion of real variability (estimated in I^2) that is explained by the covariates in the meta-regression model. This, in turn, can be used to gauge the quality of the model, and to compare the abilities of different models to explain variance in the dependent variable.

Finally, confidence intervals and prediction intervals can both be calculated for predicted outcome at a given value of one or more of the moderators in meta-regression model. Confidence intervals are calculated by substituting the 95% lower and upper limits of the unstandardized moderator coefficient[83] in the prediction equation,[84] and solving for the respective predicted lower and upper values of the outcome. Prediction intervals, on the other hand, can be calculated by adding and subtracting 1.96 * *T* (standard deviation of the real study effects) to the predicted value of an outcome at a given value of one or more of the moderators. Together, these statistics indicate the range within which the real value of a predicted outcome is likely to be found (95% of the time) at a given value of one or more of the moderators, and the range within which predicted values of an outcome at a given value of one or more of the moderators are likely to fall (95% of the time) in future studies.

Given the interpretative value added by the measures of heterogeneity discussed here, each will be presented or discussed along with the results from each meta-regression model presented in the following chapter.

[83] These 95% lower and upper limits are provided in the CMA meta-regression output file.
[84] The standard form of the prediction equation is $Y_{pred} = a + bx$, where a = the Y-intercept and b = the slope of the line associated with the moderator "x."

CHAPTER IV

RESULTS

The primary goals of this study include: (a) Estimating the prevalence of attendance in America; (b) ascertaining the degree to which question wording and data collection methodology shape the answer; and, (c) understanding how the socio-demographic characteristics of samples shape the answer. Accordingly, this chapter sets out to address these three research questions. Before doing so, the data obtained for this study will be described along with the transformation and harmonization procedures used to place the data on comparable scales.

Addressing the Literature Search Problem for Religious Service Attendance Data

Meta-analyses typically follow the literature review search methods described in Chapter II. In part, this is because the available literature on a topic is relatively limited (e.g., see Higgins et al., 2003) and extensive efforts are needed to acquire enough data points for a meaningful statistical summary of the literature. Religious service attendance, however, is one of the most commonly used indicators of religiosity and religious behavior in the literature. This is because, apart from belief, study and prayer, attendance represents a major part of what it means to "be religious" in America (Mockabee et al., 2001); attendance has been found to be a key explanatory variable for a wide array of outcomes in a number of fields (as alluded to in Chapter II); and it "costs" researchers very little to measure given that it typically involves using only a single item. These factors have led religious, political, social, economic, medical and criminological researchers to treat attendance like a demographic variable in studies that have a primary, secondary or even a tertiary focus on religion. Consequently, even preliminary searches of several electronic databases using two very specific sets of search criteria (i.e., "Religious" AND "Service" AND "Attendance," as well as "Church" AND "Attendance"), a time frame that coincides with the life of the Gallup Poll (1939 to 2017) and parameters on language (English) and location (United States) yielded hundreds, and in some cases, tens of thousands of results. For instance, a search of *EBSCO*'s databases yielded 570 results for religious service attendance and 15,310 for church attendance! Using these two sets of search terms, *OVID* yielded 2,134 results, *Web of Science* yielded 1,383 results, *ProQuest* yielded 943 results and *FirstSearch* yielded 14,510 results. Searches for only religious service attendance on *JSTOR* and *Google Scholar* yielded an astronomical 46,969 results and 56,600 results, respectively. Furthermore, some of the key religious research journals that are candidates for manual scans—given the high probability that they contain religious service attendance data—date back to the late 1950s and early 1960s,[85] and contain hundreds or even thousands of articles. Even primary data on attendance are voluminous. For example, the Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan contains over 4,000 survey data points pertaining to religious service attendance.

Given the sheer volume of data unearthed by even a casual search of the literature, coupled with the time frame for this study, it became apparent that the use of traditional meta-analytic methods of searching the literature for religious service attendance frequency were not going to be feasible given the time and resource limitations allotted for this project. That is not to say that a thorough and complete search of the literature would not yield a rich set of data from which a number of valuable discoveries might be made, but the scope of this project necessitates that a practical and feasible approach be taken to collect the data needed to answer the research questions posed here.

[85] The Review of Religious Research and the *Journal for the Scientific Study of Religion* date back to 1959 and 1961, respectively.

Revised Scope of the Literature Search

Recall that the primary foci for this study are to estimate the prevalence of attendance in America, ascertain the extent to which question wording and data collection methodology shape the answer and understand how the socio-demographic characteristics of samples influence the answer. Accordingly, the literature search focused on acquiring data sources that would allow this study to address these goals. To estimate the prevalence of attendance in America, nationally representative polls and surveys were sought. These types of surveys, which typically use probability sampling, are designed to represent the opinions, beliefs, attitudes and behaviors of Americans, and they tend to do so with relatively small margins of error. Moreover, these types of surveys have been asking the same questions for decades, which allows researchers to study trends over time and associations with other variables at the meta-analytic level. Therefore, nationally representative surveys seem ideally suited to allow for an estimation of the prevalence of religious service attendance in America.

Because the researchers and organizations that administer national surveys tend to stay current on the latest measurement trends, their surveys are also likely to reflect new variations in question wording and data collection methodology. Thus, these types of surveys also seem ideally suited to explore the influence of item and method variants on estimates of attendance. Not all data collection methods can be represented by national surveys, however. The count-based method, for instance, cannot be carried out via survey, thereby making it difficult to investigate at a national level. In addition, the paucity of studies using the count method underscores the need to be over-inclusive with regard to this method. In order to arrive at a relatively robust estimate of count-based attendance, then, it was necessary to incorporate all known sources of count data, whether they represent a local area or the nation.

Finally, surveyors seeking to represent the nation's population must also collect demographic data to verify whether the sample does, in fact, represent the nation. This latter characteristic happens to varying degrees, with some surveys measuring more demographic variables than others, but the ubiquitous presence of demographics in the national polls suggests that it will be possible to examine their influence on estimates of attendance.

Thus, while the traditional method of reviewing the literature is not feasible for this study, the use of national polling data and select local studies for the count-based method should provide ample data for addressing the research questions posed here.

Data Sources

In preparation for the writing of the literature review for this study, this author read and reviewed over 1,000 peer-reviewed articles, theses, dissertations, books and book chapters. These sources were obtained largely from (non-systematic) electronic bibliographic database searches and manual scans of articles referencing other potentially relevant studies. In addition, this author has subscribed to two prominent religious research journals, *Journal for the Scientific Study of Religion* and the *Review of Religious Research*, since 2003 and to a newer, but still prominent, journal, *Psychology of Religion and Spirituality*, since 2014, which, together, contain hundreds of articles. A large portion of these sources were scanned for studies focusing on religious behavior, involvement and attendance. This search process helped identify a number of data sources, including nationally representative surveys, data warehouses and data consortiums containing religious attendance data dating back decades.

The data sources to be included in this study are summarized in Table 6. Twenty-six data sources were located for this study via electronic bibliographic searches and manual scans of the literature (n = 8), survey-specific data sites (n = 3), data consortiums (n = 3) and the grey literature (n = 1). The manual scans identified nationally representative data from three peer-reviewed publications: *American Sociological Review, Journal for the Scientific Study of Religion* and the *Review of Religious Research*. The data sites included those for the *American Heritage Time Use Study* (AHTUS; Fisher & Gershuny, 2015), *Pew Research Center* (Pew Forum on Religion & Public Life, 2007) and *NORC at the University of Chicago* (Smith et al., 2016). The data consortiums include *Michigan's Inter-University Consortium for Political and Social Research* (ICPSR), the *Association of Religion Data Archives* (ARDA) and the *Roper Center* at Cornell University. The lone grey literature data source came from following a lead in a newsletter—*Crossroads: Exploring Research on Religion, Spirituality and Health* published by the Duke Center for Spirituality, Theology & Health (2017)—to the website of the Black Women's Health Study (BWHS, 2017) and finally to the contact information of Dr. Lynn Rosenberg, BWHS principle investigator, who provided data from the BWHS (personal communication, June 1, 2017).[86]

Collectively, the datasets span 65 years and include 271 estimates of attendance (see Table 6 and Figure 8). These estimates are drawn from the responses of 1,230,237 participants. In addition, 248,524,980 Americans are represented in 11 count-based attendance estimates, bringing the total number of Americans represented by these data sources to 249,755,217. At

[86] A number of additional data sources were available, but either did not draw data from a nationally representative sample or provided data that could not be used in this study. For example, a number of studies failed to report the n-sizes or proportion of sample members selecting each response option. Some studies reported these figures for what they identified as regular attenders; other studies reported figures for combined response option categories; and a number did not report any data at the response option level, a large number of which reported mean attendance instead of the proportion attending.

least some demographic data are available in 19 of the 26 data sources. Gender, race and ethnicity and age are each found in 19 sources; education and marital status are available in 14 and 13 data sources, respectively; income is available in 11 data sources; and, both parental status and employment status are available in 10 data sources.

Item Wording and Mode of Delivery Coding

Table 7 provides information on question wording and data collection methodology. With regard to question wording, 14 of the 26 data sources measure attendance using "How Often" questions (e.g., the GSS asks, "How often do you attend religious services?"); three data sources[87] measure attendance in the last week (e.g., the Gallup Poll asks, "Did you, yourself, happen to attend church or synagogue in the last seven days?"); three data sources provide count-based attendance; and, four data sources provide time-use attendance.

Each of the items and methods are coded to indicate whether they attempt to address the criticisms levied against the gold standard items (see Table 7). For example, among the "How Often" questions, both the High School & Beyond survey and the National Education Longitudinal Study (NELS) modify the GSS gold standard item by asking, "*In the past year* [emphasis added], about how often have you attended religious services?" This wording modification addresses the ambiguous time-frame criticism by focusing the respondent's answer on their behavior during the last year. These two sets of items are also delivered to respondents via paper-and-pencil self-administered surveys. Because self-administered surveys potentially incur less socially desirable responding than either face-to-face or phone interviews, these items also address the social desirability criticism. The American National Election Study (ANES) item (see Table 7), on the other hand, addresses social desirability not by mode of delivery

[87] Two of the three data sources are from the Gallup Organization. The wording difference between the two items is slight (i.e., one item omits any mention of "synagogue"), but may alter response patterns. Accordingly, the two item versions are treated as different data sources.

(phone interview), but through question wording. Specifically, the ANES item opens with a statement that normalizes religious service absenteeism ("Lots of things come up that keep people from attending religious services even if they want to.") While respondents to the ANES may experience heightened social desirability motivations given that they are interacting with a person over the phone, this statement is designed to lessen that impact. The ANES item also addresses the problem of ambiguous wording by defining activities that are to be omitted from the definition of a religious service ("weddings, baptisms, funerals"). Thus, Table 7 includes codes for "T-F" (ambiguous time frame) and "SD" (social desirability) next to the High School & Beyond and NELS items, and "W" (ambiguous wording) and "SD" next to the ANES item to indicate the problems these items and methods address. Items that measure attendance in the past week inherently address the ambiguous time-frame problem (see Table 7). A variant of this item (i.e., the GSS Branching Item) also addresses ambiguous wording. Finally, the count-based method and time-use items address all three criticisms levied against the gold standard items and are coded accordingly in Table 7.

In all, 11 data sources used items or methods that potentially limit socially desirable responding; 10 data sources use items that at least partially address the problem of ambiguous wording; and 9 data sources use items or methods that address the ambiguous time-frame problem (see Table 7). Six data sources, however, do not address any of the three criticisms levied against the gold standard items, and will serve as baseline items from which comparisons can be made.

Religious Service Attendance Data Harmonization

The items and methods in Table 7 confirm that there are essentially two types of attendance measurement scales: (a) those that measure attendance in the past week (or yesterday); and, (b) those that ask respondents how often they attend. The former types of items and methods yield dichotomous outcomes (did not attend or attended). The latter types of items not only offer more response options, but can be further distinguished by the number and wording of these response options. For example, the ANES item uses the following six-point response scale: “Never,” “A few times a year,” “Once or twice a month,” “Almost every week,” Every week” and “More often than once a week.” The GSS, on the other hand, uses the following nine-point scale: “Never,” “Less than once a year,” “About once or twice a year,” “Several times a year,” “About once a month,” “2 to 3 times a month,” “Nearly every week,” “Every week” and “Several times a week.” Not only do the two items differ with respect to the number of response options, but two of the scale points are also worded differently (i.e., “A few times a year” vs “Several times a year” and “Once or twice a month” vs “2 to 3 times a month”). This measurement variability creates a problem for those wanting to cumulate attendance data across studies and those wanting to compare attendance rates across items. If this variability remains, any attempt to cumulate or compare will yield fuzzy and uninterpretable results.

Fortunately, researchers have faced this dilemma before and have developed a method for cumulating and comparing disparate response scales. This method has been referred to as harmonization (see Rossi & Scappini, 2014), and involves converting disparate response scales to a common metric. Presser and Stinson (1998), and more recently, Rossi and Scappini (2014) used a harmonization method to compare GSS and time-use data. Specifically, they converted the GSS response options into probabilities representing the likelihood of attending in any given week. Because the probabilities estimated attendance at the weekly level, they could then be compared to the time-use items, which directly measure attendance in a given week.

In this study, a similar approach was taken. Specifically, the attendance data were harmonized such that a common metric equates to the probability of attending in any given week. Like Presser and Stinson (1998) and Rossi and Scappini (2014), this study began the harmonization process by using the GSS item as a starting point. The GSS item is ideal because it offers more response options than most other items and because the wording of the response options offers a great deal of overlap with other items. Because there was some disagreement in the weights assigned to the GSS by Presser and Stinson and Rossi and Scappini, however, this study will use a hybrid approach (see Table 6). First, Rossi and Scappini's practice of using three decimal places (as opposed to the two decimal places used by Presser and Stinson) was used. Second, Presser and Stinson's assumption that even the most frequent attenders will miss on occasion leads to the assignment of the first two weights. Specifically, "Several times a week" (0.999) and "Every week" (0.990)[88] were assigned weights that are very close to, but slightly less than 1.0 (see Table 8). For all but one of the remaining categories, however, the Rossi and Scappini weights were used (see Table 8).[89] The exception stems from the weight applied to "Several times a year." Rossi and Scappini (2014) apply a weight of 0.115, which

---

[88] Presser and Stinson used a weight of 0.99 for both of the two most frequent response options. Here, the use of three decimals allows for a distinction to be made between the two. A slightly lower weight is applied to "Every week" given that a single miss would mean the respondent did not attend "in the past 7 days," whereas a single miss by someone who attends "Several times a week" is less likely to lead to such an outcome.
A weight of 0.999 is applied to "Several times a week," the use of two slightly different weights

[89] The additional discrepancy comes from the weight applied to "Several times a year." Rossi and Scappini (2014) apply a weight of 0.115, which translates to 6 weeks out of the year or every other month. Presser and Stinson, on the other hand, use a weight of 0.080, which translates to just over four weeks out of the year. This study uses a weight of 0.077, which translates to exactly four weeks out of 52. The rationale for this is that other attendance items also include the response options, "A few times a year" and "Less than once a month" Because "Several times a year" is more than "A few times a year," and because the halfway point between "Never" and "Less than once a month" is 6 times per year, we set the following weights for these respective response options: "A few times a year" (0.058 or 3 times per year), "Several times a year" (0.077 or 4 times per year) and "Less than once a month" (0.115 or 6 times per year).
The former researchers used weights of 0.846 and 0.115, respectively, which represent the halfway points between the next highest and lowest response option weights. The latter researchers used weights of 0.75 and 0.08, respectively, which seem somewhat arbitrarily chosen. Furthermore, the former value of 0.75 overlaps with part of the next lower response option, where "2 to 3 times a month" would convert to weights ranging from 0.50 to 0.75.

translates to 6 weeks out of the year or every other month. Presser and Stinson, on the other hand, use a weight of 0.080, which translates to *just over* four weeks a year (i.e., 4.16), which is an arbitrary figure that results from using only two decimal places. This study used a weight of 0.077, which translates to *exactly* four weeks a year. The rationale for this weight comes from juxtaposing "Several times a year" to response options from other attendance items, namely "A few times a year" and "Less than once a month." According to *The American Heritage College Dictionary* (2000), *a few* refers to more than one while *several* refers to more than two or three. Therefore, "Several times a year" can be considered slightly more than "A few times a year." "Less than once a month," on the other hand, seems to fall somewhere between "Never" and "Once a month." The halfway point would be six times per year, which would be slightly more than "Several times a year." Accordingly, this study assigned the following weights to these response options: "A few times a year" = 0.058 (or 3 times per year), "Several times a year" = 0.077 (4 times per year) and "Less than once a month" = 0.115 (6 times per year). The weights for the remaining response options were developed using a similar approach and are presented in Table 8.

After weights were assigned to all response options, the sample proportion that can be expected to be in attendance on any given Sunday can be calculated. This calculation can be accomplished by multiplying the response option weights by the frequency with which each response option was selected, summing the products and dividing by the sample size to yield an average probability of attending in any given week. For example, if we consider an item with five response options (with their respective weights in parentheses): "Never" (0.000), "Several Times a Year" (0.077), "Once a Month" (0.231), "Two to Three Times a Month" (0.577) and "Every Week or More" (0.999); and, if each response option is selected by 10, 5, 15, 20 and 15 respondents, respectively, then we first multiply the weights by their respective n-sizes: (10 * 0 = 0), (5 * 0.077 = 0.385), (15 * 0.231 = 3.465, (20 * 0.577 = 11.54) and (15 * .999 = 14.985); sum the products (0 + 0.385 + 3.465 + 11.54 + 14.985 = 30.375); and divide the sum by the total sample size (30.565 / 65 = 0.467). In this example, the probability of a sample member attending in a given week equals 0.467, or put differently, approximately 46.7% of the sample could be expected to attend on any given Sunday. In this way, the attendance data from the "how often" questions were harmonized so that comparisons could be made with questions and methods that measure attendance in the past week.

Demographic Data Transformations and Harmonization

Each demographic variable was transformed to control for any undue influence that sample size might have on the analyses. To accomplish this, the number of respondents in most of the respective demographic categories were transformed into proportions. The exceptions to this included education and income where means and medians were used to represent these variables. In addition, the number and type of categories for some of the demographic variables were inconsistently reported across studies. Thus, data harmonization techniques were applied where necessary in order to prepare the respective demographics for the analysis. These procedures are described below.

Gender. The preparation of the gender demographic was straightforward given that only male and female categories were reported in the sampled studies. Furthermore, the mutually exclusive nature of these categories means that only one variable needs to be introduced into the analysis to represent both genders. In this study, *proportion female* was selected as the indicator for a sample's gender composition. The use of a single variable to represent gender is statistically efficient given that it represents all of the information reported about gender while minimizing the degrees of freedom needed for the analysis. Across all studies, gender was available for 248 study observations, with 52.8% of the 1,326,838 respondents being female.

Race and Ethnicity. Studies varied with respect to the racial and ethnic groups reported. Earlier studies, particularly those conducted by the Gallup Poll, included only two response options (White, Black). Within the last few decades, however, the reporting of race and ethnicity grew increasingly more representative of the diversity in the U.S. For this study, the racial and ethnic data were collapsed into five categories: White respondents (represented in 246 study observations), Black, African Americans respondents (234 study observations), Hispanic respondents (139 study observations), Asian respondents (94 study observations) and respondents falling into the "Other" category (146 study observations). The sample sizes in each category were transformed into proportions to control for any undue influence of sample size.

Age. As expected, quite a bit of variability exists in the way researchers report their sample age characteristics. Among the study observations available for this study, 218 offered age category frequency data; 92 offered both age category frequency data and either a mean or median; and 14 offered only a mean or median. Even among the age category data, there was a good deal of variability with 37 age categories reported across all study observations. Because age is expected to be non-linearly related to attendance, and because attendance patterns seemed to have emerged in the literature across four broad age groups, the 37 age categories were recoded as closely as possible to the following: 13 to 17 years, 18 to 30 years, 31 to 64 years and 65 years-and-over. Study categories that spanned two of the latter theoretical categories were classified by using the mid-point of the original study category. For example, the 25- to 34-year-old group was used in some studies. To reclassify this category, the midpoint (29.5 years) was used to place this group into the 18- to 30-year-old group. In studies providing only a measure of central tendency, the sample participants were placed in the age category corresponding with the central tendency value. For example, a mean sample age of 44.58 was recoded into the 31- to 64-year-old group. In all, 230 study observations were available for the age group classifications.

Marital Status. Marital status was reported for just over one-third of all study observations (n = 108). The categories used to report marital status were fairly consistent across studies, especially for the category representing marriage. In fact, just one study used a category that included persons other than those who were married (i.e., “married or in a domestic partnership”). Given the small number of domestic partnerships observed in other studies, this combined category was placed into the married-only category for this study. The n-sizes for the married category were transformed into proportions to yield a proportion married variable.

Parental Status. Parental status was infrequently reported with just 57 study observations providing information on parental status and 56 observations providing information on whether respondents were parents of school-age children (defined loosely here as age 5 to 17). The number of respondents in each study indicating that they are parents (or parents of school-age children) was transformed into the proportion of the sample indicating as much.

Education. A number of educational categories were reported by the sampled studies. Each of these educational categories, however, was transformed into a number of educational years completed. Specifically, the following transformations were made to the respective educational categories: “Dropout” was recoded as 8 years of education; “Less Than a High School Diploma” was recoded as 10 years of education; “Diploma or Less” was recoded as 11 years; “High School Diploma” was recoded as 12 years; “Diploma or Some College” was recoded as 12.5 years; “Some College” was recoded as 13 years, “Associates or Technical Degree was recoded as 14 years; “Bachelor’s Degree was recoded as 16 years, “Some Graduate School” was recoded as 17 years; and “Graduate Degree was recoded as 18 years. Once the study categories were harmonized into years of education, the mean years of education achieved for each sample was calculated. In all, 150 study observations include sample data on years of education achieved.

Income. The categories reported for income are highly dependent on the year in which the data were collected. The Gallup data collected in 1977, for example, used categories ranging from “Under $2,000 a year” to “$20,000 and more a year.” By 2015, the range of these categories started at “Less than $10,000” and went to “$500,000 and over.” In order to harmonize the income data across years, it was necessary to first adjust for inflation. To accomplish this, the Bureau of Labor Statistics CPI Inflation Calculator (2017) was used. The mid-point of each survey income category was adjusted to reflect June 2016 dollars. The original survey income categories were then re-classified into the following 2016 income categories: “Less than $25,000,” “$25,000 to $49,999,” “$50 to $74,999,” “$75,000 to $99,999” and “$100,000 or more.” Once the original income categories had been adjusted for inflation and realigned into one of the new categories, the sample’s median income category was used to represent the sample’s average income. In all, 124 study observations provided data on income.

Employment Status. Employment status was reported in 104 study observations, and 10 categories were used across all studies to report on this demographic. One category, however, represented those working both full-time and part-time, but because full-time status and part-time status are expected to relate differently to attendance, the data from these studies were excluded, leaving 94 study observations. When converting the n-sizes per category into a proportion employed, respondents who indicated that they were disabled, retired or in school were excluded from the number of potentially employed persons. In this way, the employment status variable is an indicator of the proportion of persons employed full-time out of the persons who could potentially be in the workforce.

Data Analysis and Hypothesis Testing

Sixteen hypotheses, which were formulated in Chapter II, are used to guide the analyses for this study. Six hypotheses focus on assessing the impact of item and method differences on estimates of religious service attendance frequency, and ten focus on determining whether there are differences in attendance across the socio-demographic variables discussed in this study. Each of the 16 hypotheses are restated below, and are followed by the meta-analyses and findings associated with each hypothesis test (for a summary of hypotheses and findings, see Table 27).

Hypothesis 1. Studies that utilize items and methods that (explicitly or implicitly) provide a specific time frame will yield lower attendance estimates than studies that fail to do so.

Before testing Hypothesis 1, it is necessary to first identify the data sources that can best address the problem. To begin, baseline attendance needs to be established. Baseline attendance can best be established by drawing from data sources that have not implemented items or methods designed to minimize or eliminate the three criticisms levied against the gold standard items. Six data sources fit this description (see Table 7), but two of these focus on special populations, namely youth (National Study of Youth and Religion or NSYR) and African Americans (National Survey of Black Americans or NSBA), and were removed from consideration here. The remaining four data sources that can provide a baseline attendance rate include the Americans' Changing Lives Survey (Waves I – V), the new Gallup item, the new Gallup item that uses religiously inclusive language (i.e., the item adds mention of "Mosque" to "Church" and "Synagogue") and the original GSS item.

There are 12 data sources that address the ambiguous time frame problem. Two of these sources, however, focus on a special population (i.e., high school students in the High School & Beyond Study and National Education Longitudinal Study or NELS) and were removed from consideration. Of the remaining 10 data sources, four approach the time-frame problem by using modified item language. The original Gallup item, the original Gallup item that omits mention of "Synagogue" and the GSS branching item all accomplish this by asking respondents about their attendance in the past seven days. The 2008 ANES item accomplishes this by asking about attendance in the past year, and then by allowing respondents to indicate how many times they attended in the past year, the past month or the past week. Three data sources (AHTUS, ATUS and Presser & Stinson, 1998) use time-use methodology to gather information on attendance yesterday. Finally, three data sources (Chaves & Cavendish, 1994; Hadaway & Marler, 2005; Hadaway, Marler & Chaves, 1993) restrict the attendance time-frame by removing the participant all together (by counting) and selecting a researcher-defined time-frame. It is recognized that the latter two data sources minimize or eliminate all three item criticisms (not just the time-frame criticism) and will likely yield markedly different attendance rates than the items that just address the time-frame criticism, but their inclusion here will help address whether the specification of a time frame (via item wording or methodology) influences estimates of attendance.

After selecting the baseline and time-frame data sources, a random-effects meta-regression model, using *Comprehensive Meta-Analysis* (CMA), *Version 3.3,* was developed to test Hypothesis 1. Specifically, a dummy-coded variable was used to represent the baseline data sources (n = 60) and the time-frame data sources (n = 103). A covariate was also introduced to represent the year in which each study observation was made (or reported). These two variables

(Time-Frame and Year) were then regressed on the logit-transformed proportion of each study sample expected to be in attendance on any given Sunday.[90]

The initial results of the meta-regression indicated that four count observations (i.e., the Mainline Protestant and non-Christian observations from Hadaway & Marler, 2005; the Western Region Catholic observation from Chaves & Cavendish, 1994; and, the observation from Hadaway et al., 1993) were exerting undue influence on the results, as indicated by their studentized residuals. According to Borenstein et al. (2015), studentized residuals are distributed approximately normal with a standard deviation of about 1.0. Thus, values equal to 1.96 or greater can be expected to occur by chance about 5% of the time while values greater than 2.58 can be expected to occur by chance about 1% of the time. This study uses the latter criteria to identify data points with excessive influence given that it is desirable to retain as much data as possible. The studentized residuals associated with these four observations were above the 2.58 threshold (3.07 for non-Christian counts and 2.77 for Mainline Protestant counts in Hadaway & Marler, 2005; 3.03 for Western Region Catholics in Chaves & Cavendish, 1994; and 2.94 for the Hadaway et al., 1993 observation), and were removed from the analysis.

The results of the meta-regression with 159 observations (60 in the baseline group and 99 in the time-frame group) are presented in Table 9. The model containing Year and Time-Frame was significant, $F(2,156) = 13.89$, $p < .001$. After controlling for study year, the results of this analysis lend support to Hypothesis 1. Specifically, the use of items or methods to address the ambiguous time-frame problem is associated with a significant reduction in the attendance rate, B = -0.2660, $t = -5.23$, $p < 0.001$. If study year is held constant at its most recent value (2015),

[90] Because proportions are often skewed, the CMA program automatically converts proportions into logits. As a result, the regression model coefficients are also represented as logits. The predicted probability of attending on any given Sunday can be calculated by solving the regression equation and converting the predicted logit back into a proportion via the following formula: Exp(logit) / 1 + Exp(logit).

then the baseline attendance rate (i.e., with no controls for ambiguity or socially desirable responding) is 43.7%. The predicted attendance rate for items and methods designed to address the ambiguous time frame problem is 37.3%. This represents a 6.4% drop, and an inflation factor of 1.17 (or 17%). Thus, hypothesis 1 is supported.

Hypothesis 1a. The length of the time frame specified by an item or method will be positively related to estimates of attendance (e.g., "regular" attendance > attendance "in the past week" > attendance "yesterday").

Testing Hypothesis 1a involves breaking down the time frame observations that were grouped together for Hypothesis 1. Three data sources ask respondents to refer to the "last 7 days" (original Gallup, original Gallup omitting "Synagogue" and the GSS branching item); three data sources ask respondents to refer to "Yesterday" (AHTUS, ATUS and Presser & Stinson, 1998) and three data sources remove the time-frame from the respondent all together by counting attenders within a researcher-defined time frame (Chaves & Cavendish, 1994; Hadaway & Marler, 2005; and Hadaway et al., 1993).[91] Of the 162 observations derived from these nine data sources, two were identified as having an undue influence on the results (the Orthodox count data from Hadaway & Marler, 2005 and part one of the GSS branching item), as evidenced by extreme studentized residuals (2.92 and -2.62, respectively). Consequently, these two observations were removed, leaving a total of 160 observations for the analysis (60 baseline observations, 60 observations for the "Last 7 days," 30 for "Yesterday" and 10 for "Researcher defined counts").

The results of the random-effects meta-regression are presented in Table 10. The model containing Year and three dummy variables representing time-frame length was significant,

[91] One observation also asked respondents to refer to the "last 12 months" (2008 ANES), but because it is a single data point it is not included in a regression analysis that relies on clusters of observations to determine the line of best fit.

$F(4,155) = 99.16$, $p < .001$.[92] After controlling for study year, the results indicate partial support for Hypothesis 1a. Compared to baseline attendance, the specification of a time frame in the "last 7 days" was not associated with a significant reduction in attendance, B = -0.0301, $t = -0.80$, $p = 0.211$. Attendance "Yesterday," however, was associated with a significant reduction, B = -0.6784, $t = -14.03$, $p < 0.001$, as were the "Researcher Defined Counts," B = -0.9273, $t = -14.25$, $p < 0.001$. According to these results, then, weekly attendance is projected to be 43.1% when estimated by the gold standard items, 42.3% when estimated by items asking about attendance in the past 7 days, 27.7% when estimated by asking respondents about yesterday and 23.1% when researchers remove the reporting of attendance from the respondent all together and count those present in the pews. These latter effects are associated with inflation factors of 1.56 and 1.87, respectively.

These results are generally consistent with the literature, and suggest that attendance rates fall when shorter time-frames are used. The non-significant finding associated with the items measuring attendance in the past week, however, is unexpected given that they clearly specify a relatively short time frame. The difference (-0.8%) is in the expected direction, however, and could suggest that wording modifications simply yield small effect sizes, which were not able to be detected in this analysis. In fact, in separate analyses that did not include the count data, the indicator for "last 7 days" was significantly and negatively associated with the attendance rate. Thus, the "last 7 days" time frame deserves further investigation. The data from the present model, however, only provide partial support for Hypothesis 1a.

[92] Surprisingly, the R^2 for this model was estimated to be 0.0%. According to Borenstein et al. (2015), this is a unique situation to meta-regression where the variance estimates with and without covariates are conducted separately and so are subject to error from time-to-time. When the model clearly explains some of the variance, but the R^2 statistic is zero, it is likely a case where the variance explained by the model without covariates was underestimated while the variance estimated with covariates was overestimated. Such cases represent an anomaly, however, and do not indicate that the model is incapable of explaining variance in the dependent variable.

Hypothesis 2. Studies that utilize items and methods that control for, or eliminate the use of ambiguous terminology (e.g., "religious services," "church" or "synagogue") will yield lower attendance estimates than studies that fail to do so.

The same baseline data sources (i.e., ACLS, Waves 1-V, both new Gallup items and the original GSS item) that were used for Hypotheses 1 and 1a will also be used here to test the hypothesis that items and methods designed to clarify ambiguous language will lower estimates of attendance. The data sources that address the ambiguous wording problem include the American National Election Studies (ANES), the GSS branching item, the Religious Landscape Survey and the World Values Surveys (WVS). The count-based (Chaves & Cavendish, 1994; Hadaway & Marler, 2005; and Hadaway et al., 1993) and time-use studies (ATUS, AHTUS, Presser & Stinson, 1998) also minimize or eliminate problems with ambiguous wording and were included in the analysis. Given that item wording changes are expected to have a relatively small impact on attendance compared to the latter two sources, however, this analysis will look at the impact of these two data sources separately by using two dichotomous indicators.

There are 50 data sources addressing the ambiguous wording problem, 41 of which are approached methodologically (via time-use and the count-based method) and nine of which are approached via question wording clarifications. Four observations were identified as having an undue influence on the results with studentized residuals equal to -3.56 (GSS branching item), 2.68 (1965 time-use study reported in Presser & Stinson, 1998), 2.68 (1966 AHTUS) and -2.66 (count data for non-Christians reported by Hadaway & Marler, 2005). After removing these observations, there were 106 available observations for the analysis (60 baseline, 8 item wording and 38 method sources).

The results of the random effects meta-regression are presented in Table 11. Although the statistical model is significant, $F(3,102) = 115.40$, $p < .001$,[93] there is only partial support for Hypothesis 2. Specifically, the coefficient for item wording was not significant, B = 0.0780, t = 1.07, p = 0.1427, and it was in the wrong direction. Not surprisingly, the coefficient associated with the methodological approach to addressing the ambiguous wording problem was significant and in the expected direction, B = -0.7537, t = -17.77, $p < .001$. These results can be interpreted as follows: 45.4% of the American population reported attending in response to the gold standard items; this rate increased slightly, but not significantly, to 47.3% when responding to items that made wording changes designed to correct ambiguous wording; and, 28.1% of respondents indicated attending in response to the time-use and count-based methods that address the ambiguous wording problem. The latter finding was 17.3% lower than the estimate obtained by the baseline items, and translates to an inflation factor of 1.68 (or 68%).

The unexpected finding associated with the item wording changes could be the result of low statistical power for this group of study observations given that were only eight. Furthermore, the effect size for these wording differences was expected to be small, if one existed (i.e., the literature review suggested a decrease of only 7% over the gold standard items). Sampling error (in the wrong direction) was likely present as well given that three of the eight wording change observations reflected high rates of attendance (between 0.50 and 0.51) and all three were from the WVS. Thus, the small sample size may have led to several related deleterious conditions that negatively influenced the findings. Still, the methodological approach to minimizing or eliminating the ambiguous wording problem did provide at least partial support for Hypothesis 2.

[93] The R^2 for this model was estimated at 0.0%, which is likely an anomaly (see Borenstein et al., 2015).

Hypothesis 3. Studies that utilize items and methods that minimize or eliminate opportunities for socially desirable responding will yield lower attendance estimates than studies that fail to do so.

The analysis for Hypothesis 3 mirrors that for Hypothesis 2. Specifically, the baseline data sources from the ACLS (Waves 1-V), new Gallup items and the original GSS item were used as the referent. Two indicators were also used to represent the items and methods designed to reduce the impact of socially desirable responding on the attendance rate. The first indicator represents data sources that either made wording modifications (ANES 2000, 2002, 2004) or used the same item but a different mode of delivery (the self-administered Baylor Religion Surveys). The second indicator represents the same time-use and count-based data sources as in Hypothesis 2. Even though the latter sources have been compared to the referent already, it is appropriate to test and report the outcomes associated with these methods here given that their primary purpose is to reduce the effects of social desirability.

Just six study observations are available from two data sources that address the effect of item wording and delivery mode on attendance. On the other hand, there are 41 study observations from six data sources available to test the effect of the methodological approaches to reducing social desirability (i.e., time-use and counting). Together with the baseline observations (k = 60), there are 107 study observations available for this analysis. Three observations were identified as having an undue influence on the results with studentized residuals equal to 2.78 (1966 AHTUS), 2.71 (1965 time-use study reported in Presser & Stinson, 1998) and -2.65 (count data for non-Christians reported by Hadaway & Marler, 2005) and were removed, leaving 104 observations for the analysis (60 baseline, 6 item wording and delivery mode and 38 method sources).

The results of the random effects meta-regression are presented in Table 12. Here again, the full model is significant, $F(3,100) = 110.12$, $p < .001$,[94] but Hypothesis 3 receives only partial support. The coefficient associated with item wording and delivery mode changes was in the right direction, but not significant, B = -0.1067, $t = -1.25$, $p = 0.1066$. As before, the coefficient associated with the methodological approach was significant and in the expected direction, B = -0.7601, $t = -17.90$, $p < .001$. These results can be interpreted as follows: 43.9% of the American population reported attending in response to the gold standard items; this rate decreased slightly, but not significantly, to 40.0% for those responding to items that made wording changes designed to address social desirability or that were self-administered; and, 26.8% of respondents indicated attending in response to the time-use and count-based methods that address the socially desirability problem. The latter finding was 17.1% lower than the estimate obtained by the baseline items, and translates to an inflation factor of 1.64 (or 64%).

While the finding associated with the wording and delivery-mode changes was in the right direction, it failed to reach significance. Here again, this may have been due to low statistical power given that only six studies were available to represent this effect. Thus, further exploration is warranted before a solid determination can be made as to whether wording changes and the self-administration mode are viable options for minimizing socially desirable responding. For now, Hypothesis 3 can only be partially supported.

Hypothesis 3a. Minimizing or eliminating opportunities for socially desirable responding will suppress attendance estimates to a greater extent than controlling for, or eliminating ambiguous language (i.e., ambiguous time frame frames and terminology).

To address Hypothesis 3a, it is necessary to focus only on the data sources that address either the ambiguity problems or the social desirability problem, but not both. This excludes the

[94] The R^2 for this model was estimated at 0.0%, which is likely an anomaly (see Borenstein et al., 2015).

time-use and count-based methods because they address all three problems. The eligible data sources that address the ambiguous time-frame come from the original Gallup items (with and without "Synagogue" in the question text). The eligible data sources that address ambiguous wording come from the Religious Landscape Survey and the WVS. Two data sources address both the ambiguous time frame and ambiguous wording (2008 ANES and the GSS branching item), although these were later flagged and removed for exerting too much undue influence on the model (their respective studentized residuals were 2.60 and -2.60). Finally, one data source addresses only social desirability (Waves 1-3 of Baylor Religion Survey). A total of 67 study observations are available to represent the attempts to address ambiguity and social desirability, although 60 of these are from the Gallup Poll and address the ambiguous time-frame problem. In this analysis, the attendance rate for each group was compared to the attendance rate for the baseline referent group (k = 60).

The results of the random effects meta-regression are presented in Table 13. The full model is significant, $F(4,122) = 6.26$, $p < .001$, and accounts for approximately 14% of the variance in attendance. The coefficient associated with items attempting to address the ambiguous time-frame problem was not significant, B = -0.0394, $t = -1.24$, $p = 0.1084$. The coefficient associated with items attempting to clarify ambiguous language was significant, but, in the wrong direction, B = -0.2378, $t = 2.82$, $p = .0028$! Finally, the coefficient associated with items attempting to address the social desirability problem was also not significant, B = -0.1136, $t = -1.14$, $p = .1277$. These results simply compare each effect to the referent group, but predicted attendance rates and confidence intervals can be constructed and used to compare the three sets of items.

Using the model to predict 2015 attendance rates for the three types of items indicates that the social desirability items yield the lowest estimate of attendance at 45.1% (95% CI = 40.3% to

50.0%), followed by the ambiguous time-frame items at 46.9% (95% CI = 45.4% to 48.5%) and the ambiguous wording items at 53.9% (95% CI = 49.7% to 58.0%).[95] Although the direction of effects is consistent with Hypothesis 3a, the sample size for the social desirability and ambiguous wording groups are too small to yield a precise estimate. Consequently, the confidence interval for the social desirability items spans nearly 10 percentage points and overlaps with the confidence intervals for both sets of ambiguity items. Therefore, Hypothesis 3a is not supported.

Hypothesis 4. Among the items and methods reviewed here that attempt to reduce socially desirable responding, the count-based method will yield lower attendance estimates than the time-use item, which will yield lower estimates than the self-administered item.

The analysis for Hypothesis 4 is the first not to use the referent group of gold standard items. Instead, the time-use items will serve as the referent given that they are expected to produce attendance rates that are significantly lower than the self-administered items and significantly higher than the count-based studies. Again, this study suffers from a shortage of nationally representative adult population studies using self-administered surveys. The Baylor Religion Surveys of 2005, 2007 and 2010 serve as the only source of data pertaining to the self-administration mode. The time-use items, however, are well represented with 30 study observations from AHTUS, ATUS and the Presser and Stinson (1998) study. The count-based method also offers only about a dozen studies (n = 11), but the n-sizes within those studies are so large that the within-study variance is virtually zero, and the effect size is also expected to be large enough not to worry about low statistical power. All together there were 44 study observations available for the analysis, but the Orthodox Christian count data provided by Hadaway and Marler (2005) was flagged as having undue influence on the results (studentized residual = 2.68) and was removed, leaving 43 observations.

[95] Baseline attendance was 47.9%

The results of the random effects meta-regression model are presented in Table 14. The full model was significant, $F(3,39) = 17.30$, $p < .00$,[96] and the results support Hypothesis 4. The time-use items yielded attendance rates that were significantly lower than the self-administered item as represented in the Baylor Religion Survey, B = 0.6792, $t = 5.44$, $p < 0.001$, and significantly higher than the count-based studies, B = -0.2863, $t = -3.91$, $p < .001$. These results can be interpreted as follows: 40.2% of the American population reported attending in response to the self-administered items. This finding, while based on just a few studies, is consistent with the literature. The time-use items indicated that just over one-quarter (25.5%) of the American population attends religious services on a weekly basis, while the count-based method yielded an average weekly attendance estimate of just 20.4%. These findings provide clear support for Hypothesis 4.

Socio-Demographic Hypothesis Testing

The socio-demographic data were analyzed using the full set of 271 study observations. In case a socio-demographic category tends to be more heavily concentrated in studies using a particular type of attendance item or method, a set of three dichotomous indicators was used to indicate whether a study observation was made using an item or method designed to control for ambiguous wording, an ambiguous time-frame and social desirability. The inclusion of these indicators as control variables should allow for any bias associated with item wording or methodology to be removed prior to assessing the relationship between a socio-demographic and attendance. For example, the Monitoring the Future Survey (MTF) uses a self-administered item to gather attendance data from adolescents (8^{th}, 10^{th} and 12^{th} grade students). If self-administered surveys are associated with a reduction in attendance compared to phone or face-to-face surveys, and if adults are more likely to respond to the latter types of surveys, then any age

[96] The R^2 for this model was estimated at 0.0%, which is likely an anomaly (see Borenstein et al., 2015).

comparisons between adolescents and adults will minimize or obscure differences in their attendance patterns. Accordingly, the three item and method controls were included in each socio-demographic model.

Hypothesis 5. Studies with higher percentages of female participants will yield higher attendance estimates, on average, than studies with higher percentages of male participants.

It is expected that a higher proportion of female study participants will be associated with higher attendance rates. The proportion female statistic was available from a wide array of data sources, including those that measure attendance using the gold standard items (original GSS, both new Gallup items and the ACLS), "in the past week" items (both original Gallup items and the GSS branching item), self-administered items (Baylor Religion Survey and the MTF) and time-use items (AHTUS). The data sources also include studies focused on adolescents (NSRY, NELS, High School & Beyond and MTF) and special populations (Black Women's Health Study). In all, there are 247 study observations available for this analysis. Preliminary analyses indicated that five observations were exerting undue influence on the results given studentized residuals of -2.90 (GSS branching item), 2.84 (2004 ANES), 2.68 (2008 ANES), 2.65 (2002 ANES) and 2.64 (Religious Landscape Survey) and were removed, leaving 242 observations.

The results of the random effects meta-regression model are presented in Table 15. The full model was significant, $F(5,236) = 42.69$, $p < .001$,[97] and the results support Hypothesis 5. After controlling for study year and efforts to address the three attendance item criticisms, proportion female was found to be positively and significantly related to religious service attendance rates, B = 1.3734, $t = 4.91$, $p < 0.001$. Interpreting these results, the attendance rate for a sample

[97] The R^2 for this model was estimated at 0.47, indicating that 47% of the between-studies variation was accounted for by the variables in this model.

comprised of 52.8% females[98] is expected to be 41.9%.[99] For each 1% increase in the female participation rate, attendance is expected to increase by 0.33%. Thus, if the proportion of women increases by 5% so that women comprise 57.8% of a sample, the attendance rate would be expected to increase to 43.5%. An all-female sample would lead to an expected attendance rate of 57.9%, whereas an all-male sample would lead to an expected attendance rate of 25.9%, a difference of 32%! This difference is larger than what has typically been reported in the literature, but it is derived from a model that only controls for study year and item wording and methodology so perhaps a more fully specified model might yield a discrepancy more in line with what has been reported.

Hypothesis 6. The percentage of African American participants in studies will be positively related to the attendance estimates that these studies yield.

It is expected that a higher proportion of African American participants will yield higher attendance rates. The proportion African American statistic was available from the same data sources as for proportion female. In all, there are 233 study observations available for this analysis. Preliminary analyses indicated that two observations were exerting undue influence on the results given studentized residuals of 2.83 (1990 NELS) and 2.66 (2002 ANES) and were removed, leaving 231 observations.

The results of the random effects meta-regression model are presented in Table 16. The full model was significant, $F(5,225) = 29.78$, $p < .001$,[100] and the results support Hypothesis 6. After controlling for study year and efforts to address the three attendance item criticisms, the

---

[98] This proportion represents the arithmetic mean participation rate across the sampled study observations available for this analysis.
[99] This estimate was obtained by using parameter values equal to the average study characteristic. The lone exception was Year, 2015, which was selected because it is the most proximal.
[100] The R^2 for this model was estimated at 0.28, indicating that 28% of the between-studies variation was accounted for by the variables in this model.

proportion of a sample that is African American was found to be positively and significantly related to religious service attendance rates, B = 0.7385, $t = 4.58$, $p < 0.001$. Interpreting these results, the attendance rate for a sample with a 13.6% African American participation rate[101] is expected to be 44.2%.[102] For each 1% increase in the African American participation rate, attendance is expected to increase by 0.18%. Thus, if the proportion of African Americans in a sample increased by 5% so that African Americans comprise 18.5% of sample, the expected attendance rate would increase to 45.1%. If a sample were composed entirely of African Americans, the expected attendance rate would increase to 60.0%, whereas a sample composed of all non-African Americans would be expected to yield an attendance rate of 41.8%, a difference of 18.2%! This difference is high, but consistent with the literature.

The model above was limited to African Americans given that there was not enough available literature to formulate a hypothesis for other racial and ethnic minority groups. A meta-analysis can overcome this limitation, however, by examining covariation between sample composition rates (even if they are small) and study outcomes. Thus, an analysis with proportion of sample indicators for African Americans, Asians, Hispanics and Whites was undertaken and included 86 study observations. Four observations (2002 ACLS Wave IV, 2004 ANES, 1992 NELS, Religious Landscape Survey) were discarded due to high studentized residual values, leaving 82 for the analysis.

The results of the full race and ethnicity random effects meta-regression model are presented in Table 17. The full model was significant, $F(8,73) = 131.50$, $p < .001$.[103] After

[101] This proportion represents the arithmetic mean participation rate across the sampled study observations available for this analysis.

[102] This estimate was obtained by using parameter values equal to the average study characteristic. The lone exception was Year, 2015, which was selected because it is the most proximal.

[103] Surprisingly, the R^2 for this model was estimated to be 98%. This likely represents a unique situation to meta-regression where the variance estimates with and without covariates are conducted separately and so are subject to error from time-to-time. When the R^2 statistic is close to unity, it is likely a case where the variance explained by

controlling for study year and efforts to address the three attendance item criticisms, the proportion of African Americans, B = 2.6137, $t = 4.19$, $p < 0.001$, and Whites, B = 2.1614, $t = 3.33$, $p < 0.001$, were significantly and positively associated with attendance rates. Given that Hispanics are predominantly Catholic, and Catholics tend to have higher attendance rates (Gallup & Lindsay, 1999; Lazerwitz, 1961), it was surprising that the coefficient for proportion Hispanic was not significant, B = 0.6166, $t = 0.76$, $p = 0.224$. Interpreting the results for African Americans and Whites, however, the expected attendance rate for a sample with an average proportion of Asian, African American, Hispanic and White participants is 44.5%. Each percentage increase in the proportion of African Americans and Whites sampled is expected to produce an increase in attendance equal to 0.65% and 0.5%, respectively. Increasing these sample compositions by five percent will increase the attendance rate by 3.25% and 2.7%, respectively. Finally, if a sample were composed entirely of African Americans, the attendance rate would be 59.5%, whereas the attendance rate for an all-White sample would be 48.3%. This is a difference of 11.2%, which is consistent with what has been reported in the literature.

Hypothesis 7. The average age of a sample will be non-linearly related to religious service attendance estimates, such that studies of young adults (mean age = 18 to 30 years) will yield the lowest estimates, followed by adolescents (13 to 17 years), middle-aged adults (31 to 64 years) and elderly adults (65 years and older), respectively.

The same data sources that were available for gender and race and ethnicity are also available for age. The proportion of respondents who were between the ages of 13 to 17, 18 to 30, 31 to 64 and 65 and over were calculated for each study observation. In all, there are 228 observations available for this analysis. Preliminary analyses indicated, that four observations

---

the model without covariates was overestimated while the variance estimated with covariates was underestimated. Such cases represent an anomaly, however, and do not indicate that the model has fully explained the between-studies variance (see Borenstein et al., 2015).

were exerting undue influence on the results given studentized residuals of 3.16 (NSBA), 3.04 (2004 ANES), 2.98 (1990 NELS) and -2.72 (1992 NELS Dropouts) and were dropped from the analysis, leaving 224 observations.

The results of the random effects meta-regression model are presented in Table 18. The full model was significant, $F(8,215) = 24.78$, $p < .001$.[104] After controlling for study year and efforts to address the three attendance item criticisms, the proportion of a sample that is age 13 to 17 was found to be marginally and positively related to attendance, B = 0.4895, $t = 1.39$, $p = 0.083$; the proportion age 18 to 30 was significantly and negatively related to attendance, B = -1.0240, $t = -2.57$, $p = 0.005$; the proportion age 31 to 64 is marginally and positively related to attendance, B = 0.5360, $t = 1.61$, $p = 0.054$; and, the proportion age 65 and over was not related to attendance, B = -0.4796, $t = -1.19$, $p = 0.118$. Given the average age distribution of respondents in the observed studies, the expected attendance rate is 43.9%. A one percent increase in the proportion of each age group would result in an increase of 0.1% for 13 to 17 year-olds and 0.13% for 31 to 64 year-olds and a decrease of -0.25% for 18 to 30 year-olds.[105] If the proportion of each of these age groups were to increase by 5%, the attendance rates would be expected to also increase by 0.6% and 0.66% for 13 to 17 year-olds and 31 to 64 year-olds, respectively, but to decrease by -1.26% for 18 to 30 year-olds.

To test Hypothesis 7, we use the meta-regression results to estimate the attendance rate if the entire sample were composed of each respective age group, and then construct confidence intervals around those figures. If the confidence intervals do not overlap, then the differences between age groups are significant. Consistent with Hypothesis 7, the 18- to 30-year-old group

---

[104] R^2 was equal to 0.29 in this analysis, indicating that 29% of the between-studies variation was accounted for by the variables in this model.

[105] The coefficient for those 65 and over was not significant, and so it is assumed that the attendance rate will not fluctuate meaningful in response to changes in the proportion of 65-and-older participants.

had the lowest expected attendance rate of 17.7% (95% CI = 8.9% to 32.0%). As expected, the 13- to 17-year-old age group had a higher rate of attendance at 49.3% (95% CI = 32.7% to 66.1%). Also as expected, the 31- to 64-year-old age group had an even higher expected attendance rate at 50.5% (95% CI = 34.6% to 66.3%), but the 65-and-older age group had a much lower attendance rate than expected at just 27.0% (95% CI = 14.3% to 45.0%). The differences between the 18- to 30-year-old group and both the 13- to 17- and 31- to 64-year-old groups were significant, but no other differences were significant. With one exception, then, the results followed the expected pattern of attendance throughout the lifespan, and so Hypothesis 7 is partially supported.

Hypothesis 8. The percentage of married participants in a study will be positively related to religious service attendance estimates.

It is expected that higher proportions of married participants will be associated with higher attendance rates. A total of 107 study observations are available from a number of data sources including the ACLS, ANES, AHTUS, Baylor Religion Survey, BWHS, original and new Gallup items, original GSS, the GSS branching item, NSBA and the Religious Landscape Survey. Preliminary analyses indicated, however, that three observations were exerting undue influence on the results given studentized residuals of 3.75 (NSBA), 3.00 (BWHS) and 2.69 (2002 ANES) and were removed from the analysis, leaving 104 observations.

The results of the random effects meta-regression are presented in Table 19. The full model was significant, $F(5,97) = 42.93$, $p < .001$,[106] and the results support Hypothesis 8. After controlling for study year and efforts to address the three attendance item criticisms, the proportion of a sample that is married was found to be positively and significantly related to

---

[106] R^2 was equal to 0.74 in this analysis, indicating that 74% of the between-studies variation was accounted for by the variables in this model.

attendance, B = 0.8583, $t = 3.76$, $p < 0.001$. Interpreting these results, the attendance rate for a sample with a marriage rate equal to 59.1%[107] is expected to be 40.9%.[108] For each 1% increase in the marriage rate, attendance is expected to increase by 0.21%. Thus, if the proportion of married participants increased by 5%, the expected attendance rate would increase to 42.0%. If a sample were composed entirely of married participants, the expected attendance rate would increase to 49.6%, whereas a sample composed of all non-married participants would yield an expected attendance rate of 29.4%, a difference of 20.2%! This finding supports Hypothesis 8, but the magnitude of the difference is much bigger than the literature suggests.

Hypothesis 9. The percentage of participants who are parents of school-aged children will be positively related to religious service attendance estimates.

It is expected that samples with higher proportions of participants with school-age children will yield higher attendance rates. Just 56 observations are available to test this hypothesis. The available data sources include the 2000 ANES, AHTUS, the original GSS and the GSS branching item. Some of the data sources did not distinguish between children aged 5 to 12 and those aged 13 to 17 and so the available data can only speak to whether there is a relationship between attendance rates and sample proportions of parents of children age 5 to 17.

The results of the random effects meta-regression model are presented in Table 20. The full model was significant, $F(5,49) = 69.84$, $p < .001$,[109] and the results support Hypothesis 9. After controlling for study year and efforts to address the three attendance item criticisms, the sample proportion having school-age children was found to be positively and significantly related to attendance, B = 1.1295, $t = 4.09$, $p < 0.001$. Interpreting the results, the attendance rate for a

---

[107] This proportion represents the arithmetic mean participation rate across the sampled observations available for this analysis.

[108] This estimate was obtained by using parameter values equal to the average study characteristic. The lone exception was Year, 2015, which was selected because it is the most proximal.

[109] R^2 was equal to 92% in this analysis, but this may have been an anomaly (see Borenstein et al., 2015).

sample of similar studies composed of roughly one-fourth (23.5%[110]) parents of school-age children is expected to be 34.8%.[111] This baseline attendance rate is lower than observed for other demographics, but that is likely the result of a smaller sample of studies that include 23 time-use observations. Using this sample as context, it was found that for each 1% increase in the school-age parenting rate, attendance was expected to increase by 0.26%. Thus, if the proportion of school-age parents increased by 5%, the expected attendance rate would increase to 36.1%. If a sample were composed entirely of parents of school-age children, the expected attendance rate would increase to 55.9%, whereas a sample composed of non-parents and non-school-age parents would be expected to attend at a rate equal to 26.8%, a difference of 29.1%! This finding supports Hypothesis 9.

Hypothesis 10. Studies that utilize samples of relatively well educated participants will yield relatively high attendance estimates.

It is expected that education will be positively associated with attendance. One hundred forty-nine study observations are available from the ACLS, AHTUS, ANES, Baylor Religion Survey, new and original Gallup, original GSS, GSS branching, NELS Dropout surveys, NSBA and NSYR. Preliminary analyses indicated that eight observations were exerting undue influence on the results given its studentized residual of 3.35 (2008 ANES), 3.24 (2002 ANES), 2.99 (2000 ANES), 2.90 (NSBA), 2.87 (2004 ANES), -2.74 (1990 NELS High School Dropouts), 2.62 (1966 AHTUS) and -2.62 (1992 NELS High School Dropouts) and were removed, leaving 141 observations available for the analysis.

[110] This proportion represents the arithmetic mean participation rate across the sampled observations available for this analysis.
[111] This estimate was obtained by using parameter values equal to the average study characteristic. The lone exception was Year, 2015, which was selected because it is the most proximal.

The results of the random effects meta-regression model are presented in Table 21. The full model was significant, $F(5,134) = 66.74$, $p < .001$,[112] but the results fail to support Hypothesis 10. After controlling for study year and efforts to address the three attendance item criticisms, mean education was significantly, but *negatively* related to attendance rates, B = -0.0325, $t = -2.45$, $p = 0.008$. Interpreting these results, the attendance rate for a sample averaging 13 years of education[113] is equal to 44.7%.[114] For every additional year of education, attendance decreases by an average of 0.80%. Thus, those with 12 years of education (High School Diploma) can be expected to attend at a rate equal to 45.5%, while those with 16 years of education (Bachelor's degree) are expected to attend at rate equal to 42.3%. This latter difference between high school grads and college grads is 3.2%, with the former group attending more frequently. These results are contrary to what was expected, and, therefore do not support Hypothesis 10.

Hypothesis 11. Studies that utilize samples of relatively wealthy participants will yield relatively high attendance estimates.

Recall that a sample's median income was classified into one of five categories: Less than \$25,000, \$25,000 to \$49,999, \$50,000 to \$74,999, \$75,000 to \$99,999 and \$100,000 or more. As might be expected with nationally representative samples, there were no studies reporting a median sample income of \$100,000 or more, and there were only two studies where the median income was less than \$25,000 (i.e., the ACLS Wave II and NSBA). Accordingly, these categories were omitted from the analysis, and two dummy variables were used to represent the three middle-class income categories, with the \$50,000 to \$74,999 group serving as the referent.

[112] R^2 was equal to 0.78 in this analysis, indicating that 78% of the between-studies variation was accounted for by the variables in this model.
[113] This proportion represents the arithmetic mean across the sampled observations available for this analysis.
[114] This estimate was obtained by using parameter values equal to the average study characteristic. The lone exception was Year, 2015, which was selected because it is the most proximal.

The data sources available for this analysis include the following: ACLS, AHTUS, ANES, Baylor Religion Survey, original and new Gallup, original GSS, GSS branching and the Religion Landscape Survey. These data sources contribute 122 study observations, although two were identified as having an undue influence on the results as evidenced by their large studentized residuals of 2.88 (1963 original Gallup that omits "Synagogue") and -2.67 (2003 AHTUS) and were removed, leaving 120 study observations.

The results of the random-effects meta-regression model are presented in Table 22. The model containing Year and three dummy variables representing time-frame length was significant, $F(6,113) = 38.45$, $p < .001$,[115] and supports Hypothesis 11. After controlling for study year and efforts to address the three attendance item criticisms, it was found that study observations reporting a median income of $50,000 to $74,999 yielded significantly higher attendance rates than observations where the median income was $25,000 to $49,999, B = -0.1108, $t = -3.44$, $p < 0.001$, but significantly lower than observations where the median income was $75,000 to $99,999, B = 0.2861, $t = 4.01$, $p < 0.001$. Interpreting the results, it appears that study observations reporting a median income between $25,000 and $49,999 yield an average attendance rate of 37.7%. This rate increases to 40.4% for studies reporting a sample median income between $50,000 and $74,999, and to 47.4% for studies reporting a sample median income above $75,000.[116] The difference in attendance rates between the lowest income category and the highest category was 9.7%.

[115] The R^2 for this model was estimated at 0.74, indicating that 74% of the between-studies variation was accounted for by the variables in this model.
[116] These attendance estimates were obtained by using parameter values equal to the average study characteristic. The lone exception was Year, 2015, which was selected because it is the most proximal.

Hypothesis 12. Studies that utilize samples of primarily employed persons will yield relatively low attendance rates.

It is expected that samples with higher proportions of employed participants will yield lower attendance rates. A total of 94 study observations are available from the following data sources: ACLS, ANES, AHTUS, original and new Gallup items, the original GSS and the NSBA. Preliminary analyses indicated that two observations were exerting undue influence on the results given their studentized residuals of 2.75 (1980 original Gallup) and 3.59 (NSBA) and were removed, leaving 92 observations.

The results of the random effects meta-regression model are presented in Table 23. The full model was significant, $F(5,85) = 52.45$, $p < .001$,[117] but the results fail to support Hypothesis 12. After controlling for study year and efforts to address the three attendance item criticisms, the sample proportion employed was found to be positively and significantly related to attendance, B = 0.7234, $t = 4.15$, $p < 0.001$. Interpreting these results, the attendance rate for a sample with an employment rate of 62.3%[118] is expected to be 41.0%.[119] For each 1% increase in the employment rate, attendance is expected to increase by 0.18%. Thus, if the proportion of employed participants increased by 5%, the expected attendance rate would be 41.9%. If a sample were composed entirely of employed participants, the expected attendance rate would increase to 47.8%, whereas a sample composed of all unemployed participants would yield an expected attendance rate of 30.7%, a difference of 17.1%! This finding is opposite of what was expected, as stated in Hypothesis 12. Moreover, the magnitude of the difference is large suggesting that this finding is not trivial.

---

[117] The R^2 for this model was estimated at 0.82, indicating that 82% of the between-studies variation was accounted for by the variables in this model.
[118] This proportion represents the arithmetic mean rate across the sampled observations available for this analysis.
[119] This estimate was obtained by using parameter values equal to the average study characteristic. The lone exception was Year, 2015, which was selected because it is the most proximal.

Hypothesis 12a. Studies that utilize samples of primarily employed males will yield relatively high attendance rates. Hypothesis 12b. Studies that utilize samples of primarily employed females will yield relatively low attendance rates.

Hypotheses 12a and 12b can be addressed simultaneously. To address these hypotheses, the proportion female variable and proportion employed variables were multiplied to produce the interaction term for proportion female by proportion employed. The proportion female, proportion employed and proportion female by proportion employed variables were then entered into a meta-regression model that also contained controls for study year and indicators representing attempts to address the three gold standard item criticisms. The data sources available for the model include the ACLS, AHTUS, ANES, original GSS, GSS branching item, original and new Gallup and the NSBA. These data sources yielded 93 study observations. Two observations were identified as exerting an undue influence on the results given their large studentized residual values of 3.67 (NSBA) and -2.63 (ACLS Wave 3) and were removed, leaving 91 observations.

The results of the random effects meta-regression model are presented in Table 24 and are depicted in Figure 9. The full model was significant, $F(7,83) = 40.50$, $p < .001$,[120] and the coefficients for both proportion female, B = -11.3684, $t = -3.70$, $p < 0.001$, and proportion employed, B = -8.9717, $t = -3.65$, $p < 0.001$, were significant. The interaction term for proportion female by proportion employed was also significant, B = 18.0089, $t = 3.92$, $p < 0.001$, but in the opposite direction of what was expected.

Inserting a combination of low, medium and high values for both proportion female and proportion employed into the regression equation allows for a visualization of the interaction

[120] The R^2 for this model was estimated at 0.83, indicating that 83% of the between-studies variation was accounted for by the variables in this model.

(see Figure 9).[121] For samples with the average gender composition (i.e., 53% female, 47% male), attendance increases linearly with the employment rate, starting at 40.0% for samples with low employment rates to 45.8% for samples with high employment rates. For primarily male samples, however, attendance is at its highest (49.8%) when the employment rate is at its lowest. As the employment rate increases, the attendance rate for primarily male samples declines markedly to 37.2%. This finding is opposite of the expectation stated in Hypothesis 12a. For primarily female samples, on the other hand, attendance is at its lowest (30.9%) when the employment rate is at its lowest, but female attendance reaches its zenith (54.8%) when the employment rate is highest. Here again, this finding is opposite of the expectation stated in Hypothesis 12b. Thus, neither Hypothesis 12a nor 12b was supported.

Fully Specified Socio-Demographic Model

With a few exceptions, the analyses undertaken here have primarily involved examining the relationships between attendance and one socio-demographic variable at a time. This approach leaves open the possibility that spurious relationships have been found where none existed, or that relationships have been obscured (via the lack of variance suppressing effects). This approach also does not adequately identify the variables that are most strongly associated with the attendance rate. Accordingly, the socio-demographics tested in Hypotheses 5 through 12 were included in a fully specified multivariate model.

Because the socio-demographics form conceptually distinct sets of predictors, and because some of the individual predictors have fewer observations associated with them than others, this analysis examines these predictors in three blocks before arriving at a final, fully-specified model. The three blocks of predictors are as follows: (1) Gender (proportion female), race and

[121] The mean, along with 2 SD below and above the mean were used to determine Low, Medium and High for graphing purposes.

ethnicity (proportion African American, Asian, Hispanic and White) and age (proportion 13-17, 18-30, 31-64 and 65 and over); (2) familial Status (proportion married, proportion with school-age children); and, (3) socioeconomic status (mean education, proportion employed and median income). The socio-demographics that emerge as significant predictors of attendance within each block were retained for the fully specified model.

The results from the random effects meta-regression model for block one can be found in Table 25. There were 77 observations available for this analysis after removing 10 outlier observations. After controlling for study year and efforts to address the three attendance item criticisms, proportion female, B = 0.5128, $t = 1.82$, $p = 0.0365$, proportion African American, B = 2.7447, $t = 5.61$, $p < 0.001$, proportion White, B = 2.3166, $t = 4.56$, $p < 0.001$, proportion age 18 to 30, B = -0.4338, $t = -2.24$, $p = 0.014$ and proportion age 31 to 64, B = 0.7514, $t = 2.35$, $p = 0.011$,[122] all emerged as significant predictors and were retained for the full model.[123]

The results from the random effects meta-regression model for block two can be found in Table 25. There were 52 observations available for this analysis. After controlling for study year and efforts to address the three attendance item criticisms, proportion married, B = 1.3856, $t = 6.67$, $p < 0.001$, emerged as a significant predictor and was retained for the full model.[124]

The results from the random effects meta-regression for block three can be found in Table 25. There were 81 observations available for this analysis after removing 1 outlier observation. After controlling for study year and efforts to address the three attendance item criticisms, mean education, B = -0.1139, $t = -4.36$, $p < 0.0001$, proportion employed, B = 0.4771, $t = 2.19$, $p = 0.016$, median income \$25,000-\$49,999, B = -0.1831, $t = -4.37$, $p < 0.001$ and median income

[122] Age 65 and over was dropped from the current block due to suspected over-specification of the age variable (i.e., assuming that the four age indicators total to 100% of a sample, the fourth indicator becomes unnecessary if you know the values of the first three).
[123] $F(12,64) = 176.48$, $p < .001$
[124] $F(6,45) = 161.01$, $p < .001$

$75,000-$99,999, B = 0.0969, $t = 1.42$, $p = 0.0799$, all emerged as significant or marginally significant predictors and were retained for the full model.[125]

The fully specified socio-demographic model included controls for year of study and efforts to address the three attendance item criticisms, along with proportion female, proportion African American, proportion White, proportion in each of three age groups (13-17, 18-30, 31-64), proportion married, mean education, proportion employed and median sample income. There were 74 observations available for this analysis after removing 3 outlier observations.

The results of the fully specified model are provided in Table 25, and indicate that five socio-demographics emerged as significant predictors of religious service attendance.[126] The proportion of African Americans sampled was positively related to attendance, B = 2.3438, $t = 6.67$, $p < 0.001$, as was the proportion of White participants, B = 1.6971, $t = 6.70$, $p < 0.001$, proportion married, B = 1.0193, $t = 3.29$, $p < 0.001$ and, surprisingly, mean years of education, B = 0.0808, $t = 2.27$, $p = 0.013$. This latter finding is opposite of what was found in the bivariate case and even in the block-regression case. If this finding is reliable, then it would suggest that one of the other variables in the model suppressed the variance in education that was driving the initial negative association. Finally, the sample proportion age 18 to 30 was negatively related to attendance, B = -0.6096, $t = -1.82$, $p = 0.037$.

A quick calculation of the standardized coefficients indicates that the two race and ethnicity variables are the strongest predictors, with proportion White (*Beta* = 6.70) and proportion African American (*Beta* = 6.67) emerging as roughly equivalent in their influence on attendance. Proportion married (*Beta* = 3.29) is the next strongest predictor in the model, and is followed by mean years of education (*Beta* = 2.27) and proportion aged 18 to 30 (*Beta* = -1.82). Surprisingly,

[125] $F(8,72) = 43.15$, $p < .001$
[126] $F(14,59) = 78.19$, $p < .001$

gender did not emerge as a significant predictor of attendance in the final model despite it being one of the most consistent predictors of attendance in the literature.

Importantly, the fully specified model also allows for an estimation of the impact that each of the indicators for ambiguous wording, ambiguous time-frame and social desirability have on attendance rates. An examination of the full model in Table 25 indicates that efforts to reduce ambiguous wording have not significantly impacted the attendance rate, B = -0.0157, $t = -0.15$, $p = 0.442$. The coefficients associated with efforts to reduce the impact of ambiguously specified time frames, B = -0.1403, $t = -4.12$, $p < 0.001$, and socially desirable responding B = -0.4928, $t = -4.91$, $p < 0.001$, however, are significant. In this model, efforts to minimize or eliminate socially desirable responding (*Beta* = -4.91) had a slightly greater impact on the attendance rate than efforts to minimize or eliminate the ambiguously specified time-frame (*Beta* = -4.11). Putting these results into more practical terms, the average religious service attendance rate, after controlling for year of study and the socio-demographic variables, was 36.7%. Without the items and methods designed to control for ambiguous wording, the projected attendance rate would have been just one-tenth of one percent higher at 36.8%. Without the items and methods designed to control for ambiguously specified time-frames, however, the projected attendance rate would be 38.1%, which is consistent with an inflation factor of 1.04 after partialling out the influence of the other variables in the model. Finally, without the items and methods designed to control for socially desirable responding, the projected attendance rate is 39.5%, which is consistent with an inflation factor of 1.07 after partialling out the influence of the other variables in the model.

Estimating the Prevalence of Attendance in America

The previous analysis demonstrated that the estimated prevalence of attendance in America is 36.7%. But, this estimate did not include the count-based attendance data,[127] and it reflects the rate across *all* data sources after controlling for all efforts to minimize the gold standard attendance item criticisms. But, the findings thus far have been very clear that the attendance rate varies greatly across items and methodologies, and a grand mean across all sources is not meaningful. Thus, the final analysis will estimate the prevalence of attendance in America as estimated by each of four major types of items and methodologies. This will allow future researchers to better understand the attendance rates that can be expected from each item and method, and to place their research into context.

The four types of items and methods that were analyzed include: (1) the gold standard items, (2) items measuring attendance within the last week (or last 7 days),[128] (3) the time-use items and (4) the count-based method. Where available, the significant socio-demographic predictors identified in the previous analysis, along with study year, were included as controls. Each of the four models are discussed below, with the results also presented in Table 26 and Figure 10.

Prevalence of Attendance According to the Gold Standard Items

Four data sources (original GSS, new Gallup, new Gallup that mentions "Mosque" and Waves 1 through 5 of the ACLS) provided 60 study observations for the gold standard attendance item. Thirteen observations did not include the full set of socio-demographic covariates, and were removed, leaving 47 observations. Preliminary analyses indicated that two

[127] Socio-demographics are not available for the count-based method so these data could not be included in the models.

[128] A version of this analysis was conducted to test Hypothesis 1a, but that analysis did not include controls for socio-demographic variation.

covariates (Age 18-30, Age 31-64) were not significantly associated with the gold standard attendance rates and so were removed from the model to improve statistical efficiency.[129]

The results from the meta-regression model are provided in Table 26. The full model was significant, $F(5,41) = 43.13$, $p < .001$,[130] and each of the covariates were significantly associated with attendance in the expected direction. The coefficient for year of study was negatively related to attendance, B = 2.3438, $t = 6.67$, $p < 0.001$, indicating that attendance has slowly, but steadily declined (by 0.44% per year) over the 43 years for which data are available from 43.3% in 1972 to 41.4% in 2015 (see Figure 10).[131] Thus, according to the gold standard items, the current prevalence of attendance in America for any given week is 41.4%.[132]

Prevalence of Attendance According to Items Measuring Attendance in the Past Week

Three data sources (original Gallup, original Gallup omitting mention of "Synagogue" and the GSS branching item[133]) provided 61 study observations for items measuring attendance in the past week. The covariate for marital status was missing data for 36 observations and was removed from the model, along with three additional covariates that were found to be non-significantly related to attendance in the past week (proportion African American, proportion White and mean education), to improve statistical power and efficiency. Fourteen observations

[129] The covariate-to-observation ratio improved from 1:7 to 1:9 after removing the non-significant covariates.
[130] The R^2 for this model was estimated at 0.92, indicating that 92% of the between-studies variation was accounted for by the variables in this model. Given that this is an extremely large figure, however, it may also have been an anomaly (see Borenstein et al., 2015).
[131] These rates assume that all covariates are at their mean level across all observations. Using the same model assumptions and the year 2015, the prediction interval (i.e., the range within which 95% of all future estimates of attendance from studies using the gold standard items will fall within) ranges from 32.0% to 50.8%.
[132] Because the attendance metric was transformed to reflect attendance in any given week the results from this analysis may deviate with research that use these items, but define "regular" attendance differently.
[133] The initial question in the GSS branching item was used here. The initial question asks about attendance in the past 7 days, which is identical to the other original Gallup items included in the analysis. The four follow-up branching questions help clarify what respondents meant by their initial answer, but those data are not included here because they address ambiguous wording as well, and maximum comparability to other items measuring attendance in the "past week" was the goal.

did not include data on the remaining covariates (proportion age 18-30, 31-64) and were removed, leaving 47 observations for the analysis.[134]

The results from the meta-regression model are provided in Table 26. The full model was significant, $F(3,41) = 21.28$, $p < .001$.[135] The coefficient for proportion age 31 to 64 was significant and in the expected direction, B = 2.3578, $t = 5.99$, $p < 0.001$, while the coefficient for proportion age 18 to 30 approached significance in the expected direction, B = -0.3841, $t = -1.60$, $p = 0.058$. The coefficient for year of study was not significant, B = 0.0014, $t = 1.05$, $p = 0.299$, but was retained in the model for consistency of reporting across models. According to this model, attendance has slowly, but not significantly increased (by 0.034% per year) over the 65 years for which data are available from 40.9% in 1950 to 43.1% in 2015 (see Figure 10).[136] Thus, according to the items measuring attendance in the past week, the current prevalence of attendance in America for any given week is 43.1%. This rate is 1.9% *higher* than the gold standard estimate of 41.4%. This is surprising given that items measuring attendance within the past week address the ambiguous time-frame problem associated with the gold standard items, which should lead to a lower rate of attendance.

Prevalence of Attendance According to the Time-Use Items

Three data sources (AHTUS, ATUS and Presser & Stinson, 1998) provided 30 study observations for the time-use item. Five of the six covariates (proportion African American, proportion White, proportion married, proportion age 18-30 and mean education) were found to be non-significantly related to attendance, leaving only year of study and proportion age 31 to

[134] As a result of the model trimming, the covariate-to-observation ratio improved from 1:7 to 1:16.
[135] The R^2 for this model was estimated at 0.71, indicating that 71% of the between-studies variation was accounted for by the variables in this model.
[136] These rates assume that all covariates are at their mean level across all observations. Using the same model assumptions and the year 2015, the prediction interval ranges from 27.3% to 58.9%.

64. Seven observations did not include age data and were removed, leaving 23 observations for the analysis.[137]

The results from the meta-regression model are provided in Table 26. The full model was significant, $F(2,20) = 6.60$, $p < .001$.[138] The covariate, age 31 to 64, was significantly associated with attendance in the expected direction. The coefficient for year of study was negatively related to attendance, B = 0.0037, $t = -2.51$, $p = 0.010$, indicating that attendance has slowly, but steadily declined (by 0.08% per year) over the 46 years for which data are available from 31.4% in 1966 to 27.8% in 2012.[139] Thus, according to the time-use items, the most current prevalence estimate of attendance in America is 27.8%. If this rate were projected to the year 2015 (27.6%), it would be 13.8% lower than the gold standard estimate of 41.4%, which translates to an inflation factor of 1.50 (see Figure 10).

Prevalence of Attendance According to the Count-Based Methodology

Three data sources (Chaves & Cavendish, 1994; Hadaway & Marler, 2005; Hadaway et al., 1993) provided 11 study observations for the count-based method. The only covariate available for all 11 observations is year of study, which was not significantly associated with the count-based attendance estimates, but was retained in the model for consistency of reporting.

The results from the meta-regression model are provided in Table 26. The full model was not significant, $F(1,9) = 0.13$, $p = .726$,[140] reflecting the lack of a significant covariate. According to this model, attendance has slowly, but not significantly, decreased (0.06% per year) from 23.9% in 1990 to 22.7% in 2011.[141] Thus, according to the count-based method, the most

[137] As a result of the model trimming, the covariate-to-observation ratio improved from 1:4.3 to 1:11.5.
[138] R^2 = 100%, but this is likely an anomaly associated with meta-regression (see Borenstein et al., 2015)
[139] These rates assume that all covariates are at their mean level across all observations. The prediction interval was incalculable because T was estimated to be 0.00 for this model.
[140] R^2 = 0.0%, but this is likely an anomaly associated with meta-regression (see Borenstein et al., 2015)
[141] Using the year 2015, the prediction interval estimated by the model ranges from 0.0% to 61.0%.

current prevalence estimate of attendance in America is 22.7%. If this rate were projected to the year 2015 (22.5%), it would be 18.7% lower than the gold standard estimate of 41.4%, which translates to an inflation factor of 1.82 (see Figure 10).

Summary of the Prevalence of Attendance Findings

The findings from the four most recent models confirm that there is a great deal of variability in the estimation of the prevalence of attendance in America. As of 2015, the gold standard items were estimating that 41.4% of Americans attend religious services on a weekly basis. The items measuring attendance in the past week yielded a similar, but slightly higher estimate of 43.1%. As expected, the time-use items and the count-based method both produced drastically different estimates of the prevalence of attendance. The time-use item estimated that 27.8% of Americans attend religious services on a weekly basis, which is 50% below the gold standard rate. The count-based method goes further, estimating that just 22.7% of Americans attend religious services on any given Sunday, which is 84% below the gold standard rate.

Furthermore, three of the four data sources indicate that the prevalence of attendance in America has been slowly, but steadily declining for the last 65 years. The exception to this is attendance in the past week, which showed a slow, but non-significant, increase over time.

In sum, the answer to the question, "What is the prevalence of attendance in America?" depends upon the method of estimation. The gold standard items and items measuring attendance in the past week both indicate a prevalence estimate of just over 40%. The "new" methods of estimating attendance, however, indicate that the rate is closer to 25%. And, three of the four data sources indicate that attendance is declining. Together, these data challenge the gold standard rate of attendance, and suggest that the true rate has yet to be definitively identified.

CHAPTER V

DISCUSSION

The questions that are central to this study stem from a “gold standard” that was established and promulgated through the literature for decades. The gold standard states that over 40% of Americans attend religious services on a regular basis and have done so since the 1930s when data were first collected. As the religious research literature began to take off, however, several criticisms were levied against the method by which attendance was measured. These criticisms, which centered on ambiguous item wording and socially desirable responding, suggested that the gold standard estimates of attendance were inflated. Several lines of research began investigating the veracity of these claims. The results were varied, with some efforts producing little change and others producing drastically different estimates of attendance. This study set out to addresses this variability by gathering, cumulating and synthesizing the literature on attendance using meta-analytic methodology. This study also attempted to identify other sources of variance—namely that from the socio-demographics characteristics of study samples—and to assess their impact, if any, on attendance rates. Finally, this study set out to address the question of just how many Americans attend religious services by cumulating data from hundreds of thousands of Americans and hundreds of studies using nationally representative samples. The answers rendered to these questions are summarized below.

Efforts to Overcome “Gold Standard” Attendance Item Criticisms and Their Effects

The three main criticisms levied against the gold standard attendance items are that the they are characterized by ambiguous wording (i.e., what is included and excluded in the definition of “religious services?”), an ambiguously specified time-frame (i.e., should the respondent report their most recent behavior or their typical behavior over the last year, five years or longer?) and socially desirable responding (i.e., data were collected via methods likely to elicit desirable responses and religious behavior, because it is important to many, is a desirable response). To test the veracity of these criticisms, researchers began looking at data from items that either had built-in wording modifications or were designed with wording modifications to address one or more of the criticisms (e.g., the ambiguous time frame is made clear in the Gallup Poll question, "Did you, yourself, attended church or synagogue in the last seven days?"). Other researchers set out to develop new data collection methodologies that would address the ambiguity and social desirability problems (e.g., the time-use and count-based methods). Six hypotheses were developed in this study to test the effects of these efforts, and the results are summarized and discussed below.

Summary of Hypotheses and Results Related to the Ambiguous Time-Frame

Hypotheses 1 and 1a were generated to test whether efforts to address the ambiguous time frame problem had an impact on the religious service attendance rate, and if so, which efforts yielded greater effects. The results provided partial support for these hypotheses, indicating that some efforts had no impact, while others had a sizeable impact. In the initial analysis, where all items and methods using a time-frame were lumped together and compared to the gold standard, the effect of including a time-frame was to reduce the attendance rate by 6.4%, which translates to an inflation rate of 17% due to an ambiguously specified time-frame.

In the follow-up analysis, each item and method was analyzed for its individual impact. The attendance rate yielded by items measuring attendance in the past week was found to be not significantly different from the gold standard rate. Although the effect of the time-frame was in the right direction (-0.8%), the difference, if real, was too small to be detected in this study, and therefore, does not support the expectation that the specification of a time-frame will reduce the attendance rate. Moreover, in later analyses when the item asking about attendance in the last week was isolated in a model with the gold standard item, the difference was actually in the opposite direction, with the latter item yielding a lower attendance rate (see Table 26 and Figure 10).

To rule out the possibility that the non-significant difference between the gold standard items and the items measuring attendance in the past week was not an artifact created by the harmonization process, the gold standard "regular" rates of attendance that would have been obtained without harmonization (M = 0.439, SD = 0.04) were compared to the rates obtained after harmonization (M = 0.436, SD = 0.04). There was virtually no difference between the two sets of rates, $t(59) = 1.35$, $p = .18$, which indicates that the harmonization process was not responsible for the non-significant difference between the gold standard rates and those associated with the items measuring attendance in the past week.

The items referring respondents to yesterday (i.e., time-use items) and those where the researcher removed the respondent from the process all together in lieu of conducting counts, however, did produce markedly lower attendance rates. Compared to the gold standard rate of 43.1%, the time-use item and count-based methods yielded estimates of 27.7% and 23.1%, respectively. These differences equate to inflation rates of 1.56 (56%) for the time-use item and 1.87 (87%) for the count-based method.

Taken together, these results indicate that the use of a time-frame may matter. Attendance rates are reduced by 17% when using any combination of items and methods designed to specify a time-frame. Yet, items that incorporate a time frame into the wording of a question (i.e., in the last seven days) do not seem to produce rates that are any different than the gold standard rates. The real differences stem from two methodologies that are very different from what has been used before. In one case, respondents are asked to describe what they did yesterday without the questioner ever mentioning anything about religion. This approach not only provides a time-frame, it also eliminates ambiguous wording, given that the respondent can use their own words to describe what they did yesterday, and it reduces socially desirable responding because the respondent is never told that their religious behavior is a focus of the study. In the other case, the respondent is removed from the situation all together in lieu of manual attendance counts, which eliminate any possibility for ambiguity or social desirability. Thus, the results indicate that the inclusion of a time-frame is associated with a reduction in the attendance rate, but that this difference might also be due to the minimization or elimination of ambiguous wording and social desirability problems.

Summary of the Hypothesis and Results Related to Ambiguous Wording

Hypothesis 2 was generated to test whether efforts to address ambiguous item wording have an impact on the religious service attendance rate, and if so, which efforts yielded greater effects. The results for ambiguous wording essentially mirror those for the ambiguous time-frame. Specifically, when all items and methods attempting to address the ambiguous wording problem are lumped together and compared to the gold standard items, the attendance rate decreases by 15.7%, which translates to an inflation rate of 53% due to ambiguous wording.

When the effects of item wording and methodological changes are teased out, however, the items that attempt to address the ambiguous wording problem by clarifying terminology actually result in a slightly, but not significantly, higher attendance rate than the gold standard items. This finding was unexpected, and could be the result of low statistical power and sampling error. Specifically, just seven study observations were available to address this approach, and three of those observations were atypically high and came from three waves of a single survey. Thus, this study did not provide a good test of whether attempts to address the ambiguous wording problem through language clarification are effective.

As evidenced in the analyses addressing the ambiguous time-frame problem, however, the time-use and count-based methods, which both minimize or eliminate the ambiguous wording problem, were associated with significant reductions in the attendance rate. In fact, their combined effect reduced the attendance estimate by 17.3%, which translates to an inflation rate of 68%. As was the case before, however, it is unclear whether this difference is due to the minimization of ambiguous wording, or to the minimization of other factors, such as an ambiguously specified time-frame or socially desirable responding. Thus, the results provide partial support for the idea that items and methods designed to minimize ambiguous wording yield lower attendance rates, but it is not clear whether the removal of ambiguous wording is the prime reason why the attendance rate drops.

Summary of the Hypothesis and Results Related to Social Desirability

Hypothesis 3 was generated to test whether efforts to address socially desirable responding have an impact on the religious service attendance rate, and if so, which efforts yielded greater effects. The results are similar to those generated for the analysis of the ambiguous time frame and wording problems. Specifically, when all items and methods attempting to address socially desirable responding are lumped together and compared to the gold standard items, the attendance rate decreases by 20.1%, which translates to an inflation rate of 85% due to socially desirable responding.

When the effects of item wording and methodological changes are teased out, the former are once again found ineffective, while the latter are found to be effective. The result associated with item wording changes, however, could have been the result of low statistical power given that just six study observations were available to address this approach. Thus, this study did not provide a good test of whether attempts to address the social desirability problem through item wording modifications are effective.

The time-use and count-based methods, which were initially touted as methods designed to reduce social desirability, were able to reduce the attendance rate. Their combined effect reduced the attendance estimate by 17.3%, which translates to an inflation rate of 64% due to social desirability. As was the case before, however, it is unclear whether this difference is due to the minimization of the ambiguous time-frame, ambiguous wording or socially desirable responding. Thus, the results provide partial support for the idea that items and methods designed to minimize the ambiguous wording problem yield lower attendance rates, but it is not clear whether it is the minimization of socially desirable responding or the other two problems that is responsible for the lower attendance rate.

Summary of the Hypothesis and Results Related to Determining Whether Ambiguous Wording or Social Desirability Impacts the Attendance Rate More

For hypothesis 3a, efforts were made to identify data sources that impacted only one of the three gold standard attendance item problems so that independent comparisons could be made. Unfortunately, there were only a few data sources available for each problem. With regard to study observations, there were 60 available for the time-frame problem, but only 8 and 6, respectively for both the ambiguous wording problem and social desirability problems. Thus, statistical power was likely too low for this analysis, which found only one significant difference (i.e., the data for items attempting to address the ambiguous wording problem yielded significantly higher rates of attendance than the items addressing the other two problems), and even that was more likely the result of sampling error (i.e., 3 of 8 observations provided abnormally high attendance rates and were from three waves of the same survey) than a systematic effect. Thus, while this research question is valuable to ask, the available data were too limited to provide a quality answer.

Summary of the Effects of Items and Methods Used to Minimize and Assess the Impact of the Three Gold Standard Item Problems

The results discussed thus far indicate three main findings: (1) Efforts to reduce the problems associated with the gold standard items through wording modifications have little support to suggest that they are effective; (2) the data available to determine whether wording modification efforts are effective are too limited to provide a definitive answer regarding their efficacy; and, (3) the time-use and count-based methods clearly have an impact on the estimated attendance rate. Together, these two measurement approaches have yielded attendance rates that are between 56% and 87% lower than the gold standard rate. Because these items and methods address each of the gold standard item criticisms, however, it is not clear by which mechanism(s) they are having their effect. Additional data on the effects of wording modifications designed to address one of the three gold standard item criticisms would, at least, shed light on which problem has the greater influence on the attendance rate.

Identifying Socio-Demographic Sources of Variation in Attendance

Measurement and methodology issues are not the only sources of variability in the attendance literature. The literature also suggests that the socio-demographic characteristics of study samples can influence attendance rates. Accordingly, this study tested 10 hypotheses pertaining to the nature and size of the relationships, if any, between 8 socio-demographic sample characteristics and attendance. The results of these tests are summarized below.

Summary of the Hypotheses and Results Related to Gender, Race and Ethnicity and Age

The literatures linking attendance with gender, race and ethnicity and age were the most fully developed of any of the socio-demographics. Perhaps the clearest finding in the literature is that women attend more frequently than men, with the difference being around 10%. This study examined the relationship between the proportion of samples that were female and the attendance rates yielded by those samples. The finding was that there is a strong positive relationship between the two. For every 1% increase in the proportion of females sampled, the attendance rate increased by one-third of one percent (0.33%). Furthermore, by projecting the expected attendance rate for all-female samples (57.9%) and all-male samples (25.9%), the disparity between the genders was found to be a whopping 32%! The direction of this finding is consistent with the literature, but the magnitude of the difference is much larger than what has typically been reported. Perhaps the projected rates of attendance for all-female and all-male samples over-estimate the actual difference, but the finding is clear: women attend religious services at a much greater rate than men.

Perhaps the second clearest finding in the literature is that African Americans attend religious services at a greater rate than Whites (and those of other races and ethnicities), with the difference reportedly around 10%. This study examined the relationship between the proportion of African Americans sampled and the attendance rate. This initial approach did not allow for a direct comparison between African Americans and Whites, but it did test whether, and by how much, the African American participation rate in samples is tied to the attendance rate. The findings suggested that for every 1% increase in the proportion of African Americans in a sample, the attendance rate increased by slightly less than one-fifth of one percent (0.18%). Projecting the expected attendance rate for an all-African American sample (60.0%) and an all-non-African American sample (41.8%), the disparity was found to be 18%, which was higher than most differences reported in the literature.

Too little comparative evidence was available for other racial and ethnic groups to develop hypotheses for this study, but by using the study as the unit of analysis instead of the individual, the meta-analytic approach allows for an exploration of the associations between sample participation rates for each racial and ethnic group and attendance. Ample data were available for African Americans, Asian Americans, Hispanic Americans and Whites. Sample proportions of African Americans and Whites were both positively related to the attendance rate, but no relationship was found for either Asian Americans or Hispanic Americans.

The latter finding is surprising given that Hispanic Americans tend to be affiliated with the Catholic Church, and Catholics attend more frequently than most other religious groups (Gallup & Lindsay, 1999; Lazerwitz, 1961). Ancillary analyses indicated that multi-colinearity may have been a problem for the proportion Hispanic variable with the proportion African American and White variables,[142] but even when proportion Hispanic was isolated in a model by itself (along with controls), the association was not as expected (in fact, it was significant in the opposite direction!).[143] Another explanation for this unexpected finding might be that a degree of selection bias exists. If there are differences between Hispanic Americans who can be reached and agree to participate in national surveys than those who cannot be reached or do not agree to participate, and those differences are tied to religious service attendance, then a self-selection threat to internal validity (see Shadish et al., 2002) may account for the non-significant association with attendance in this study. The available evidence from this analysis, however,

[142] Correlations with proportion African American (0.92) and proportion White (0.92) were both extremely high.
[143] $B = -2.43$, $t = -3.02$, $p = 0.002$

suggests that as the proportion of sampled African Americans and Whites increase, so too does the attendance rate.

Finally, the literature on age and attendance is somewhat less well developed than that for gender and race and ethnicity, but the available evidence suggested that age is non-linearly related to attendance. That is, young adults (age 18 – 30) were expected to attend at a lower rate than all other age groups, and to be followed closely by adolescents (age 13 – 17) and middle-aged adults (age 31 – 64), with those in late adulthood (65 and over) attending at the highest rate. With the exception of the latter group, the findings from this study generally supported this view of aging and attendance. Young adults did evidence the lowest levels of attendance (17.7%); and, adolescents (49.3%) did have higher rates than young adults. Middle-age adults (50.5%) attended at a higher rate than adolescents, as expected, but only by a trivial amount, which suggests that the attendance patterns for these two groups could be linked by familial patterns of attendance (i.e., parents and adolescents attend together). The primary surprise in this analysis, however, was that as the proportion of respondents in late adulthood increased, the attendance rate actually decreased. In fact, this group was associated with rate of attendance equal to 27.0%, which is much lower than what has been reported in the literature. This finding raises the possibility that functional disability and morbidity may be related to survey response rates. Specifically, if those who are unable to attend religious services (due to impairment or illness) are also more available to receive phone calls or respond to face-to-face interviews at home, then infrequent elderly attenders may be more likely to be represented in national surveys, which would at least partially explain this finding. Further investigation would be needed to determine if this is a plausible explanation, however.

In sum, samples with relatively high rates of participation by females, African Americans, Whites, adolescents and middle-aged adults should yield high rates of attendance. Conversely, samples with high rates of participation by young adults should be expected to yield low rates of attendance. The finding associated with attendance in late adulthood, however, deserves further scrutiny before a conclusion can be made about what to expect from this population. On the one hand, this study suggested that low rates of attendance should be expected, but the literature clearly suggests that late adulthood is a time of relatively intense religious participation. Thus, the findings reported here for late adulthood should be held tentatively.

Summary of the Hypotheses and Results Related to Familial Status

The results of the familial status analyses support the hypotheses that samples with higher proportions of married persons and persons with school-age children also yield higher attendance rates. In this study, for each one percent increase in the marital and parental status rates, attendance increased by 0.21% and 0.26%, respectively. Furthermore, the projected differences between samples composed entirely of married persons (49.6%) versus those of non-married persons (29.4%), and of parents of school-age children (55.9%) versus those of non-school-age children (26.8%) were large (20.2% and 29.1%, respectively)! The literature on these two outcomes is not as voluminous as it is for gender and race and ethnicity, but these findings are just as clear. Whether people who marry and have children are more likely to be attenders to begin with (e.g., Fergusson et al., 1984), or if getting married and having children changes one's perspective on attending (e.g., Stolzenberg et al., 1995), it is clear that married persons and parents of school-age children are more likely to attend religious services than non-married persons and person not having school-age children.

One caveat associated with the finding for parents of school-age children is that data were relatively rare (i.e., approximately 20% of the study observations obtained contained parental status data) and could not be broken down any further by age. The specificity of the data used in this study was limited to parents of school-age children between the ages of 5 and 17. Given that research differs somewhat on when attendance is at its highest during the school years (e.g., see Argue et al., 1999; Carroll & Roozen, 1975; Stolzenberg et al., 1995), it would be enlightening to break these data down by primary school age-children (age 5 to 12) and secondary school-age children (age 13 to 17) to see if the relationship held or changed across time. For now, however, it appears that the presence of school-age children of any age is associated with higher rates of attendance.

Summary of the Hypotheses and Results Related to Socioeconomic Status

Three outcomes were used to represent socioeconomic status (SES) in this study: median income, mean education and proportion employed. Just one of the three SES indicators related to attendance as expected. Sample median income was positively related to attendance as expected (see Hypothesis 11). Specifically, samples with median incomes between $25,000 and $49,999 yielded an average attendance rate of 37.7%, while samples with median incomes between $50,000 and $74,999 and $75,000 to $99,999 yielded average attendance rates equal to 40.4% and 47.4%, respectively. Because data were limited to these three income categories, it is not known if the relationship would project to samples averaging $25,000 or less or to samples averaging $100,000 or more. But, the evidence here suggests that income, at least middle-class income, is positively related to attendance.

Contrary to prior research and current expectations (see Hypothesis 12), employment status was positively related to attendance. Specifically, the findings indicated that for every percentage increase in the sample employment rate, attendance increased by an average of 0.18%. Samples composed entirely of employed persons were expected to attend at a rate equal to 47.8%, while the unemployed were expected to attend at a rate equal to 30.7%. This is a difference of 17.1%, which is substantial, especially considering that employment status was expected to be negatively related to attendance. It should be mentioned, however, that the published literature on attendance and employment status is limited, meaning that the review of the literature for this study could have been affected by the variability that comes with sampling error and is notorious for plaguing narrative reviews (see Hunter & Schmidt, 2004).

Both deVaus (1984) and Ulbrich and Wallace (1984) also reported employment status by gender interactions, where the relationship between employment status and attendance was negative among women, but positive among men. To explore the veracity of these findings, Hypotheses 12a and 12b were developed. The results, again, revealed just the opposite pattern. That is, a positive association between employment status and attendance was observed among women, while a negative relationship was observed among men. Again, this is opposite of what deVaus (1984) and Ulbrich and Wallace (1984) found. More weight can be given to the current finding, however, given that it is based on 92 study observations. Because both the main effect and the interaction effect turned out different than expected, however, this relationship merits continued attention.

Also unexpectedly, education was *negatively* related to attendance in the initial analysis. Each year of education was associated with a decrease in attendance of -0.80%. Those whose highest degree was a Bachelor's degree (42.3%) attended at a rate 3.2% lower than high school graduates (45.5%). This is surprising given the numerous findings in the literature reporting a positive cross-sectional and longitudinal relationship between education and attendance (e.g., see

Loury, 2004; Muller & Ellison, 2001). Among the few studies in the literature that reported either null or negative relationships, however, are two that use data from the GSS and ACLS, which both figured prominently in the analyses for this study. This might suggest that there is something systematic about the GSS and ACLS that generates the negative education-attendance association. But, this does not seem likely as there is no known reason why the two variables would be related to each other negatively in one set of surveys and positively in another set of surveys when both share the same types of sampling frames and methodologies.

Interestingly, however, when education was included in the fully specified model alongside the other socio-demographic predictors of attendance, the coefficient became positive. This suggests that, after controlling for the other socio-demographics of the sample—such as gender, race and ethnicity, age, marital and parental status, employment status and income—the variance driving the negative bivariate association between education and attendance was accounted for by one or more of the other variables in the model (i.e., this variance was suppressed). For example, if employed males tend to be relatively well educated, and well-educated men are more likely to attend, then it's possible that a positive education-attendance association will not be revealed until gender and employment status are teased out. This set of findings remains perplexing, however, and indicates that the education-attendance association is complex and deserves further attention.

Summary of the Results from the Fully Specified Socio-Demographic Model

Besides the peculiar education finding, the fully specified model revealed some other interesting findings. First, the socio-demographic variables that emerged from the fully-specified model as the strongest predictors of attendance included proportion White and proportion African American. It should be noted that this finding does not mean that Whites attend more, or even equal to, African Americans (in fact, they do not; 59.5% of African Americans and 48.3% of Whites were estimated to attend in any given week).[144] Instead, it means that proportion White covaries with attendance at a slightly greater (although essentially equal) rate than proportion African American. Gender, which was expected to be the strongest predictor of attendance given the large disparity between female and male attenders, was not a significant predictor in the full model. Proportion married emerged as the next strongest predictor, followed by mean education—which again, was positively related to attendance—and then both age variables (13-17 years-old and 31-64 years-old). Besides gender, the other variables that became non-significant include proportion employed and median income. These non-significant associations in the full model do not mean that these variables are unimportant, or that they do not predict attendance. But, it does mean that the other variables in the model were able to better account for the variance in attendance.

One of the primary implications of the findings from the full model is that they provide guidance to researchers who need to identify control variables for their own models of attendance. For example, researchers desiring to assess the relationship between religious service attendance and morbidity or mortality will need to control for variables that covary with both attendance and their outcome of interest. This table provides an indication of the variables to consider including. Because of sampling error, it might be wise to consider all of the socio-demographics included in the final model, but if resources are scarce (including degrees of freedom) then a priority can be placed on the socio-demographics that emerged as significant predictors of attendance in the full model.

[144] This finding comes from the follow-up analyses to Hypothesis 6.

Prevalence of Religious Service Attendance in America

Estimating the prevalence of attendance in America is not a straightforward proposition. Attempts to simply cumulate attendance frequency data across all surveys, items and methods would yield an average rate that would not be meaningful. Because there is so much variability in attendance across the available data sources, the only solution to estimating the prevalence of attendance in America is to provide separate estimates for each of the major items and methods that have been developed.

The final analysis in this study, then, attempted to estimate the prevalence of attendance in America—after controlling for the key socio-demographics identified in the fully specified model—by cumulating data over the life span of the gold standard items, the items measuring attendance in the past week, the time-use items and the count-based method. The ordering of the results for the latter two methods were not surprising. The count-based method produced the lowest attendance estimate, which was 84% lower than the gold standard estimate. This was followed by the time-use item, which produced an estimate that was 50% lower than the gold standard item. The ordering of the first two items, however, was unexpected. The gold standard item actually produced a lower attendance rate in its most recent year (2015) than the item measuring attendance in the last seven days, even though the latter was designed to produce a lower rate via the specification of a time frame.

This latter finding calls into question the effectiveness of item wording changes that are designed to address the ambiguity and social desirability problems. Throughout the analyses reported in this study, attempts to control for attendance "inflation" through the use of wording modifications have been shown to be ineffective. In part, this is because ample data were lacking for items that attempted to clarify ambiguous wording or address social desirability concerns. But, there were ample data for items specifying a time-frame, and as seen from this analysis, these items actually yielded a slightly *higher* rate of attendance than the gold standard items. Furthermore, even with a small number of studies, if the effect size were large enough, the difference would be observable. The count-based method, for example, provided just 11 study observations, but the effect size was so large that it clearly showed up in the analyses. If the true effect of wording modifications is on the order of a couple percentage points, then the small number of studies could explain the non-significant findings here. The literature suggests that this might be the case. For example, the inflation factors roughly estimated in Chapter II were mostly on the order of 10% or under (i.e., an inflation factor of 1.10 or less; see Table 1). Thus, it could be a combination of low power and a small effect size that prevented this study from detecting any wording modification effects. But, one finding is clear: Methodological changes have made a difference!

Study Strengths

This study was able to incorporate data from 26 data sources, 271 study observations and over a million participants. An additional 248 million Americans were represented in 11 count-based attendance estimates, bringing the total number of Americans represented by these data sources to just under 250 million. Together with the use of meta-analytic techniques, these data made it possible to effectively investigate the effects of three main item and method approaches (i.e., attendance in the “past 7 days,” as well as the time-use and count-based methods) designed to address the main criticisms levied against the gold standard items. They also made it possible to identify significant sources of variation in attendance due to the socio-demographic characteristics of samples and to explore the relationships between attendance and socio-demographic factors that have not received a lot of attention in the literature. Finally, they allowed for an estimate of the prevalence of attendance in America across four different items and methods of measuring attendance after controlling for socio-demographic factors and time. Each of these study strengths are touched on briefly below.

This study was able to investigate the effects of modifying item wording to limit the problem of the ambiguous time-frame. The items measuring attendance in the past seven days were represented by some 61 study observations. The attendance rates yielded by these items were compared to 60 study observations representing the gold standard item. Even though the items measuring attendance in the past seven days were found to be ineffective at reducing the attendance rate, just the knowledge of this effect (or lack thereof) is useful. In fact, the results of this study collectively suggest that item wording modifications may not be an effective method of reducing bias introduced by the ambiguous wording or social desirability problems associated with the gold standard items; and, if they are effective, they are only able to produce a small effect, one that was undetectable in this study. This knowledge can help guide future efforts to improve upon the measurement of attendance by discouraging item wording modifications in lieu of more powerful methods of removing bias.

Importantly, this study was able to adequately assess the impact of two powerful methods of removing bias. Specifically, the time-use item was found to produce attendance estimates that were 50% lower than the gold standard estimates, and the count-based method was found to produce estimates that were 84% lower than the gold standard estimates. While this knowledge is becoming increasingly clear in the literature, the use of meta-analytic techniques to estimate the average attendance estimate yielded by these two respective methods will help inform the literature about the average rate of attendance that can be expected from these two methods.

Additionally, by collecting data at the study level, relationships between sample characteristics and attendance rates could be investigated. A number of findings merely confirmed well-established individual-level relationships (e.g., with gender and race and ethnicity). Some findings, however, pertained to relationships that have received relatively little attention. For example, the relationship between parenting a school-age child and attendance, which has been understudied, could be examined using data from over 50 study observations! And, employment status, which had received little prior attention, could be investigated using close to 90 study observations! Moreover, prior expectations regarding some variables like employment status and its interaction with gender could be tested and modified using large amounts of nationally representative data.

The use of meta-regression as a statistical technique to summarize a literature is also relatively new. As mentioned in Chapter III, prior approaches to analyzing the data in this study would have involved using a hierarchical approach where findings for females, for example, would be meta-analyzed, and then compared with the findings for males. Essentially, a separate meta-analysis would be performed for every level of methodological or socio-demographic moderator included in the study. This approach is extremely laborious and time-intensive. Meta-regression simplifies the process in the same way that regression in primary studies is more efficient than the calculation of a series of means and confidence intervals from which to make comparisons. Meta-regression also allows for the simultaneous control of both categorical and continuous covariates, and for the identification of which predictors are the strongest. Recall that this study found that proportion White and African American were the strongest predictors of the attendance rate. And, these were followed by proportion married, education and age. This information can be useful to other researchers for both imagined (e.g., the selection of a key set of predictors or controls for future research models) and unimagined reasons, but it would not have been possible without the use of meta-regression. Thus, the meta-regression approach—which has only recently become fully functional and available via software programs like CMA (Borenstein et al., 2015)—was a very powerful asset in this study that allowed for the identification of a number of relationships and patterns that might not have otherwise been possible.

The available nationally representative data, coupled with the meta-analytic approach were also keys to this study. Because a grand mean attendance rate would not be meaningful, separate estimates were calculated for each of four major types of items and methods of assessing the prevalence of attendance. Importantly, both year of study and key socio-demographic characteristics were held constant in these analyses, thereby making the different prevalence estimates more comparable. Taken together, these characteristics allowed this study to estimate the prevalence of attendance in America while controlling for factors that might otherwise bias these estimates. Furthermore, because the estimates are relatively free of bias, they can be used as a comparison for future studies attempting to make item wording or methodological improvements in the estimation of attendance.

Study Limitations

While the amount of individual level data represented in this study is impressive, the number of study observations (k = 271) actually represents the sample size for the meta-analysis. This would be akin to conducting a primary study with 271 participants. While the comparison is not perfect because the sample size within each observation helps drive down the within-study error (or the sampling error) more than any one study could, it is important to remember that the between-studies variance is still based on the number of study observations. Thus, the precision of the estimates obtained in this study still have relatively wide confidence intervals. Moreover, while it is possible to perform a number of analyses adequately with 271 observations, when analyses begin to focus on sub-groups, the number of available observations can quickly dwindle. This is especially true given that the CMA meta-regression modeling procedure eliminates all observations with missing data. When socio-demographics like parental status (with only 56 observations) are modeled, then, the analysis is not based on the number of outcomes available for the dependent variable, but on the predictor with the fewest number of observations. This has the effect of dramatically diminishing the statistical power available for the analysis.

The lack of data for sub-groups showed up in several crucial places during the analysis. Specifically, a test of the efficacy of the self-administered item to reduce socially desirable responding was severely limited. Just three observations (from the Baylor Religion Survey) using adult samples were available for this item, making comparisons with other items and methods difficult. In addition, there were fewer than 10 observations available for items that had incorporated wording changes designed to address the ambiguous wording and social desirability problems. Coupled with an anticipated small effect size, the small number of studies made the detection of any impact these items might have had very difficult. Thus, despite the availability of a large amount of data for this project, some of the sub-group analyses were either not possible or severely underpowered.

In a related vein, the available data did not always cover the range of outcomes typically included in the measurement of a variable. For example, the income categories that were available in this study ranged from $25,000 to $99,999, which represents the middle-class well, but did not allow for an assessment of whether the income-related findings in this study would generalize to the lower and upper-class. The reporting of race in the first few decades was also limited to just two categories (White and Black). As a result, racial and ethnic groups, such as Hispanic and Asian Americans are still relatively underrepresented in terms of the number of study observations with data for these groups. The reported gradations of some socio-demographic categories were also not always ideal. For example, the data available to measure parental status of school-age children were limited to the broadest age range possible (i.e., 5 to 17). Yet, there may be important differences in the attendance patterns of parents as their children grow and move into and through adolescence. Even breaking the age groups down so that they represent children in primary school (5 to 12 years) and secondary school (13 to 17 years) would be useful. Unfortunately, the available survey data did not support this breakdown. These data problems are partly the result of a circumscribed literature search, and partly the result of the measurement and reporting used in the primary studies.

This study also failed to incorporate safeguards for coding error. Wilson (2009) argued for the importance of checking for coder drift and coding oversights (due to coder fatigue), and this study proposed to use two raters as a check against these sources of error. Unfortunately, a lack of time and resources prevented the use of these safeguards. Although a conscientious effort was made to provide high quality coding and data extraction for this project, future efforts should secure the use of a second coder to formally assess data quality.

Directions for Future Research

One of the first objectives of future research efforts in this area is to locate additional data sources that will help address some of the sub-group and ancillary analyses that could not be performed here, or could only be performed with low statistical power. Efforts should focus on identifying nationally representative surveys of adults that use the self-administered delivery mode. This would allow for an assessment of this method's impact on reducing socially desirable responding in face-to-face and phone interviews. In addition, data from items that provide an "in the last year" time frame should be located and included in the comparison of items and measures designed to address the ambiguous time-frame problem. Given the other findings associated with question wording alterations, however, it is not expected that this type of item will make much of an impact, if at all, on the attendance rate. Yet, if nationally representative adult data are available for this item, it would be beneficial to at least explore a potential impact.

Efforts to obtain data on underrepresented groups would also be beneficial. For example, locating data from samples of the poor, the wealthy, minority racial and ethnic groups, minority sexual orientation groups and others would provide an avenue for studying attendance patterns for understudied groups. Obtaining data for finely tuned age groups, particularly through middle adulthood and into late adulthood would help shed light on life-span attendance patterns. And, incorporating data on some of the other categories of variables included in this study (e.g., separated or divorced persons, disabled or retired persons) would add to our understanding of the patterns of attendance among a wide variety of Americans. In order to capture data from these lesser studied groups, however, it will likely be necessary to include data from non-national, non-representative samples. Taking this approach would also allow for an assessment of the differences between local and national samples, as well as convenience and representative samples.

Incorporating several additional potential covariates or moderators of attendance might also be useful. For example, region of residence and religious affiliation are strongly tied to religious behavior and attendance patterns (e.g., see Gallup & Lindsay, 1999), and would likely be useful predictors of the attendance rate.[145] In addition, nationally representative surveys differ in terms of their non-response bias. Therefore, including survey response rates as a covariate might be a useful way of controlling for this source of bias across studies.[146] Efforts to identify and control for those more or less likely to over-report their attendance would also be beneficial. For example, if over-reporting is tied to the distinction between intrinsic, extrinsic and quest religious motivation (see Allport, 1950; Batson et al., 1993), then a subset of studies including those measures should be located and analyzed to ascertain whether the item, method and socio-demographic differences in attendance observed in this study persist after controlling for religious motivation.

Additional analyses should also be undertaken. The interaction between employment status and gender, for instance, is very interesting and enlightening. Perhaps that same interaction is present for the other two SES variables included in this study (i.e., income and education). Additional research efforts should explore these potential interactions, as well as others, to help shed light on any nuances in the socio-demographic relationships with attendance.

Analyses could also seek to test and verify some of the unexpected findings observed in this study. For example, the relationship between employment status and gender was opposite of what was expected, as was the interaction between employment status and gender. Further testing of these relationships would add clarity to the literature. The findings associated with education were also peculiar. In simplistic models, education was negatively related to attendance, but in the full, multivariate model, education was positively related to attendance. Research efforts should continue to test this relationship to determine the intricacies, if any,

[145] These data were collected during the data extraction phase, but were not analyzed because hypotheses had not been developed to guide the analyses of these data.
[146] These data were intermittently collected during the data extraction phase, and a more concerted effort is needed to extract these data from all available studies before they can be used to control for non-response bias.

driving both the negative and positive associations found here. For instance, it would be interesting to see if the education-attendance association is moderated by schooling-type (i.e., secular vs. parochial). The lack of a relationship between the proportion of Hispanic Americans sampled and attendance was also unexpected and should continue to be tested as additional data on Hispanic Americans become available.

Studies investigating the prevalence of attendance could also benefit from using new and alternative methodological approaches. For example, experimental research that varies item wording, delivery mode and methodology would provide direct evidence of the bias introduced into the gold standard estimates by the problems of ambiguity and social desirability. Experimental studies could also be useful for developing and identifying more effective item wording, delivery mode and methodological variants that are also practical to implement for researchers lacking the resources for an intensive methodological approach such as the time-use method. In addition, new technologies could be used to provide an alternate method of counting. For example, Wiehe, Carroll, Liu, Haberkorn, Hoch, Wilson et al. (2008) explored the potential utility of providing participants with GPS-enabled cellular phones, and found that they were a reliable means of tracking location and collecting daily diary data. Studies capitalizing on this technology could use traditional methods of asking participants about their attendance behavior—where and how often they attend—and then follow them over a period of time using the GPS tracking function to both verify the accuracy of the self-reported attendance data and capture attendance counts that would be relatively free of error.

Beyond focusing on the psychometrics of attendance, it would also be interesting to see whether the biopsychosocial attendance associations noted in the literature hold across different methods of measuring attendance. For example, do the health and wellness relationships identified using “how often” questions hold when attendance is measured via the time-use method? If there is some variation in findings, then perhaps that will lead to the identification of a set of active ingredients not previously identified. For example, if associations with ‘how often” items do not hold for time-use items, then perhaps religious identity (or how someone thinks of themselves as religious or not), as opposed to religious behavior, is the key element inside the “black box” of attendance.

Religious groups might also find utility in extending the results of this study for their own growth. For example, young adult males were identified as relatively infrequent attenders. Research on marketing to, and creating a welcoming and engaging environment for, this demographic could be consulted and implemented with the hope of reaching an underrepresented segment of the market. On the other hand, parents of school-age children were identified as relatively frequent attenders. To capitalize on this finding, congregations in family-rich geographic environments could focus their resources on providing quality religious socialization and education experiences for school-age children. In this way, the identification of even small attendance differences between socio-demographic groups can be used to tap into potential growth areas for congregations and denominations.

Finally, future research should focus on improving upon the weaknesses of this study by replicating, but also filling in missing data and variable gaps; improving the statistical power available for sub-group and extreme-group analyses; increasing the sample’s representativeness of the attendance literature; and, by more thoroughly understanding the psychometric and socio-demographic factors that influence the prevalence of attendance in America.

Table 1

Summary of Attendance Estimates, Inflation Factors and Expected Estimates if Bias were Removed from the Gold Standard Items

Issue / Item	Year of Administration	Expected Regular Attendance Rate	Expected Attendance Rate in Past Week	Baseline Attendance Rate	Comparison Attendance Rate	Inflation Factor
Gold Standard						
Original Gallup item	1939 – 2009		41%			
Original GSS item	1972 – 2006	43%				
New Gallup item	1992 – 2009	44%				
Ambiguous Time Frame						
Regular Attendance vs. Att. in Past Week						
Orig GSS/new Gallup items vs Orig Gallup item	1939 – 2009	41%, 41.1%		43%, 44%	41%	1.05, 1.07
Ambiguous Terminology						
GSS Branching item (before and after removing activities other than religious services)	1996		37.6%	30.6%	28.0%	1.09
Item from Hadaway et al. (1993)	1993		40.2%	35.8%	35.2%	1.02
Item from Marler & Hadaway (1999)	1996		36.3%	69.7%	61.0%	1.13
Social Desirability						
Time-Use items						
5 Gallup/GSS items vs EPA Time-Use item	1992 – 1994		29%	41%	29%	1.41
Orig GSS & Gallup items vs SRC Time-Use item (18-29 year-olds)	1992 – 1994		28.7%, 24.0%	30%, 36%	21%	1.43, 1.71
Orig Gallup vs ATUS items	2003 – 2005		27%, 26.7%, 26.3%	41%	27%, 26.7%, 26.3%	1.52, 1.54, 1.56
Self-Administered item						
Orig GSS & Gallup items vs Monitoring the Future Survey (MFS) item (18-29 year-olds)	1993 – 1994		31.5%, 26.1%[1]	30%, 36%	23%	1.30, 1.57
Count Method						
Local survey vs counts of Protestants	1993		24.3%	33.2%	19.6%	1.69
Local survey vs counts of Catholics	1997		22.5%	51%	28%	1.82
National surveys vs counts of Catholics	1993		18.7% - 35.0%			1.17 – 2.19
Orig Gallup item vs count estimate from hypersmaple of U.S. congregations	2005		21.1%	41%	21.1%	1.94

[1] Although the self-administered item measures regular attendance, it was converted to a measure of attendance in the past week by Presser & Stinson (1998).

Table 2

Indicators of Religiosity for Professors and Scientists, Graduate Students and the General Public

	Monthly Attendance	Daily Prayer	"Strong" Member of Religion	Feel Close to God	Believe in Afterlife	Believe in Bible	Affiliated with a Religion
Professors and Scientists	38%	37%	27%	64%	66%	62%	81%
Graduate Students	43%	42%	34%	75%	73%	66%	87%
General Public	45%	57%	34%	85%	78%	85%	93%

Note. This table was reproduced from Stark et al.'s (1997) Table 2 (p. 36), which summarized data on over 30,000 GSS respondents who participated between 1972 and 1990. Approximately 300 of the GSS respondents were professors or scientists, 1,300 were graduate students and the rest were classified as being members of the general public.

Table 3

Key Word Search Terms Used in Electronic Search Engines

Key Word Categories / Search Terms
Primary (Stand-Alone) Prefixes and Root Words
“Relig,” “Spirit”
Secondary Prefixes and Root Words (To be Used With Other Terms and Prefixes)
“Activ,” “Attend,” “Behav,” “Commit,” “Engag,” “Involv,” “Org,” “Participat,” “Practic”
Names for Religious Services
Chapel, Church, Mass, Service, Worship
Religious Places Where Services are Held
Congregation, Mosque, Parish, Sanctuary, Shrine, Synagogue, Tabernacle, Temple
Terms Representing High Probability Sources
“Meta,” Review, Synthesis
Religiously Extreme Groups
Agnostic, “Apost,” “Atheis,” Born Again, Conservative, Cult, Demand Side, Evangelical, “Fundamental,” Nones, Sect, Supply Side, Unchurched
Religious/Spiritual Instruments
Baylor Religion Survey, Duke University Religion Index (DUREL), Spirituality/Religiousness Index, See Hill & Hood (1999); also Fetzer Institute (1999); Hill & Pargament (2003)
Religious Terminology Related to Attendance
Demand Side, Supply Side
Primary Search Terms
“Religious Service Attendance,” “Church Attendance,” “Religious Participation,” “Church Participation,” “Religious Behavior,” “Religious Involvement”

Note. All terms are placed in only one category to avoid redundancy. Prefixes and root words are enclosed in quotes.

Table 4

Electronic Search Engines and Bibliographic Databases Used to Locate and Retrieve Study Data on Religious Service Attendance

Search Engine	Description of Search Engine	Relevant Bibliographic Databases and Known Fields of Study
EBSCO	Includes 50 bibliographic databases covering a variety of academic disciplines	*Academic Search Complete, America: History & Life, American Doctoral Dissertations, Business Source Complete, Business Source Elite, CAB Abstracts 1990-Present, Communication & Mass Media Complete, Consumer Health Complete - EBSCOhost, eBook Collection (EBSCOhost), EconLit, Education Abstracts (H.W. Wilson), ERIC, Health and Psychosocial Instruments, Health Source – Consumer Edition, Health Source: Nursing/Academic Edition, Humanities International Complete, LGBT Life with Full Text, MasterFILE Premier, MEDLINE, Mental Measurements Yearbook with Tests in Print, Military & Government Collection, OmniFile Full Text Select (H.W. Wilson), Philosopher's Index, Professional Development Collection, PsycARTICLES, PsycCRITIQUES, PsycINFO, SocINDEX, SocINDEX with Full Text, SPORTDiscus, Teacher Reference Center*
FirstSearch	Includes 13 bibliographic databases covering journal articles, books, e-books, conference papers and proceedings, government documents, theses, dissertations and other materials available from a variety of disciplines	*ArticleFirst, Ebooks, ECO, ERIC, GPO, IllinoisCatalog, MEDLINE, OAIster, PapersFirst, Proceedings, WorldCat and WorldCatDissertations*
JSTOR	Contains millions of documents published in over 2,300 journals, thousands of monographs and 45,000 ebooks representing 58 disciplines, including African American, American, American Indian and Asian Studies, Anthropology, Biological Sciences, Business, Criminology, Economics, Education, Epidemiology, Women's Studies, Health Sciences, Jewish Studies, Latin American Studies, Nursing, Philosophy, Political Science, Psychology, Religion, Sociology and Theology	*JSTOR*
LexisNexis	Contains billions of records from a variety of sources, including blogs, consumer information reports, magazines, market and industry reports, news media, publications and wire services. The database also contains contact and other information for people and companies. Topical areas include education, energy, financial services, government, healthcare, insurance, legal, life sciences, manufacturing, media, non-profits, political, retail and sales and technology	*LexisNexis*
ProQuest	Includes 22 databases in the areas of art, business, health and medicine, history, literature and language, science and technology and the social sciences	*Ethnic NewsWatch, Historical Newspapers: The Wall Street Journal, Chicago Defender, Chicago Tribune, The New York Times, PILOTS: Published International Literature on Traumatic Stress, ProQuest Newsstand and ProQuest Dissertations & Theses* (by subject area)

Table 4 (Continued)

Search Engine	Description of Search Engine	Relevant Bibliographic Databases
Web of Science	Provides access to 18 searchable databases covering scientific literature in the arts and humanities, biological sciences, health and medicine and social sciences. Users also have access to conference proceedings, symposia, seminars, colloquia, workshops, conventions and author information. Forward citation searching is a key feature	*Web of Science Core Collection (Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Conference Proceedings Citation Index – Science, Conference Proceedings Citation Index – Social Science & Humanities, Book Citation Index – Science, Book Citation Index – Social Sciences & Humanities, Emerging Sources Citation Index), BIOSIS Citation Index, Current Contents Connect, Data Citation Index, MEDLINE* and *SciELO Citation Index*
Google Scholar	Search engine designed to help users locate published and unpublished literature from a wide array of scientific fields. Examples of document types include abstracts, books, court opinions, journal articles, theses and dissertations	*Google Scholar*

Table 5

Operationally Defining the Four Geographic Regions by Their Constituent States

Geographic Region	Constituent States
Northeast	Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
Midwest	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin
South	Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia
West	Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming

Table 6

Nationally Representative Data Sources, Years of Administration, Delivery Mode, Number of Cases and Available Demographics

Instrument or Method	Literature Source	Citation	Date Range (Observations)	Total Cases	Gender	Race / Ethnicity	Age	Marital Status	Parental Status	Ed.	Emply. Status	Income
American Heritage Time Use Study	Data Site: AHTUS	Fisher & Gershuny (2015)	1965 – 2012 (21)	35,233	Y	Y	Y	Y	Y	Y	Y	Y
American National Election Study	Consortium: ARDA	Burns et al. (2000, 2002, 2004, 2008)	2000 – 2008 (4)	5,603	Y	Y	Y	Y	--	Y	Y	Y
American Time Use Study	Manual Scan	Presser & Chaves (2007)	2003 – 2005 (3)	10,590	--	--	--	--	--	--	--	--
Americans' Changing Lives: Waves, I – V	Consortium: ICPSR	House (2014)	1986 – 2011 (5)	11,850	Y	Y	Y	Y	Y	Y	Y	Y
Baylor Religion Survey, Waves I – III	Consortium: ARDA	Bader et al. (2005, 2007, 2010)	2005 – 2010 (3)	5,030	Y	Y	Y	Y	Y	Y	--	Y
Black Women's Health Study	Newsletter (*Crossroads*)	Rosenberg (2017)*	2005	36,999	Y	Y	Y	Y	--	--	--	--
Counts of 48 Catholic Dioceses	Manual Scan	Chaves & Cavendish (1994)	1990 (1)	24,332,324	--	--	--	--	--	--	--	--
Counts of Five Religious Groups	Manual Scan	Hadaway & Marler (2005)	2001-02 (5)	224,126,091	--	--	--	--	--	--	--	--
Counts of Protestants in Ashtabula Co., OH	Manual Scan	Hadaway et al. (1993)	1992 (1)	66,565	--	--	--	--	--	--	--	--

Table 6 (Continued)

Instrument or Method	Literature Source	Citation	Date Range (Observations)	Total Cases	Gender	Race / Ethnicity	Age	Marital Status	Parental Status	Ed.	Emply. Status	Income
EPA Time-Use Study	Manual Scan	Presser & Stinson 1998)	1993	1,442	--	--	--	--	--	--	--	--
Gallup Poll – Original Item	Consortium: Roper Cntr	Gallup Org. (1950-2015)	1950 – 2015 (45)	58,392	Y	Y	Y	Y	Y	Y	Y	Y
Gallup Poll – Original Item (Omits "Synagogue")	Consortium: Roper Cntr	Gallup Org. (1954-1976)	1954 – 1976 (15)	29,588	Y	Y	Y	Y	Y	Y	Y	Y
Gallup Poll – New Item	Consortium: Roper Cntr	Gallup Org. (1950-2015)	1992 – 2011 (16)	16,155	Y	Y	Y	Y	Y	Y	Y	Y
Gallup Poll – New (Mosque) Item	Consortium: Roper Cntr	Gallup Org. (1950-2015)	2010 – 2015 (9)	9,418	Y	Y	Y	Y	Y	Y	Y	Y
GSS – Branching Item	Data Site: NORC	Smith et al. (2016)	1996	950	Y	Y	Y	Y	Y	Y	Y	--
GSS – Original Item	Data Site: NORC	Smith et al. (2016)	1972 – 2014 30 Yrs	59,268	Y	Y	Y	Y	Y	Y	Y	Y
High School & Beyond	Consortium: ICPSR	US DOE/NCES (2001)	1980 2 Obs.	53,433	Y	Y	Y	--	--	--	--	--
U. of Michigan SRC Time-Use Studies	Manual Scan	Presser & Stinson (1998)	1965 – 1975 2 Yrs	606	--	--	--	--	--	--	--	--
Monitoring the Future Survey – 8th Grade	Consortium: ICPSR	Bachman et al. (2008); Johnston et al. (1999, 2006)	1992 – 2015 (24)	155,644	Y	Y	Y	--	--	--	--	--
Monitoring the Future Survey – 10th Grade	Consortium: ICPSR	Bachman et al. (2008); Johnston et al. (1999, 2006)	1992 – 2015 (24)	141,940	Y	Y	Y	--	--	--	--	--

Table 6 (Continued)

Instrument or Method	Literature Source	Citation	Date Range (Observations)	Total Cases	Gender	Race / Ethnicity	Age	Marital Status	Parental Status	Ed.	Emply. Status	Income
Monitoring the Future Survey – 12th Grade	Consortium: ICPSR	Bachman et al. (2008); Johnston et al. (1999, 2006)	1976 – 2015 (40)	530,436	Y	Y	Y	--	--	--	--	--
National Ed. Long. Survey	Consortium: ICPSR	US DOE/NCES (2002)	1990 – 1992 2 Yrs / 4 Obs.	20,975	Y	Y	Y	--	--	Y	--	--
National Study of Youth and Religion, Wave I	Manual Scan	Pearce et al. (2013)	2002-03 (1)	3,285	Y	Y	Y	--	--	Y	--	--
National Survey of Black Americans	Consortium: ICPSR	Jackson & Gurin (1999)	1980 (1)	1,920	Y	Y	Y	Y	Y	Y	Y	Y
Religious Landscape Study	Data Site: Pew Res Cntr	Pew Forum on Religion & Public Life (2007)	2007 (8)[a]	35,556	Y	Y	Y	Y	--	Y	--	Y
World Values Survey (U.S.)	Manual Scan	Aarts et al. (2008)	1981 – 2000 (3)	4,637	--	--	--	--	--	--	--	--

Note. “Y” denotes available data

* Lynn Rosenberg, Personal Communication, June 1, 2017

[a] The Religious Landscape Study provided one overall study observation, but also provided seven sub-group observations where attendance was reported by five age groups (18-29, 30-39, 40-49, 50-64 and 65 and over) and gender (male, female). Only the overall study observation was used in the primary analysis.

Table 7

Listing of the Question or Method Types, Question Text and Response Options for Each Data Source

Source	Question or Method Type	Delivery Mode[a]	Question or Procedure	Response Options	Item Problems Addressed
American National Election Study	How Often	Phone	Lots of things come up that keep people from attending religious services even if they want to. Thinking about your life these days, do you ever attend religious services, apart from occasional weddings, baptisms or funerals? IF YES: Do you go to religious services every week, almost every week, once or twice a month, a few times a year or never? IF EVERY WEEK: Would you say you go to religious services once a week or more often than once a week?	Never, A few times a year, Once or twice a month, Almost every week, Every week, More often than once a week	W, SD
Americans' Changing Lives Survey, Waves I-V	How Often	FTF / Phone	How often do you usually attend religious services?	Never, Less than once a month, About once a month, 2 or 3 times a month, Once a week, More than once a week	--
Baylor Religion Survey	How Often	Self-Admin.	How often do you attend religious services?	Never, Less than once a year, Once or twice a year, Several times a year, Once a month, 2-3 times a month, About weekly, Weekly, Several times a week	SD
*Black Women's Health Study	How Often	Self-Admin.	How often do you attend religious services?	Never, Less than once a month, About once a month, 2-3 times a month, Once a week, Several times a day	SD
Gallup Poll – New Item	How Often	Phone	How often do you attend church or synagogue?	Never, Seldom, About once a month, Almost every week, At least once a week	--
Gallup Poll – New (Mosque) Item	How Often	Phone	How often do you attend church, synagogue or mosque?	Never, Seldom, About once a month, Almost every week, At least once a week	--
General Social Survey – Original Item	How often	FTF	How often do you attend religious services?	Never, Less than once a year, About once or twice a year, Several times a year, About once a month, 2-3 times a month, Nearly every week, Every week, Several times a week	--

Table 7 (Continued)

Source	Question or Method Type	Delivery Mode[a]	Question or Procedure	Response Options	Item Problems Addressed
*High School & Beyond	How Often	Self-Admin.	In the past year, about how often have you attended religious services?	Not at all, Several times a year or less, About once a month, 2 or 3 times a month, About once a week, More than once a week	T-F, SD
*Monitoring the Future Survey (8th–12th Grades)[b]	How Often	Self-Admin.	How often do you attend religious services?	Never, Rarely, Once or twice a month, About once a week or more	SD
*National Education Longitudinal Study	How Often	Self-Admin.	In the past year, about how often have you attended religious services?	Not at all, Several times a year, About once a month, 2 or 3 times a month, About once a week, More than once a week	T-F, SD
*National Survey of Black Americans	How Often	FTF / Self-Admin.	How often do you usually attend religious services?	Never, Less than once a year, A few times a year, A few times a month (1-3 times), At least once a week (1-3 times/week), Nearly every day (4 or more times/day	--
*National Study of Youth and Religion	How Often	Phone	About how often do you usually attend religious services?	Never, Few times to many times a year, One to three times a month, Once a week or more	--
Religious Landscape Study	How Often	Phone	Aside from weddings and funerals, how often do you attend religious services?	Never, Seldom, A few times a year, Once or twice a month, Once a week, More than once a week	W
World Values Survey	How Often	FTF	How often do you attend religious services, apart from weddings, funerals and festivities?	Never, Once a year, On holy days, Once a month, Once a week, More than once a week	W
Gallup Poll – Original Item	Last Week	FTF / Phone	Did you, yourself, happen to attend church or synagogue in the last seven days?	No, Yes	T-F
Gallup Poll – Original Item (Omits "Synagogue")	Last Week	FTF / Phone	Did you, yourself, happen to attend church in the last seven days?	No, Yes	T-F

Table 7 (Continued)

Source	Question or Method Type	Delivery Mode[a]	Question or Procedure	Response Options	Item Problems Addressed
General Social Survey – Branching Item	Last Week	FTF	Now I'm going to ask you about things you did during the last seven days. I'm only interested in what you did during the last seven days. From last (DAY OF WEEK) to today did you: Attend religious services? -- During the last seven days did you do the following: (A) Attend a regular, weekly worship service at a church/synagogue (e.g., Mass or Sunday morning services). Don't include watching a service on TV or listening on the radio. (B) Watch a religious program on television or listen to a religious program on the radio? (C) Attend some other type of religious event or meeting (e.g., prayer breakfasts, Bible study groups, choir practices, church sponsored lectures, adult fellowship meetings)?	No, Yes	W, T-F
*Counts of Catholic Dioceses (Chaves & Cavendish, 1994)	Counts	Count	Survey data were used to estimate the proportion of a county's population that is Catholic. Congregations reported average attendance in a month and reported up to the Diocese.	Estimated Catholic population, Average Catholic Mass attendance	W, T-F, SD
*Counts of 5 American Religious Groups (Hadaway & Marler, 2005)	Counts	Count	Used directories and surveys to estimate the total number of congregations in the U.S., the number of congregations in each of five denominations and the average number of attendees per week for each denomination.	Estimated Denominational Population, Estimated Number of Weekly Attendees per Denomination	W, T-F, SD
*Counts of Protestants in Ashtabula Co., OH (Hadaway et al., 1993)	Counts	Count	Representative sample was used to estimate the number of Protestants and the number of weekly attenders in Ashtabula County, Ohio. Average attendance for a month was then reported, counted or estimated by the congregations and the research team.	Estimated Number of Protestants, Average number of Protestants counted or estimated to be in attendance	W, T-F, SD
American Heritage Time Use Study	Time-Use	Phone / Diary	We'd like to have you keep a list of all of your activities starting at midnight, running through the daytime on [Sunday], up to midnight again [Sunday] night. We'd also appreciate it if you could write in the times when you stopped one activity and started on a different one.	Did not report attending, Reported attending	W, T-F, SD

Table 7 (Continued)

Source	Question or Method Type	Delivery Mode[a]	Question or Procedure	Response Options	Item Problems Addressed
American Time-Use Study	Time-Use	Phone	Describe all of your primary activities from 4:00 am the previous day until 4:00 am the interview day.	Did not report attending, Reported attending	W, T-F, SD
EPA Time-Use Study	Time-Use	FTF	I would like to ask you about the things you did yesterday—from midnight Saturday to midnight last night. Let's start with midnight Saturday. What were you doing? What time did you finish? Where were you? What did you do next?	Did not report attending, Reported attending	W, T-F, SD
U. of Michigan SRC Time Use Studies	Time-Use	FTF	Respondents were asked to keep a diary of their activities; follow-up interviews were used to help clarify and fill-in missing information.	Did not report attending, Reported attending	W, T-F, SD

* Denotes samples that focus on specific populations (e.g., minors, specific religious racial and ethnic groups)

[a] Delivery Modes: Self-Admin = Self-Administered Survey; FTF = Face-to-Face; Phone = Phone Survey

[b] The same question is used for the 8th, 10th and 12th grade Monitoring the Future Surveys; hence, these data sources are collapsed here.

Table 8

Harmonization of "How Often" Attendance Item Response Options

	Response Option Weights													
Item	0.000	0.010	0.019	0.029	0.058	0.077	0.115	0.231	0.346	0.461	0.577	0.846	0.990	0.999
American National Election Study (ANES)	Never				A few times a year				Once or twice a month			Almost Every week	Every Week	More often than once a week
Americans' Changing Lives Survey, Waves I-V	Never						Less than once a month	About once a month			2 or 3 times a month		Once a week	More than once a week
Baylor Religion Survey	Never	Less than once a year		Once or twice a year		Several times a year		Once a month			2-3 times a month	About weekly	Weekly	Several times a week
Black Women's Health Study	Never						Less than once a month	About once a month			2-3 times a month		Once a week	Several times a day
Gallup Poll – New Item	Never	Seldom						About once a month				Almost every week		At least once a week
Gallup Poll – New (Mosque) Item	Never	Seldom						About once a month				Almost every week		At least once a week
General Social Survey – Original Item	Never	Less than once a year	About once or twice a year			Several times a year		About once a month			2-3 times a month	Nearly every week	Every week	Several times a week
High School & Beyond	Not at all				Several times a year or less			About once a month			2 or 3 times a month		About once a week	More than once a week
Monitoring the Future Survey (8^{th} – 12^{th} Grades)	Never	Rarely							Once or twice a month					About once a week or more

Table 8 (Continued)

	Response Option Weights													
Item	0.000	0.010	0.019	0.029	0.058	0.077	0.115	0.231	0.346	0.461	0.577	0.846	0.990	0.999
National Education Longitudinal Study	Not at all					Several times a year		About once a month			2 or 3 times a month		About once a week	More than once a week
National Survey of Black Americans	Never	Less than once a year			A few times a year						A few times a month (1-3 times)		At least once a week (1-3 times/week)	Nearly every day (4 or more times/day)
National Study of Youth and Religion	Never						Few times to many times a year			One to three times a month				Once a week or more
Religious Landscape Study	Never	Seldom			A few times a year				Once or twice a month				Once a week	More than once a week
World Values Survey	Never		Once a year		On holy days			Once a month					Once a week	More than once a week

Note. Items measuring attendance within the past week will use dichotomous weights for did not attend (0.000) and attended (1.000).

Table 9

Meta-Regression Results Assessing the Impact of the use of "Time-Frame" Language on Attendance Estimates (Hypothesis 1)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 156)	*p*-value
Constant	-6.9032	2.8740	-12.5802	-1.2262	-2.40	0.0175*
Year	0.0033	0.0014	0.0005	0.0062	2.31	0.0222*
Time-Frame	-0.2660	0.0509	-0.3664	-0.1655	-5.23	0.0000***
Model Summary:	$F(2,156) = 13.89$, $p < 0.001$***					
	$R^2 = 0.01$					
	$Tau^2 = 0.0112$, $I^2 = 99.93\%$					
	Q = 22,7993.62, df = 156, p < .001***					

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test, whereas the coefficient for Time-Frame was evaluated against the critical value for a one-tailed test given that the use of a time-frame was expected to *reduce* estimates of religious service attendance frequency. The "gold standard" items served as the reference group. The B coefficients in this table are logits.

Table 10

Meta-Regression Results Assessing the Impact of the Length of Time Specified in Item Language or Methodology on Attendance Estimates (Hypothesis 1a)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 155)	*p*-value
Constant	-5.5186	1.9424	-9.3556	-1.6815	-2.84	0.0051**
Year	0.0026	0.0010	0.0007	0.0046	2.71	0.0076**
Time-Frame Length: "Last 7 Days"	-0.0301	0.0375	-0.1041	0.0439	-0.80	0.2111
Time-Frame Length: "Yesterday"	-0.6784	0.0484	-0.7740	-0.5829	-14.03	0.0000***
Time-Frame Length: "Researcher Defined Count"	-0.9273	0.0651	-1.0559	-0.7987	-14.25	0.0000***

Model Summary: $F(4,155) = 99.16$, $p < 0.001$***
$R^2 = 0.00$
$Tau^2 = 0.0356$, $I^2 = 99.9\%$
Q = 1105419.21, df = 155, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The "gold standard" items served as the reference group. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test, whereas the coefficients for Time-Frame Length were evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 11

Meta-Regression Results Assessing the Impact of Items and Methods Designed to Address the Ambiguous Wording Problem (Hypothesis 2)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 102)	*p*-value
Constant	-5.8284	2.0542	-9.9029	-1.7539	-2.84	0.0027**
Year	0.0028	0.0010	0.0008	0.0049	2.71	0.0039**
Item Wording Changes	0.0780	0.0726	-0.0660	0.2219	1.07	0.1427
Methodology Changes	-0.7537	0.0424	-0.8378	-0.6696	-17.77	0.0000***
Model Summary:	$F(3,102) = 115.40$, $p < .001$***					
	$R^2 = 0.00$					
	$Tau^2 = 0.0337$, $I^2 = 99.9\%$					
	Q = 1,005,057.72, df = 102, p < .001***					

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The "gold standard" items served as the reference group. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test, whereas the coefficient for Less Ambiguous Wording was evaluated against the critical value for a one-tailed test given that unambiguous wording is expected to *reduce* estimates of attendance frequency. The B coefficients in this table are logits.

Table 12

Meta-Regression Results Assessing the Impact of Items and Methods Designed to Limit Social Desirability Effects on Attendance Estimates (Hypothesis 3)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 100)	*p*-value
Constant	-7.0964	2.0997	-11.2621	-2.9307	-3.38	0.0005***
Year	0.0034	0.0011	0.0013	0.0055	3.26	0.0008
Wording and Delivery Mode	-0.1067	0.0852	-0.2756	0.0623	-1.25	0.1066
Methodology (Time-Use / Counts)	-0.7601	0.0425	-0.8444	-0.6758	-17.9	0.0000***
Model Summary:	$F(3,100) = 110.12$, $p < 0.001$***					
	$R^2 = 0.00$					
	$Tau^2 = 0.0334$, $I^2 = 99.9\%$					
	Q = 1004,890.83, df = 100, p < .001***					

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The “gold standard” items served as the reference group. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test, whereas the coefficients for Wording and Delivery Mode and Methodology were evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 13

Meta-Regression Results Assessing the Impact of Items Designed to Minimize Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 3a)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 122)	*p*-value
Constant	-7.3374	1.7412	-10.7843	-3.8905	-4.21	0.0000***
Year	0.0036	0.0009	0.0018	0.0053	4.06	0.0000***
Social Desirability	-0.1136	0.0994	-0.3103	0.0832	-1.14	0.1277
Ambiguous Time-Frame	-0.0394	0.0317	-0.1022	0.0234	-1.24	0.1084
Ambiguous Wording	0.2378	0.0842	0.0711	0.4045	2.82	0.0028**

Model Summary: $F(4,122) = 6.26$, $p < 0.001$***
$R^2 = 0.14$
$Tau^2 = 0.0244$, $I^2 = 90.3\%$
Q = 1255.10, df = 122, p < .001***

^ *p* < .10 * *p* < .05 ** *p* < .01 *** *p* < .001

Note. The “gold standard” items served as the reference group. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test, whereas the coefficients for Ambiguous Wording, Ambiguous Time-Frame and Social Desirability were evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 14

Meta-Regression Results Comparing the Items and Methods Designed to Reduce Social Desirability: Self-Administered Items, Time-Use Items and the Count-Based Method (Hypothesis 4)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 39)	*p*-value
Constant	6.7840	5.1582	-3.6495	17.2174	1.32	0.0981^
Year	-0.0039	0.0026	-0.0091	0.0014	-1.49	0.0719^
Self-Administered Items	0.6792	0.1249	0.4265	0.9319	5.44	0.0000***
Count-Based Method	-0.2863	0.0733	-0.4345	-0.1381	-3.91	0.0000***
Model Summary:	$F(3,39) = 17.30$, $p < 0.001$***					
	$R^2 = 0.00$					
	$Tau^2 = 0.0357$, $I^2 = 100.0\%$					
	Q = 110,4295.68, df = 39, p < .001***					

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The time-use items served as the reference group. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test, whereas the coefficients for Self-Administered Items and Count-Based Method were evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 15

Meta-Regression Results Assessing the Association Between Gender and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 5)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 236)	*p*-value
Constant	-4.2176	1.7682	-7.7011	-0.7341	-2.39	0.0179*
Year	0.0016	0.0009	-0.0001	0.0034	1.83	0.0690^
Ambiguous Wording	-0.7485	0.0633	-0.8732	-0.6239	-11.83	0.0000***
Ambiguous Time-Frame	0.0511	0.0357	-0.0192	0.1214	1.43	0.0769^
Social Desirability	-0.0025	0.0336	-0.0686	0.0637	-0.07	0.4709
Proportion Female	1.3734	0.2797	0.8225	1.9243	4.91	0.0000***

Model Summary: $F(5,236) = 42.69$, $p < 0.001$***
$R^2 = 0.47$
$Tau^2 = 0.0360$, $I^2 = 97.6\%$
Q = 9725.09, df = 236, $p < .001$***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 16

Meta-Regression Results Assessing the Association Between African American Sample Proportions and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 6)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 225)	*p*-value
Constant	-4.4372	1.9878	-8.3542	-0.5202	-2.23	0.0133*
Year	0.0021	0.0010	0.0001	0.0040	2.07	0.0200*
Ambiguous Wording	-0.4437	0.0591	-0.5603	-0.3272	-7.50	0.0000***
Ambiguous Time-Frame	-0.0743	0.0397	-0.1525	0.0039	-1.87	0.0312*
Social Desirability	-0.1079	0.0362	-0.1792	-0.0366	-2.98	0.0016**
Proportion African American	0.7385	0.1611	0.4211	1.0560	4.58	0.0000***

Model Summary: $F(5,225) = 29.78$, $p < 0.001$***
$R^2 = 0.28$
$Tau^2 = 0.0468$, $I^2 = 98.2\%$
Q = 12440.89, df = 225, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 17

Meta-Regression Results Assessing the Association Between Race and Ethnicity Sample Proportions and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 73)	*p*-value
Constant	-6.5950	1.0698	-8.7271	-4.4630	-6.16	0.0000***
Year	0.0022	0.0007	0.0009	0.0035	3.34	0.0013**
Ambiguous Wording	-0.6221	0.0805	-0.7826	-0.4616	-7.73	0.0000***
Ambiguous Time-Frame	-0.0866	0.0315	-0.1494	-0.0238	-2.75	0.0038**
Social Desirability	0.0560	0.0706	-0.0848	0.1967	0.79	0.2153
Proportion Asian	-0.4855	0.9485	-2.3758	1.4048	-0.51	0.3051
Proportion African American	2.6137	0.6231	1.3718	3.8556	4.19	0.0000***
Proportion Hispanic	0.6166	0.8087	-0.9950	2.2283	0.76	0.2241
Proportion White	2.1614	0.6491	0.8677	3.4550	3.33	0.0007***

Model Summary: $F(8,73) = 131.50$, $p < 0.001$***
$R^2 = 0.98$
$Tau^2 = 0.0050$, $I^2 = 65.9\%$
Q = 213.82, df = 73, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 18

Meta-Regression Results Assessing the Association Between Age Group Sample Proportions and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 7)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 215)	*p*-value
Constant	-2.440	1.8404	-5.8715	1.3835	-1.22	0.2241
Year	0.0010	0.0009	-0.0009	0.0028	1.03	0.3040
Ambiguous Wording	-0.2841	0.0931	-0.4677	-0.1006	-3.05	0.0013**
Ambiguous Time-Frame	-0.0429	0.0408	-0.1234	0.0377	-1.05	0.1476
Social Desirability	-0.4607	0.0959	-0.6498	-0.2716	-4.80	0.0000***
Proportion Age 13-17	0.4895	0.3522	-0.2047	1.1836	1.39	0.0830
Proportion Age 18-30	-1.0240	0.3978	-1.8080	-0.2400	-2.57	0.0054**
Proportion Age 31-64	0.5360	0.3324	-0.1190	1.1911	1.61	0.0541
Proportion Age 65+	-0.4796	0.4038	-1.2755	0.3163	-1.19	0.1181

Model Summary: $F(8,215) = 24.78$, $p < 0.001$***

$R^2 = 0.29$

$Tau^2 = 0.0368$, $I^2 = 97.8\%$

Q = 9535.19, df = 215, p < .001***

^ *p* < .10 * *p* < .05 ** *p* < .01 *** *p* < .001

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 19

Meta-Regression Results Assessing the Association Between Sample Proportions of Married Respondents and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 8)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 97)	*p*-value
Constant	-7.112	1.7081	-10.5021	-3.722	-4.16	0.0000***
Year	0.0032	0.0009	0.0015	0.0049	3.69	0.0002***
Ambiguous Wording	-0.2117	0.0736	-0.3578	-0.0655	-2.87	0.0025
Ambiguous Time-Frame	-0.2075	0.0403	-0.2875	-0.1274	-5.15	0.0000***
Social Desirability	-0.2722	0.0693	-0.4098	-0.1345	-3.93	0.0001***
Proportion Married	0.8583	0.2281	0.4056	1.3111	3.76	0.0001***
Model Summary:	$F(5,97) = 42.93$, $p < 0.001$***					
	$R^2 = 0.74$					
	$Tau^2 = 0.0226$, $I^2 = 89.43\%$					
	Q = 917.59, df = 97, p < .001***					

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 20

Meta-Regression Results Assessing the Association Between Sample Proportions of Parents with School-Age Children and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 9)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 49)	*p*-value
Constant	2.1713	1.8838	-1.6143	5.9569	1.15	0.1273
Year	-0.0014	0.001	-0.0034	0.0005	-1.47	0.0745^
Ambiguous Wording	0.4306	0.1803	0.0682	0.793	2.39	0.0104*
Ambiguous Time-Frame	-0.8070	0.1110	-1.0301	-0.584	-7.27	0.0000***
Social Desirability	-0.1988	0.1273	-0.4546	0.057	-1.56	0.0624^
Proportion Parent of School-Age Children	1.1295	0.2761	0.5747	1.6843	4.09	0.0001***

Model Summary: $F(5,49) = 69.84$, $p < 0.001$***
$R^2 = 0.92$
$Tau^2 = 0.0062$, $I^2 = 71.18\%$
Q = 170.01, df = 49, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 21

Meta-Regression Results Assessing the Association Between Mean Sample Education Years and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 10)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 134)	*p*-value
Constant	-5.9134	1.5458	-8.9708	-2.8560	-3.83	0.0001***
Year	0.0031	0.0008	0.0015	0.0046	3.85	0.0001***
Ambiguous Wording	-0.6306	0.0825	-0.7937	-0.4674	-7.64	0.0000***
Ambiguous Time-Frame	-0.0358	0.0304	-0.0958	0.0243	-1.18	0.1205
Social Desirability	-0.0449	0.0770	-0.1971	0.1074	-0.58	0.2806
Mean Education Years	-0.0325	0.0133	-0.0587	-0.0062	-2.45	0.0078**

Model Summary: $F(5,134) = 66.74$, $p < 0.001$***
$R^2 = 0.78$
$Tau^2 = 0.0018$, $I^2 = 87.00\%$
Q = 1030.70, df = 134, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 22

Meta-Regression Results Assessing the Association Between Sample Median Income and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 11)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 113)	*p*-value
Constant	-4.6401	1.5882	-7.7867	-1.4935	-2.92	0.0021**
Year	0.0022	0.0008	0.0007	0.0038	2.80	0.0030**
Ambiguous Wording	-0.0372	0.0755	-0.1867	0.1124	-0.49	0.3117
Ambiguous Time-Frame	-0.1687	0.0306	-0.2294	-0.1081	-5.51	0.0000***
Social Desirability	-0.4544	0.0794	-0.6116	-0.2972	-5.73	0.0000***
Median Income: $25,000 - $49,999	-0.1108	0.0322	-0.1746	-0.0471	-3.44	0.0004***
Median Income: $75,000 - $99,999	0.2861	0.0713	0.1448	0.4274	4.01	0.0001***

Model Summary: $F(6,113) = 38.45$, $p < 0.001$***
$R^2 = 0.74$
$Tau^2 = 0.0178$, $I^2 = 86.2\%$
Q = 816.61, df = 113, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The referent category was Median Income: $50,000 - $74,999. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 23

Meta-Regression Results Assessing the Association Between Sample Proportion Employed and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 12)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 85)	*p*-value
Constant	-5.4816	1.7102	-8.8819	-2.0813	-3.21	0.002**
Year	0.0024	0.0009	0.0006	0.0041	2.72	0.0078**
Ambiguous Wording	-0.3576	0.1173	-0.5909	-0.1244	-3.05	0.0015**
Ambiguous Time-Frame	-0.0735	0.0421	-0.1572	0.0103	-1.74	0.0423*
Social Desirability	-0.1671	0.1199	-0.4054	0.0712	-1.39	0.0835
Proportion Employed	0.7234	0.1743	0.3768	1.07	4.15	0.0000***

Model Summary: $F(5,85) = 52.45$, $p < 0.001$***
$R^2 = 0.82$
$Tau^2 = 0.1360$, $I^2 = 87.1\%$
Q = 660.85, df = 85, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 24

Meta-Regression Results Assessing the Interaction Between Proportion Female and Proportion Employed with Regard to Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame and Social Desirability (Hypothesis 12a and 12b)

	Model					
Variable	B	SE(B)	95% Lower CI	95% Upper CI	t-value (df = 83)	*p*-value
Constant	-0.2997	2.4654	-5.2032	4.6038	-0.12	0.9035
Year	0.0029	0.0009	0.0011	0.0047	3.16	0.0022**
Ambiguous Wording	-0.4583	0.1271	-0.7111	-0.2056	-3.61	0.0003***
Ambiguous Time-Frame	-0.0777	0.0453	-0.1678	0.0123	-1.72	0.0448*
Social Desirability	-0.0169	0.1247	-0.2649	0.2311	-0.14	0.4462
Proportion Female	-11.368	3.0700	-17.475	-5.2622	-3.70	0.0002***
Proportion Employed	-8.9717	2.4556	-13.856	-4.0877	-3.65	0.0002***
P-Female by P-Employed	18.009	4.5904	8.8787	27.1391	3.92	0.0001***

Model Summary: $F(7,83) = 40.50$, $p < 0.001$***
$R^2 = 0.83$
$Tau^2 = 0.0170$, $I^2 = 86.0\%$
Q = 594.76, df = 83, p < .001***

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 25

Meta-Regression Models Assessing the Relationships Between Socio-Demographics and Religious Service Attendance on any Given Sunday After Controlling for Study Year, Ambiguous Wording, Ambiguous Time-Frame, Social Desirability and Other Socio-Demographics

	Block One		Block Two		Block Three		Fully Specified Model	
Variable	B	*t-value*	B	*t-value*	B	*t-value*	B	*t-value*
Constant	-3.69	-3.26***	4.49	3.49***	-4.38	-2.73**	-3.32	-2.00*
Year	0.00	0.55	-003	-4.30***	0.003	3.22**	-.0002	-0.22
Ambiguous Wording	0.23	4.89***	0.44	3.21**	-0.04	-0.26	-0.02	-0.15
Ambiguous Time-Frame	-0.12	-4.40***	-0.79	-9.85***	-0.07	-1.63^	-0.14	-4.12***
Social Desirability	-0.75	-12.81***	-0.18	-1.81*	-0.55	-3.92***	-0.49	-4.91***
Female	0.51	1.82*					0.77	1.19
African American	2.75	5.61***					2.34	6.67***
Asian	-0.61	-0.81						
Hispanic	0.30	0.48						
White	2.32	4.56***					1.70	6.70***
Age 13 – 17	0.96	1.04						
Age 18 – 30	-0.43	-2.24*					-0.61	-1.82*
Age 31 – 64	0.75	2.35*					-0.43	-1.56^
Married			1.39	6.67***			1.02	3.29***
Parent of School-Age Children			0.23	0.62				
Mean Education (Yrs)					-0.11	-4.36***	0.08	2.27*
Employed (Full-Time)					0.48	2.19*	0.13	0.67
Median Income: $25,000-$49,999					-0.18	-4.37***	0.02	0.67
Median Income: $75,000-$99,99					0.10	1.42^	-0.00	-0.01
Model df	64		45		72		59	

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 26

Meta-Regression Models Estimating the Prevalence of Religious Service Attendance in America on any Given Sunday After Controlling for Study Year and Available Socio-Demographics

	Gold Standard		Last 7 Days		Time-Use		Count-Based Method	
Variable	B	*t-value*	B	*t-value*	B	*t-value*	B	*t-value*
Constant	-0.34	-0.24	-4.64	-1.66^	6.33	2.12*	5.01	0.30
Year	-0.002	-2.14*	0.001	1.05	-0.004	-2.51*	-.003	-0.36
African American	2.45	7.28***						
White	1.45	5.42***						
Age 18 – 30	--	ns	-0.38	-1.60^				
Age 31 – 64	--	ns	2.36	5.99***	0.22	1.86*		
Married	1.48	4.71***						
Mean Education (Yrs)								
Prevalence of RSA in First Year	1972 = 43.3%		1950 = 40.9%		1966 = 31.4%		1979 = 24.5%	
Prevalence of RSA in 2015[a]	2015 = 41.4%		2015 = 43.1%		2015 = 27.6%		2015 = 22.5%	
Inflation Factor (*c.* Gold Standard)	--		0.96		1.50		1.84	

^ $p < .10$ * $p < .05$ ** $p < .01$ *** $p < .001$

[a] The last years of data collection for the time-use item and count-based method were 2012 and 2011, respectively, but these figures were projected forward to 2015 for ease of comparison.

Note. The coefficients for Constant and Year were evaluated against the critical value for a two-tailed test; the remaining covariates are evaluated against the critical value for a one-tailed test. The B coefficients in this table are logits.

Table 27

Summary of Study Hypotheses and Findings

#	Hypothesis	Outcome	Summary Finding
1	Studies that utilize items and methods that (explicitly or implicitly) provide a specific time frame will yield lower attendance estimates than studies that fail to do so.	Supported	Use of a time-frame lowers RSA by 6.4% (1.17[a])
1a	The length of the time frame specified by an item or method will be positively related to estimates of attendance (e.g., “regular” attendance > attendance “in the past week” > attendance “yesterday”).	Partially Supported	RSA drops non-significantly by 0.8% for attendance in past week, but significantly by 15.4% (time-use; 1.56[a]) and 20.0% (count; 1.87[a]) for attendance “yesterday”
2	Studies that utilize items and methods that control for, or eliminate the use of ambiguous terminology (e.g., “religious services,” “church” or “synagogue”) will yield lower attendance estimates than studies that fail to do so.	Partially Supported	RSA increases non-significantly by 1.9% for items that use wording changes to clarify ambiguous terminology, but decreases significantly by 17.3% (1.68[a]) for methods that minimize or eliminate ambiguous wording
3	Studies that utilize items and methods that minimize or eliminate opportunities for socially desirable responding will yield lower attendance estimates than studies that fail to do so.	Partially Supported	RSA drops non-significantly by 3.9% for items using wording and delivery mode changes to minimize social desirability, but significantly by 17.1% (count; 1.64[a]) for methods that minimize or eliminate social desirability
3a	Minimizing or eliminating opportunities for socially desirable responding will suppress attendance estimates to a greater extent than controlling for, or eliminating ambiguous language (i.e., ambiguous time frame frames and terminology).	Not Supported	Items addressing social desirability through wording and delivery mode changes (45.1%) produce non-significantly lower RSA estimates than items addressing the ambiguous time-frame (46.9%) and terminology (53.9%) through wording changes
4	Among the items and methods reviewed here that attempt to reduce socially desirable responding, the count-based method will yield lower attendance estimates than the time-use item, which will yield lower estimates than the self-administered item.	Supported	RSA from the self-administered items (40.2%) was significantly higher than for time-use (25.5%), which was significantly higher than count-based (20.4%)
5	Studies with higher percentages of female participants will yield higher attendance estimates, on average, than studies with higher percentages of male participants.	Supported	RSA for all female and all male samples are expected to be 57.9% and 25.9%, respectively
6	The percentage of African American participants in studies will be positively related to the attendance estimates that these studies yield.	Supported	RSA for all African American and all non-African American samples are expected to be 60.0% and 41.8%, respectively
7	The average age of a sample will be non-linearly related to religious service attendance estimates, such that studies of young adults (mean age = 18 to 30 years) will yield the lowest estimates, followed by adolescents (13 to 17 years), middle-aged adults (31 to 64 years) and elderly adults (65 years and older), respectively.	Partially Supported	RSA for samples composed entirely of those between the ages of 18-30, 13-17, 31-64 and 65+ are expected to be 17.7%, 49.3%, 50.5% and 27.0%, respectively. All but the RSA for 65+ follows the expected pattern

Table 27 (Continued)

#	Hypothesis	Outcome	Finding Summary
8	The percentage of married participants in a study will be positively related to religious service attendance estimates.	Supported	RSA for all married and all non-married samples are expected to be 49.6% and 29.4%, respectively
9	The percentage of participants who are parents of school-aged children will be positively related to religious service attendance estimates.	Supported	RSA for all parents of school-age children and all non-parents of school-age children samples are expected to be 55.9% and 26.8%, respectively
10	Studies that utilize samples of relatively well educated participants will yield relatively high attendance estimates.	Partially Supported	Education was inversely related to RSA in simple models (controlling for Year and item/method), but positively related in the full socio-demographic model, such that samples of high school grads (34.3%) attend at a significantly lower rate than college grads (41.9%)
11	Studies that utilize samples of relatively wealthy participants will yield relatively high attendance estimates.	Supported	RSA was positively related to income, with those making \$25,000-\$49,999 (37.7%) attending less than those making \$50,000-\$74,999 (40.4%) or \$75,000-\$99,999 (47.4%)
12	Studies that utilize samples of primarily employed persons will yield relatively low attendance rates.	Not Supported	RSA for all employed and all non-employed samples are expected to be 47.8% and 30.7%, respectively
12a & 12b	Studies that utilize samples of primarily employed males will yield relatively high attendance rates, while studies that utilize samples of primarily employed *females* will yield relatively low attendance rates.	Not Supported	The interaction between RSA and employment status was significant, but in opposite direction, with employed females and unemployed males reporting higher RSA

Note. RSA = Religious Service Attendance

[a] Indicates an inflation rate, which is computed by dividing the baseline or gold standard rate by the specified item/method rate of RSA.

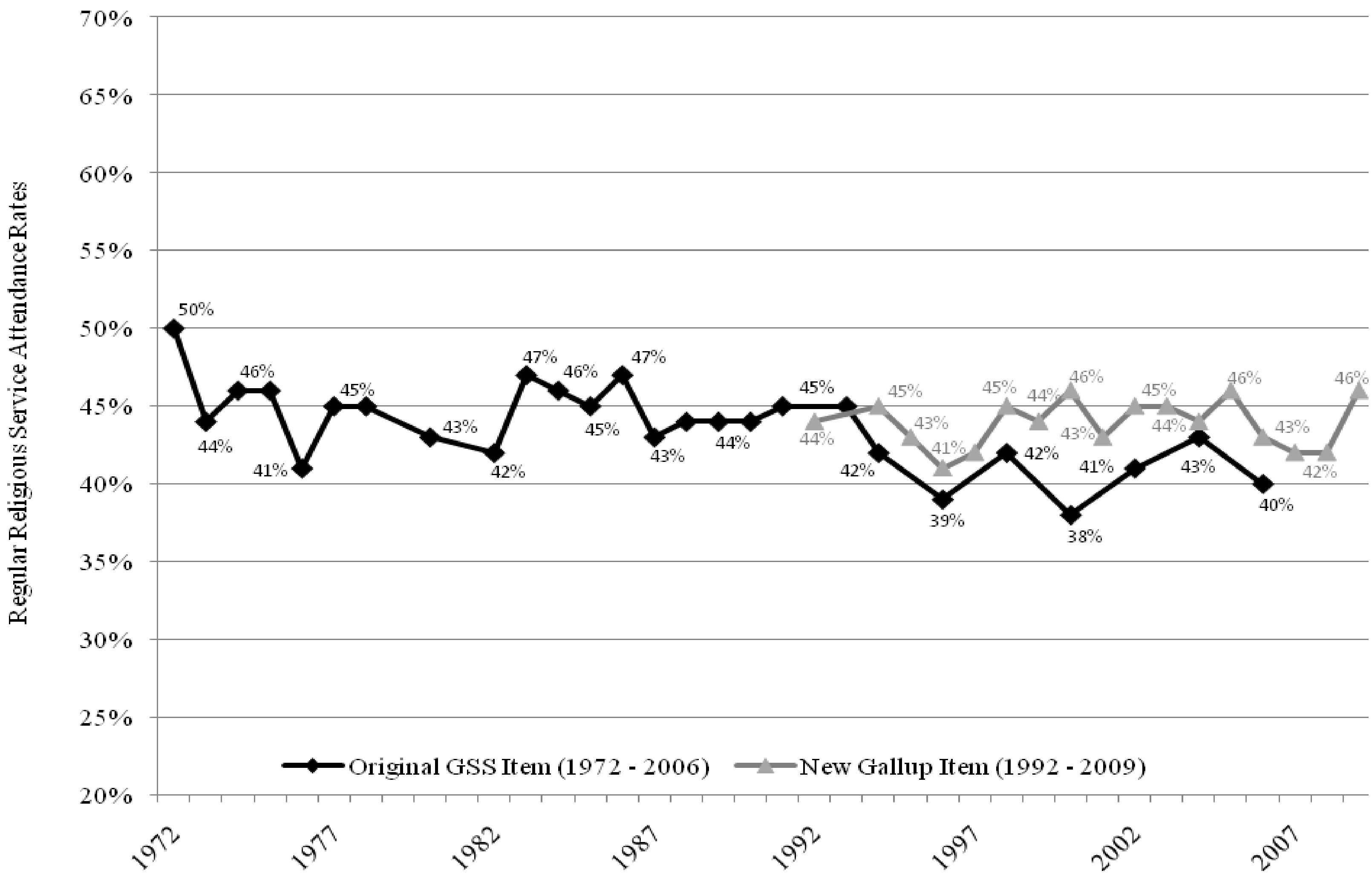

Figure 1. Regular religious service attendance rates yielded by the original GSS item (1972 – 2006) and the new Gallup item (1992 – 2009) (*Sources*: Davis, Smith & Marsden, 2007; F. Newport, personal communication, October 26, 2009).

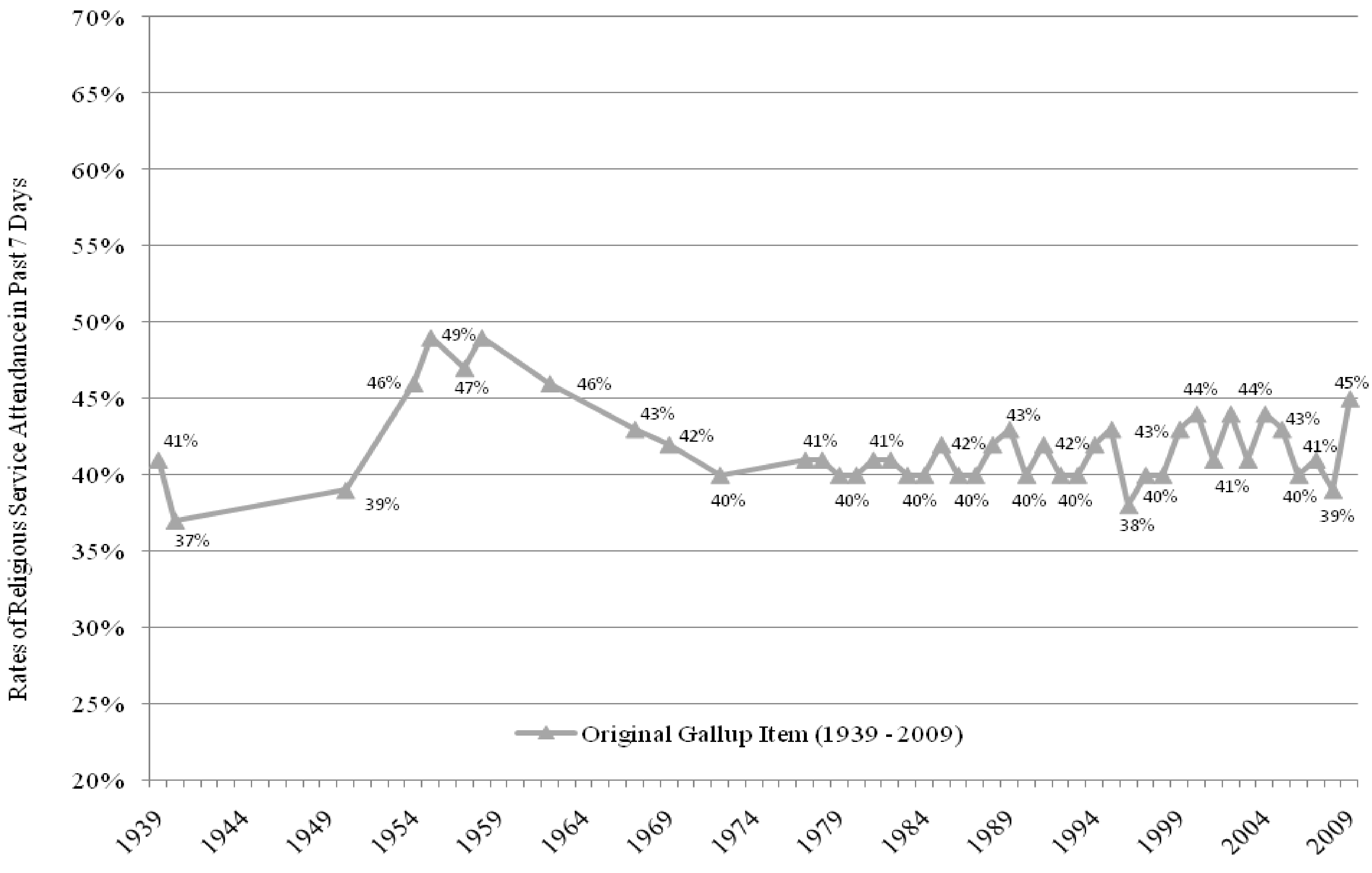

Figure 2. Rates of religious service attendance in the last week yielded by the original Gallup item (1939 – 2009) (*Sources*: Gallup & Jones, 1989; F. Newport, personal communication, October 26, 2009).

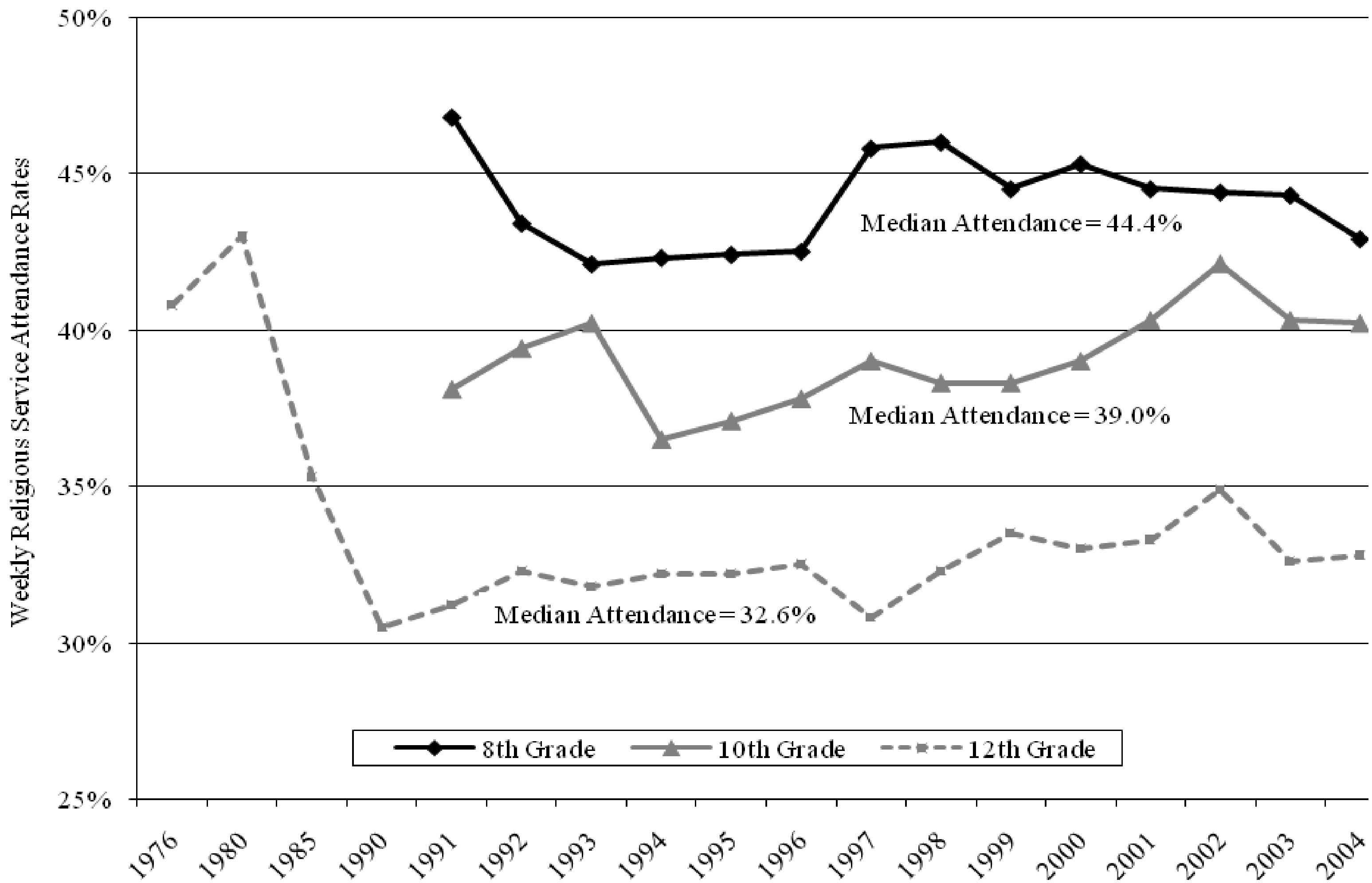

Figure 3. Percentage of Monitoring the Future Survey respondents in 8th, 10th and 12th grades who reported attending religious services at least once a week: 1976 – 2004 (*Source*: Child Trends Databank, 2007).

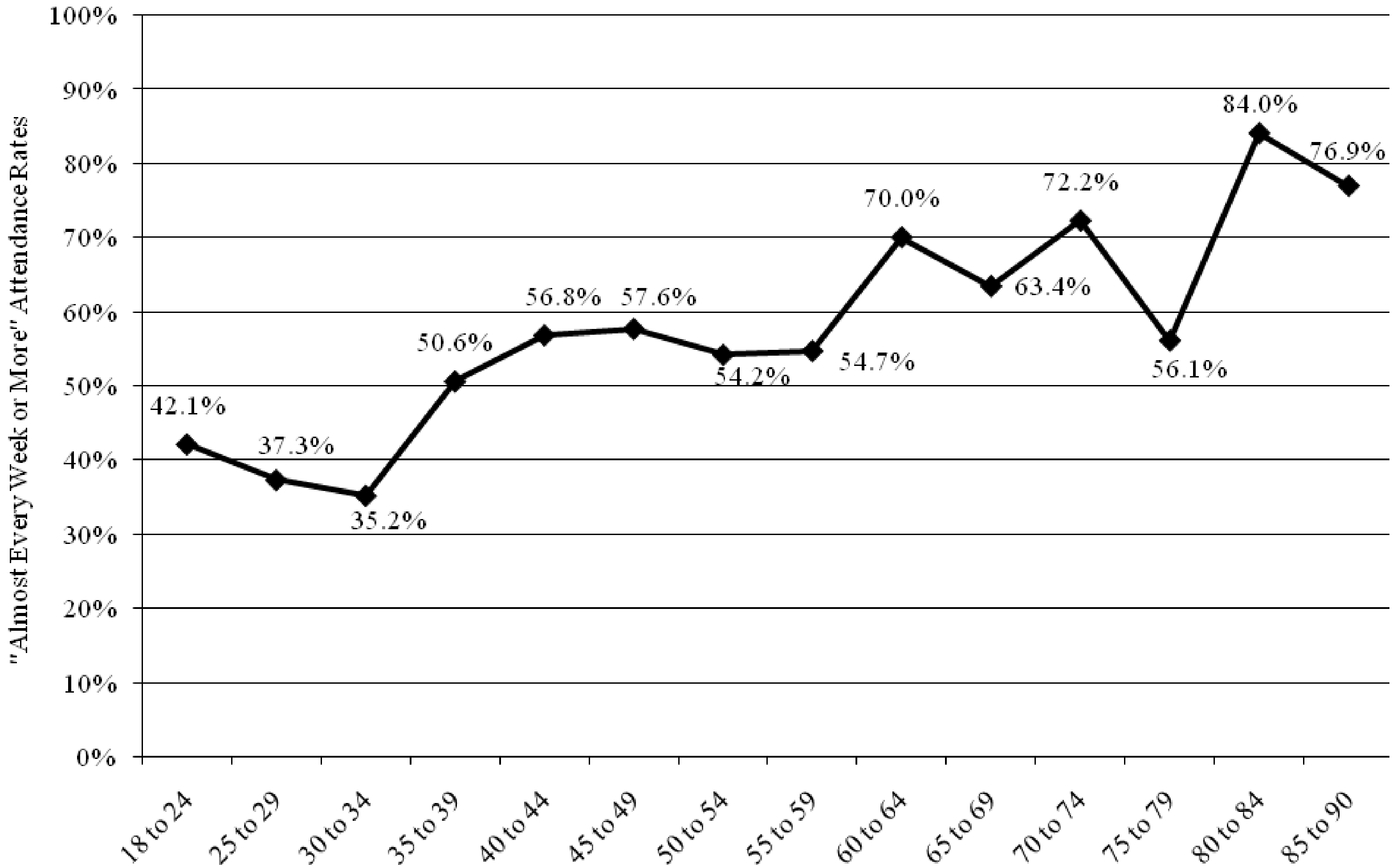

Figure 4. "Almost every week or more" attendance rates across the lifespan (*Source*: 2004 American National Election Study; ANES, 2007).

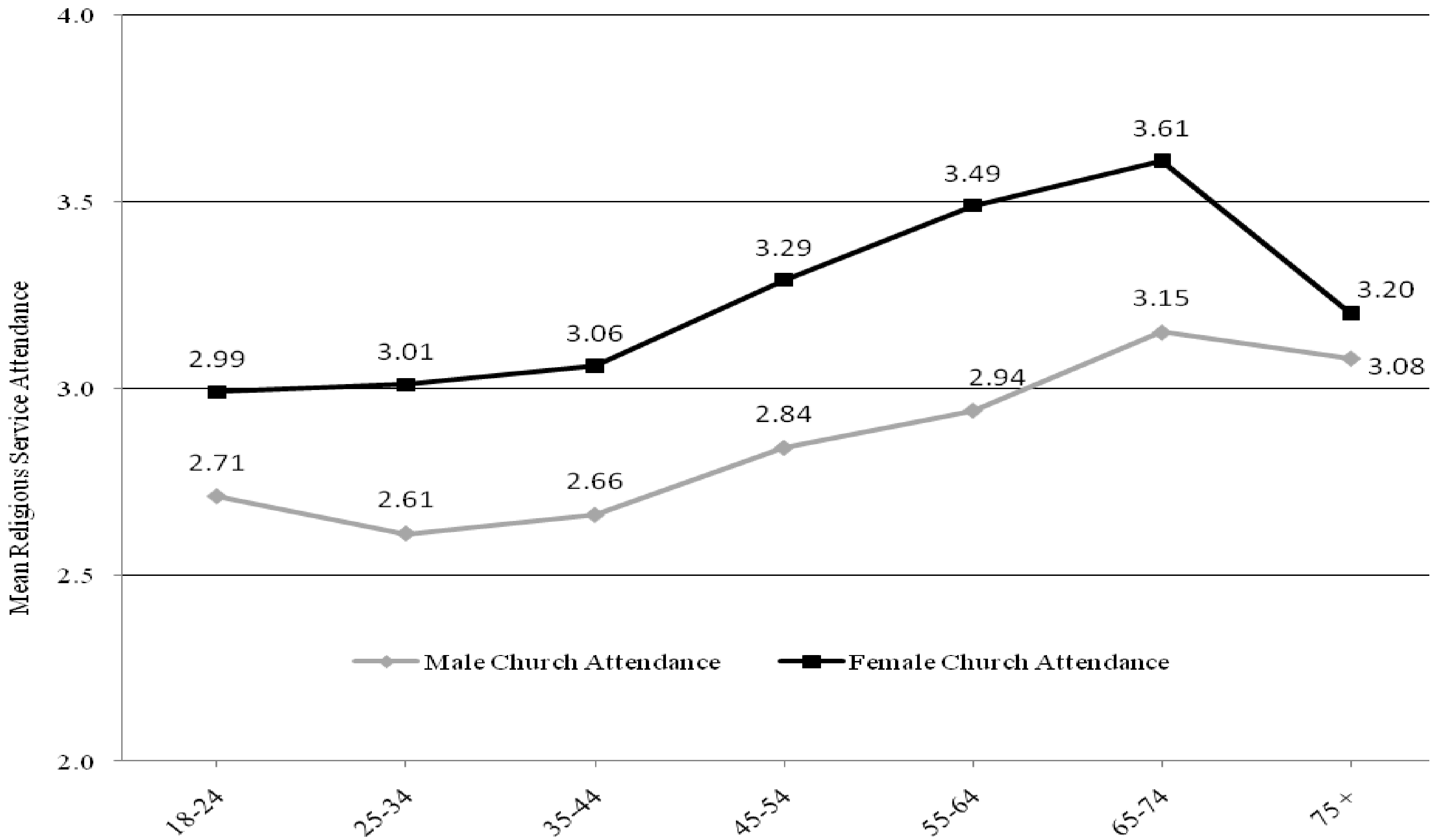

Figure 5. Mean religious service attendance across the lifespan for African Americans responding to the 1979-1980 National Survey of Black Americans item: "How often do you usually attend religious services? 1 = Less than once a year; 2 = A few times a year; 3 = A few times a month; 4 = At least once a week; 5 = Nearly every day (*Source*: Chatters & Taylor, 1989).

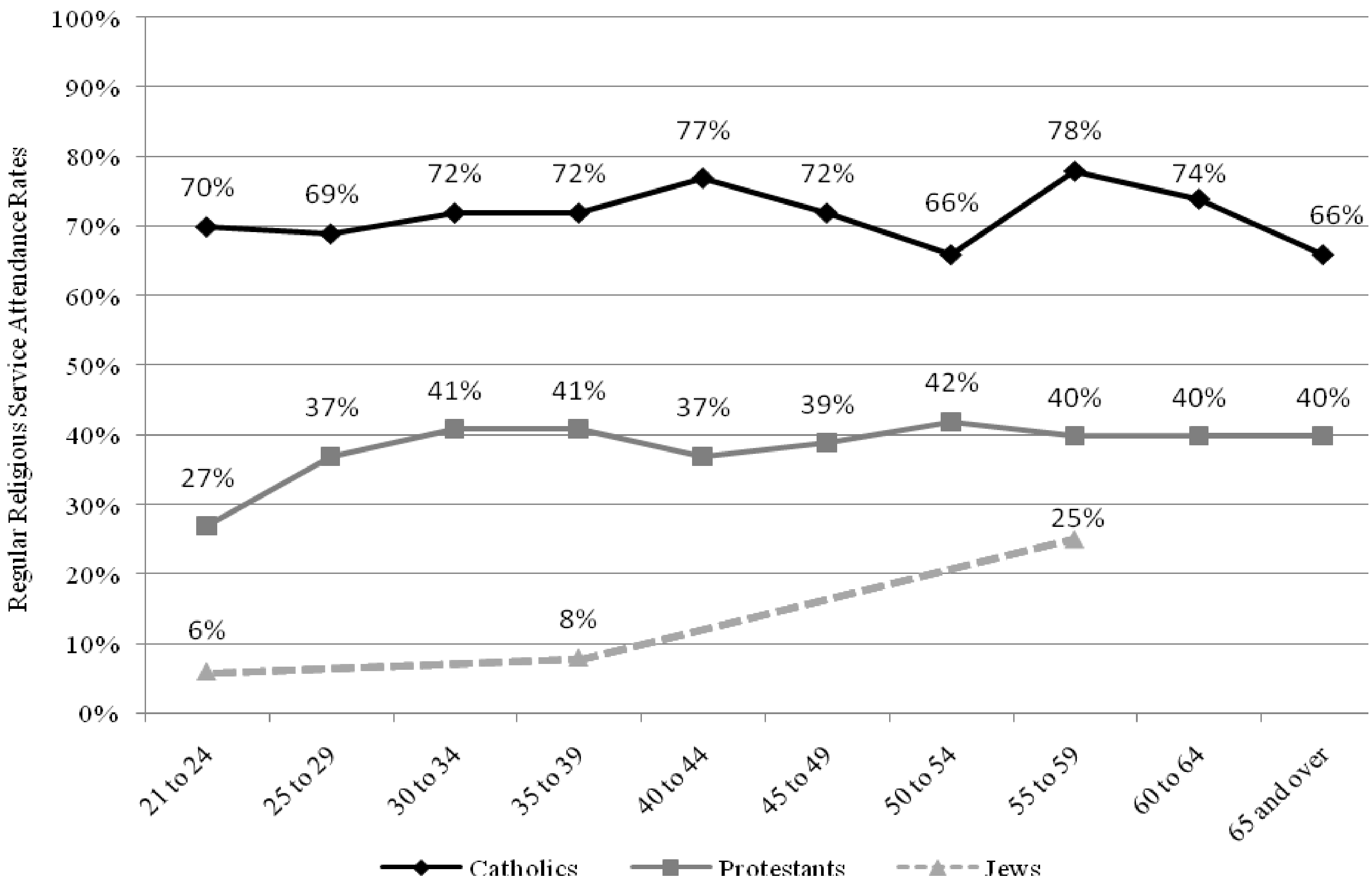

Figure 6. Regular religious service attendance rates across the lifespan for Catholics, Protestants and Jews. Age categories for Jewish respondents were collapsed into the following: 21 to 34 years, 35 to 49 years and 50 years and older (*Source*: Lazerwitz, 1961).

Study name	Statistics for each study		
	Event rate	Lower limit	Upper limit
Study 1	0.411	0.259	0.583
Study 2	0.470	0.230	0.725
Study 3	0.526	0.415	0.634
Study 4	0.349	0.228	0.493
Study 5	0.361	0.222	0.527
Study 6	0.237	0.152	0.350
Study 7	0.423	0.313	0.541
	0.398	0.346	0.451

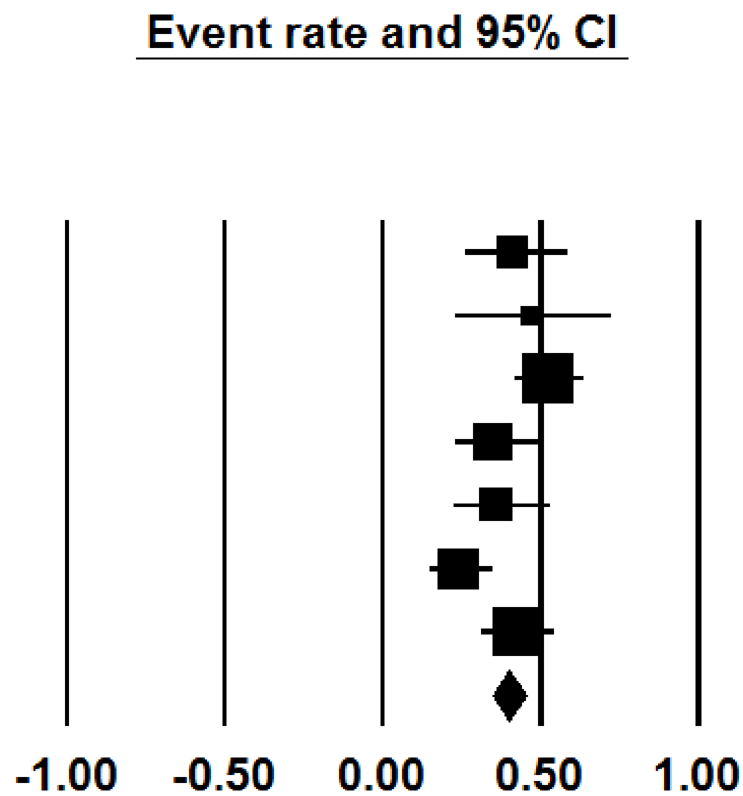

Figure 7. Example forest plot depicting proportions (boxes) and confidence intervals (whiskers) for seven fictitious studies, along with the cumulative grand mean proportion (bottom row, center of diamond) and the associated confidence intervals (width of diamond).

Study name	Statistics for each study		
	Event rate	Lower limit	Upper limit
AHTUS, 1966	0.406	0.335	0.480
AHTUS-W1, 1975-76	0.301	0.250	0.356
AHTUS-W2, 1975-76	0.318	0.274	0.365
AHTUS-W3, 1975-76	0.288	0.242	0.338
AHTUS-W4, 1975-76	0.387	0.294	0.489
AHTUS, 1985a	0.265	0.219	0.316
AHTUS, 1992-94a	0.268	0.245	0.292
AHTUS, 1994-95	0.263	0.208	0.327
AHTUS, 1998	0.268	0.204	0.344
AHTUS, 1999	0.353	0.249	0.473
AHTUS, 2000	0.269	0.166	0.405
AHTUS, 2003	0.279	0.266	0.292
AHTUS, 2004	0.270	0.255	0.285
AHTUS, 2005	0.273	0.257	0.289
AHTUS, 2006	0.280	0.264	0.296
AHTUS, 2007	0.277	0.261	0.294
AHTUS, 2008	0.275	0.259	0.291
AHTUS, 2009	0.277	0.261	0.293
AHTUS, 2010	0.265	0.249	0.280
AHTUS, 2011	0.277	0.261	0.294
AHTUS, 2012	0.271	0.255	0.287
GSS - Orig, 1972	0.476	0.452	0.501
GSS - Orig, 1973	0.419	0.395	0.445
GSS - Orig, 1974	0.432	0.407	0.457
GSS - Orig, 1975	0.428	0.403	0.453
GSS - Orig, 1976	0.406	0.382	0.432
GSS - Orig, 1977	0.423	0.398	0.448
GSS - Orig, 1978	0.415	0.390	0.440
GSS - Orig, 1980	0.417	0.392	0.442
GSS - Orig, 1982	0.419	0.397	0.442
GSS - Orig, 1983	0.440	0.416	0.464
GSS - Orig, 1984	0.444	0.418	0.469
GSS - Orig, 1985	0.429	0.404	0.454
GSS - Orig, 1986	0.441	0.416	0.467
GSS - Orig, 1987	0.431	0.408	0.454
GSS - Orig, 1988	0.408	0.383	0.433
GSS - Orig, 1989	0.415	0.391	0.440
GSS - Orig, 1990	0.416	0.390	0.442
GSS - Orig, 1991	0.418	0.394	0.443
GSS - Orig, 1993	0.411	0.387	0.435
GSS - Orig, 1994	0.388	0.371	0.406
GSS - Orig, 1996	0.368	0.350	0.385
GSS - Orig, 1998	0.382	0.364	0.400
GSS - Orig, 2000	0.355	0.337	0.372
GSS - Orig, 2002	0.378	0.360	0.397
GSS - Orig, 2004	0.400	0.381	0.418
GSS - Orig, 2006	0.378	0.364	0.392
GSS - Orig, 2008	0.375	0.354	0.396
GSS - Orig, 2010	0.366	0.345	0.387
GSS - Orig, 2012	0.372	0.351	0.394
GSS - Orig, 2014	0.349	0.331	0.368
AHTUS, 1985b	0.361	0.223	0.527
AHTUS, 1992-94b	0.267	0.223	0.316

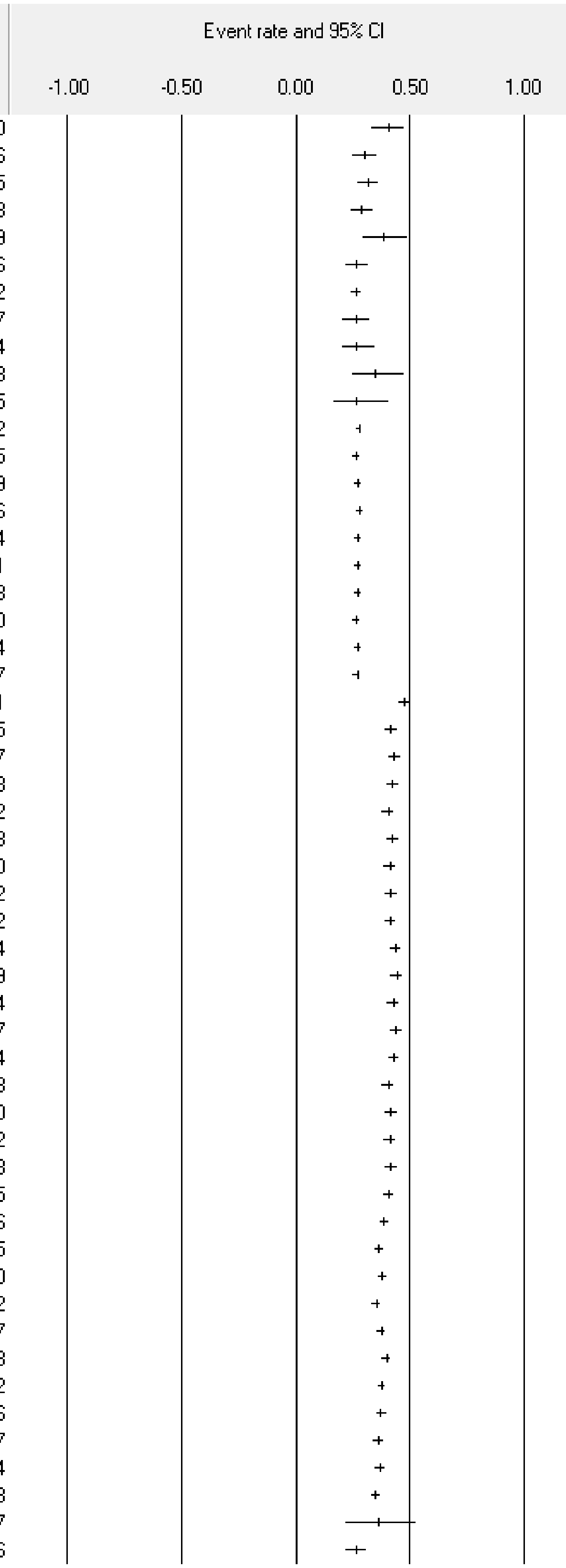

Figure 8. Forest plot of the 271 study observations available for analysis.

Study name	Event rate	Lower limit	Upper limit
WVS, 1981	0.503	0.483	0.524
WVS, 1990	0.512	0.488	0.537
WVS, 2000	0.511	0.477	0.545
NSYR-W1, 2002-03	0.505	0.488	0.522
Hadaway & Marler, 2005a	0.192	0.192	0.192
Hadaway & Marler, 2005b	0.251	0.251	0.252
Hadaway & Marler, 2005c	0.251	0.251	0.252
Hadaway & Marler, 2005d	0.355	0.355	0.356
Hadaway & Marler, 2005e	0.249	0.249	0.250
Hadaway & Marler, 2005f	0.180	0.180	0.180
Gallup - Orig, 1950	0.411	0.385	0.438
Gallup - Orig, 1954	0.470	0.444	0.496
Gallup - Orig_Omit	0.446	0.420	0.471
Gallup - Orig, 1955a	0.476	0.451	0.501
Gallup - Orig, 1955b	0.493	0.468	0.519
Gallup - Orig, 1955c	0.490	0.465	0.515
Gallup - Orig, 1955d	0.489	0.464	0.515
Gallup - Orig, 1956a	0.461	0.439	0.483
Gallup - Orig, 1956b	0.470	0.449	0.491
Gallup - Orig, 1956c	0.476	0.455	0.497
Gallup - Orig, 1956d	0.479	0.454	0.505
Gallup - Orig, 1957	0.509	0.485	0.533
Gallup - Orig_Omit	0.486	0.461	0.512
Gallup - Orig_Omit	0.490	0.465	0.515
Gallup - Orig_Omit	0.472	0.454	0.490
Gallup - Orig_Omit	0.475	0.457	0.493
Gallup - Orig_Omit	0.465	0.449	0.481
Gallup - Orig_Omit	0.505	0.487	0.522
Gallup - Orig_Omit	0.425	0.400	0.450
Gallup - Orig_Omit	0.417	0.392	0.442
Gallup - Orig_Omit	0.417	0.393	0.442
Gallup - Orig_Omit	0.414	0.389	0.439
Gallup - Orig_Omit	0.395	0.371	0.420
Gallup - Orig_Omit	0.388	0.364	0.412
Gallup - Orig_Omit	0.394	0.370	0.419
Gallup - Orig_Omit	0.402	0.378	0.427
Gallup - Orig, 1977	0.408	0.383	0.433
Gallup - Orig, 1978	0.384	0.360	0.409
Gallup - Orig, 1979	0.410	0.385	0.435
Gallup - Orig, 1980	0.494	0.470	0.519
Gallup - Orig, 1981	0.425	0.400	0.450
Gallup - Orig, 1982	0.412	0.387	0.437
Gallup - Orig, 1983	0.376	0.351	0.400
Gallup - Orig, 1984	0.425	0.401	0.450
Gallup - Orig, 1985	0.396	0.372	0.421
Gallup - Orig, 1986	0.393	0.369	0.418
Gallup - Orig, 1987	0.406	0.382	0.430
Gallup - Orig, 1990	0.410	0.383	0.438
Gallup - Orig, 1991	0.393	0.363	0.424
Gallup - Orig, 1992	0.413	0.383	0.444
Gallup - Orig, 1993	0.402	0.372	0.432
Gallup - Orig, 1994	0.388	0.358	0.418
Gallup - New, 1994	0.449	0.419	0.480

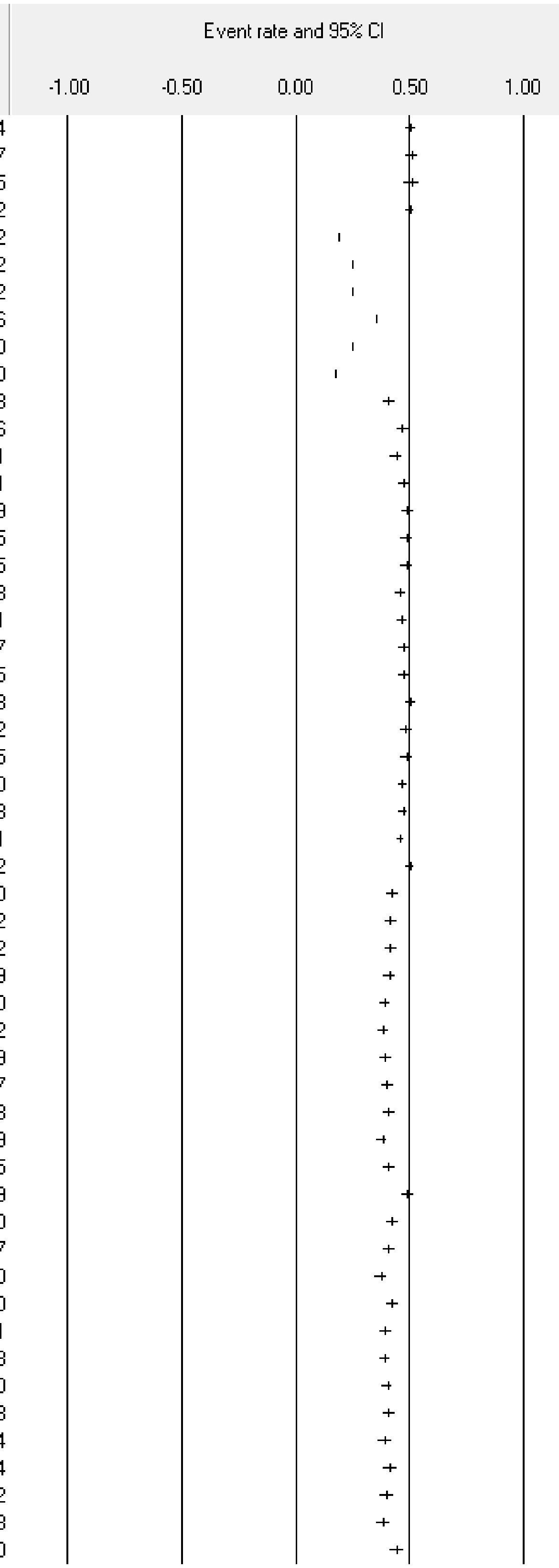

Figure 8. (Continued).

Study name	Statistics for each study		
	Event rate	Lower limit	Upper limit
Gallup - Orig, 1995	0.401	0.371	0.432
Gallup - New, 1995	0.441	0.411	0.472
Gallup - Orig, 1996	0.385	0.356	0.416
Gallup - New, 1996	0.412	0.382	0.443
Gallup - Orig, 1997	0.433	0.403	0.464
Gallup - New, 1997	0.449	0.419	0.480
Gallup - Orig, 1998	0.427	0.397	0.458
Gallup - New, 1998	0.485	0.454	0.516
Gallup - Orig, 1999	0.452	0.422	0.483
Gallup - New, 1999	0.493	0.462	0.523
Gallup - Orig, 2000	0.446	0.416	0.477
Gallup - New, 2000	0.479	0.449	0.510
Gallup - Orig, 2001	0.473	0.442	0.504
Gallup - Orig, 2002	0.432	0.402	0.463
Gallup - New, 2002	0.472	0.441	0.503
Gallup - Orig, 2003	0.438	0.408	0.469
Gallup - New, 2003	0.465	0.435	0.496
Gallup - Orig, 2004	0.434	0.404	0.465
Gallup - New, 2004	0.466	0.435	0.497
Gallup - Orig, 2005	0.447	0.416	0.478
Gallup - New, 2005	0.489	0.458	0.520
Gallup - Orig, 2007	0.443	0.413	0.474
Gallup - New, 2007	0.477	0.447	0.508
Gallup - Orig, 2008	0.411	0.381	0.441
Gallup - New, 2008	0.463	0.432	0.494
Gallup - Orig, 2009	0.481	0.450	0.512
Gallup - New, 2009	0.494	0.463	0.525
Gallup - Orig, 2010	0.432	0.402	0.463
Gallup - New, 2010	0.484	0.454	0.515
Gallup - Orig, 2011	0.392	0.362	0.422
Gallup - New, 2011	0.423	0.392	0.453
Gallup - Orig, 2012	0.425	0.395	0.456
Gallup - New (Mosque),	0.463	0.433	0.494
Gallup - Orig, 2015	0.375	0.346	0.405
Gallup - New (Mosque),	0.398	0.369	0.429
BWHS, 2005	0.593	0.588	0.598
Gallup - New (Mosque),	0.477	0.451	0.503
Gallup - New (Mosque),	0.444	0.414	0.475
Gallup - New (Mosque),	0.442	0.411	0.473
Gallup - New (Mosque),	0.437	0.407	0.468
Gallup - New (Mosque),	0.443	0.412	0.474
Gallup - New (Mosque),	0.399	0.369	0.429
Gallup - New (Mosque),	0.417	0.387	0.449
NELS, 1990	0.560	0.550	0.569
NELS HS Dropouts, 1990	0.302	0.254	0.355
NELS, 1992	0.475	0.465	0.485
NELS HS Dropouts, 1992	0.286	0.261	0.313
HS&Beyond Senior Survey,	0.506	0.500	0.512
HS&Beyond Sophomore	0.537	0.531	0.543
ACLS-W1, 1986	0.519	0.502	0.535
ACLS-W2, 1989	0.521	0.502	0.539
ACLS-W3, 1994	0.523	0.503	0.543
ACLS-W4, 2002	0.534	0.510	0.558

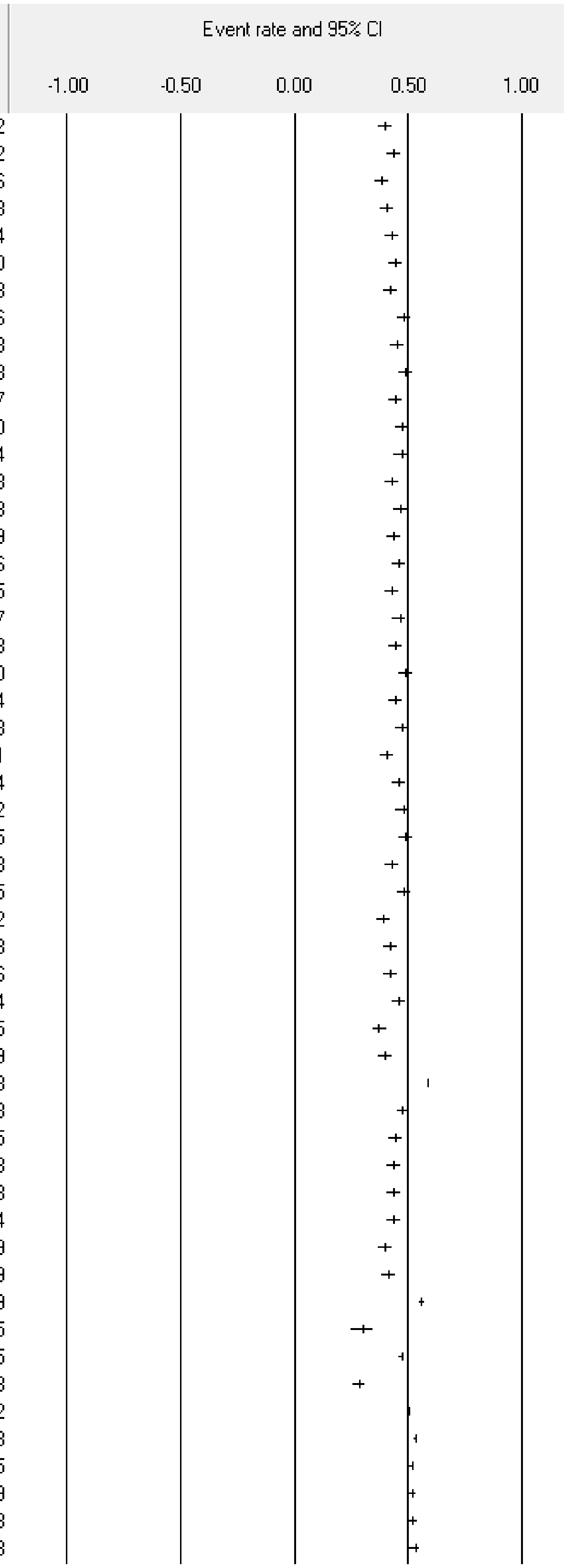

Figure 8. (Continued).

Study name	Statistics for each study		
	Event rate	Lower limit	Upper limit
ACLS-W5, 2011	0.490	0.463	0.517
GSS Branching, Pt1, 1996	0.311	0.282	0.341
GSS Branching, Pt2, 1996	0.284	0.255	0.314
Chaves & Cavendish, 1994	0.261	0.261	0.262
Chaves & Cavendish, 1994	0.265	0.264	0.265
Chaves & Cavendish, 1994	0.322	0.322	0.322
Chaves & Cavendish, 1994	0.214	0.214	0.215
Hadaway, Marler & Chaves,	0.197	0.193	0.200
Presser & Stinson, 1998	0.290	0.267	0.314
Presser & Stinson, 1998	0.420	0.350	0.494
Presser & Stinson, 1998a	0.270	0.206	0.346
Presser & Stinson, 1998b	0.310	0.258	0.367
ATUS, 2003	0.270	0.257	0.283
ATUS, 2004	0.268	0.253	0.284
ATUS, 2005	0.263	0.247	0.280
NSBA, 1980	0.586	0.563	0.607
MTF-12th, 1976	0.452	0.445	0.460
MTF-12th, 1977	0.453	0.445	0.460
MTF-12th, 1978	0.454	0.447	0.462
MTF-12th, 1979	0.473	0.466	0.481
MTF-12th, 1980	0.480	0.472	0.488
MTF-12th, 1981	0.454	0.447	0.462
MTF-12th, 1982	0.441	0.434	0.449
MTF-12th, 1983	0.466	0.458	0.474
MTF-12th, 1984	0.435	0.427	0.442
MTF-12th, 1985	0.421	0.413	0.428
MTF-12th, 1986	0.414	0.406	0.422
MTF-12th, 1987	0.380	0.373	0.388
MTF-12th, 1988	0.377	0.370	0.385
MTF-12th, 1989	0.379	0.371	0.386
MTF-12th, 1990	0.360	0.352	0.367
MTF-12th, 1991	0.362	0.354	0.369
MTF-12th, 1992	0.382	0.374	0.389
MTF-8th, 1992	0.493	0.483	0.503
MTF-10th, 1992	0.448	0.437	0.460
MTF-8th, 1993	0.488	0.477	0.498
MTF-10th, 1993	0.457	0.446	0.468
MTF-12th, 1993	0.369	0.350	0.388
MTF-8th, 1994	0.492	0.481	0.502
MTF-10th, 1994	0.430	0.419	0.441
MTF-12th, 1994	0.384	0.376	0.392
MTF-8th, 1995	0.482	0.471	0.492
MTF-10th, 1995	0.434	0.423	0.444
MTF-12th, 1995	0.374	0.366	0.382
MTF-8th, 1996	0.490	0.480	0.501
MTF-10th, 1996	0.441	0.430	0.452
MTF-12th, 1996	0.376	0.368	0.384
MTF-8th, 1997	0.512	0.498	0.526
MTF-10th, 1997	0.452	0.437	0.467
MTF-12th, 1997	0.370	0.362	0.379
MTF-8th, 1998	0.519	0.505	0.533
MTF-10th, 1998	0.448	0.432	0.463
MTF-12th, 1998	0.383	0.374	0.391

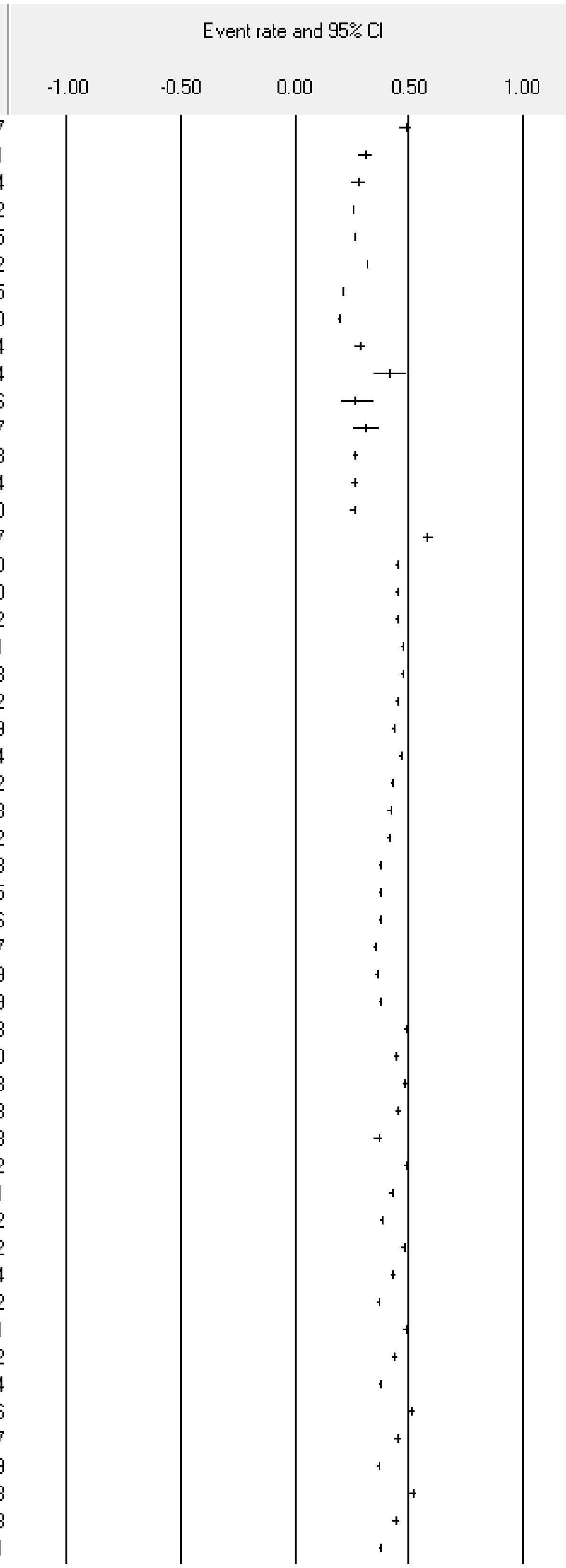

Figure 8. (Continued).

Study name	Statistics for each study		
	Event rate	Lower limit	Upper limit
MTF-8th, 1999	0.502	0.487	0.517
MTF-10th, 1999	0.450	0.434	0.466
MTF-12th, 1999	0.406	0.397	0.415
MTF-8th, 2000	0.509	0.494	0.524
MTF-10th, 2000	0.455	0.439	0.471
MTF-12th, 2000	0.410	0.400	0.419
MTF-8th, 2001	0.509	0.494	0.524
MTF-10th, 2001	0.460	0.444	0.475
MTF-12th, 2001	0.399	0.389	0.408
MTF-8th, 2002	0.489	0.474	0.505
MTF-10th, 2002	0.467	0.451	0.483
MTF-12th, 2002	0.397	0.388	0.407
MTF-8th, 2003	0.475	0.460	0.489
MTF-10th, 2003	0.467	0.452	0.483
MTF-12th, 2003	0.385	0.376	0.394
MTF-8th, 2004	0.480	0.465	0.494
MTF-10th, 2004	0.448	0.433	0.463
MTF-12th, 2004	0.385	0.376	0.394
MTF-8th, 2005	0.467	0.452	0.481
MTF-10th, 2005	0.433	0.418	0.449
MTF-12th, 2005	0.388	0.379	0.397
MTF-8th, 2006	0.469	0.455	0.484
MTF-10th, 2006	0.421	0.407	0.436
MTF-12th, 2006	0.378	0.369	0.387
MTF-8th, 2007	0.493	0.477	0.508
MTF-10th, 2007	0.428	0.413	0.443
MTF-12th, 2007	0.359	0.350	0.367
MTF-8th, 2008	0.475	0.460	0.490
MTF-10th, 2008	0.411	0.396	0.426
MTF-12th, 2008	0.352	0.342	0.361
MTF-8th, 2009	0.434	0.419	0.449
MTF-10th, 2009	0.427	0.412	0.442
MTF-12th, 2009	0.356	0.347	0.366
MTF-8th, 2010	0.428	0.414	0.443
MTF-10th, 2010	0.385	0.371	0.400
MTF-12th, 2010	0.352	0.344	0.361
MTF-8th, 2011	0.455	0.440	0.470
MTF-10th, 2011	0.398	0.382	0.413
MTF-12th, 2011	0.344	0.335	0.353
MTF-8th, 2012	0.445	0.436	0.454
MTF-10th, 2012	0.405	0.396	0.414
MTF-12th, 2012	0.361	0.352	0.371
MTF-8th, 2013	0.440	0.431	0.449
MTF-10th, 2013	0.398	0.388	0.408
MTF-12th, 2013	0.361	0.351	0.371
MTF-8th, 2014	0.446	0.437	0.455
MTF-10th, 2014	0.387	0.377	0.396
MTF-12th, 2014	0.343	0.333	0.352
MTF-8th, 2015	0.437	0.428	0.447
MTF-10th, 2015	0.378	0.370	0.387
MTF-12th, 2015	0.339	0.330	0.348

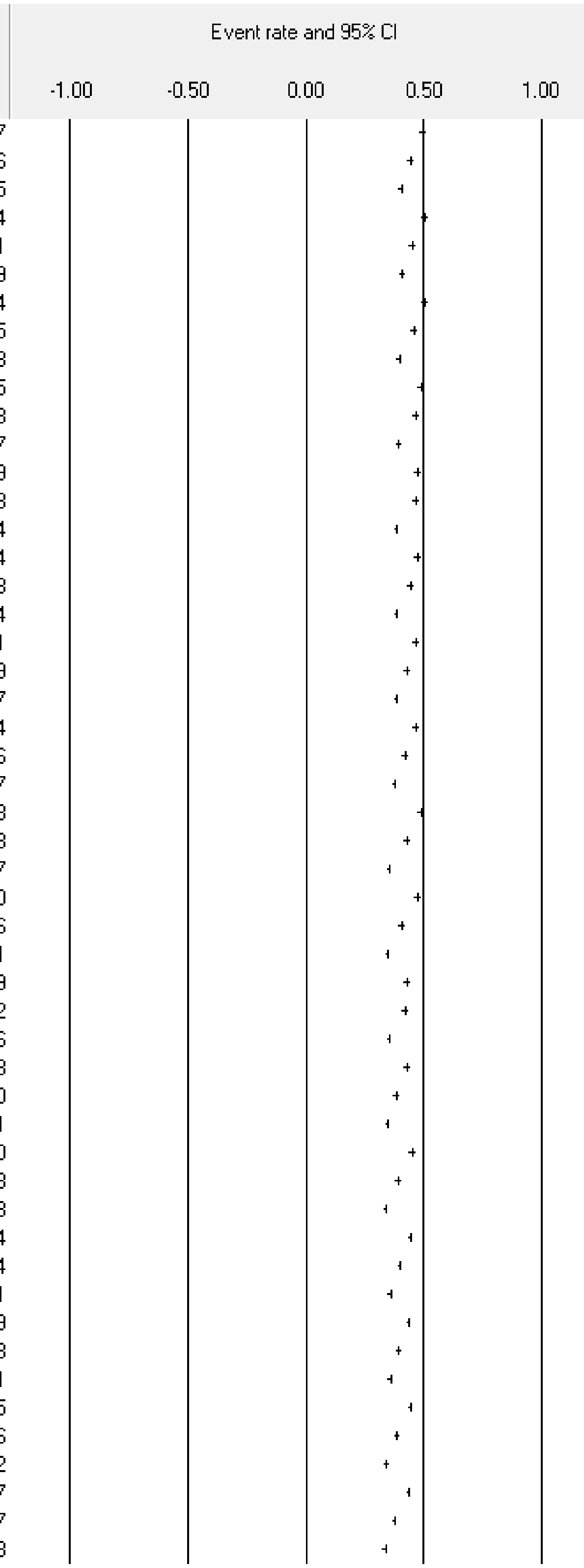

Figure 8. (Continued).

Study name	Statistics for each study		
	Event rate	Lower limit	Upper limit
BRS-W1, 2005	0.449	0.425	0.472
BRS-W2, 2007	0.419	0.395	0.443
BRS-W3, 2010	0.420	0.396	0.443
ANES, 2000	0.424	0.402	0.447
ANES, 2002	0.457	0.432	0.482
ANES, 2004	0.399	0.372	0.427
ANES, 2008	0.429	0.400	0.458
Religious Landscape	0.456	0.451	0.462

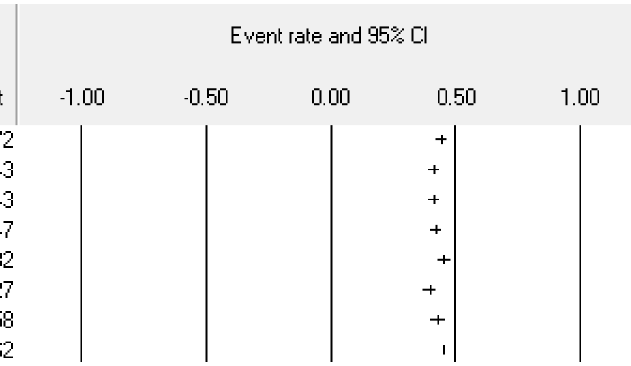

Figure 8. (Continued).

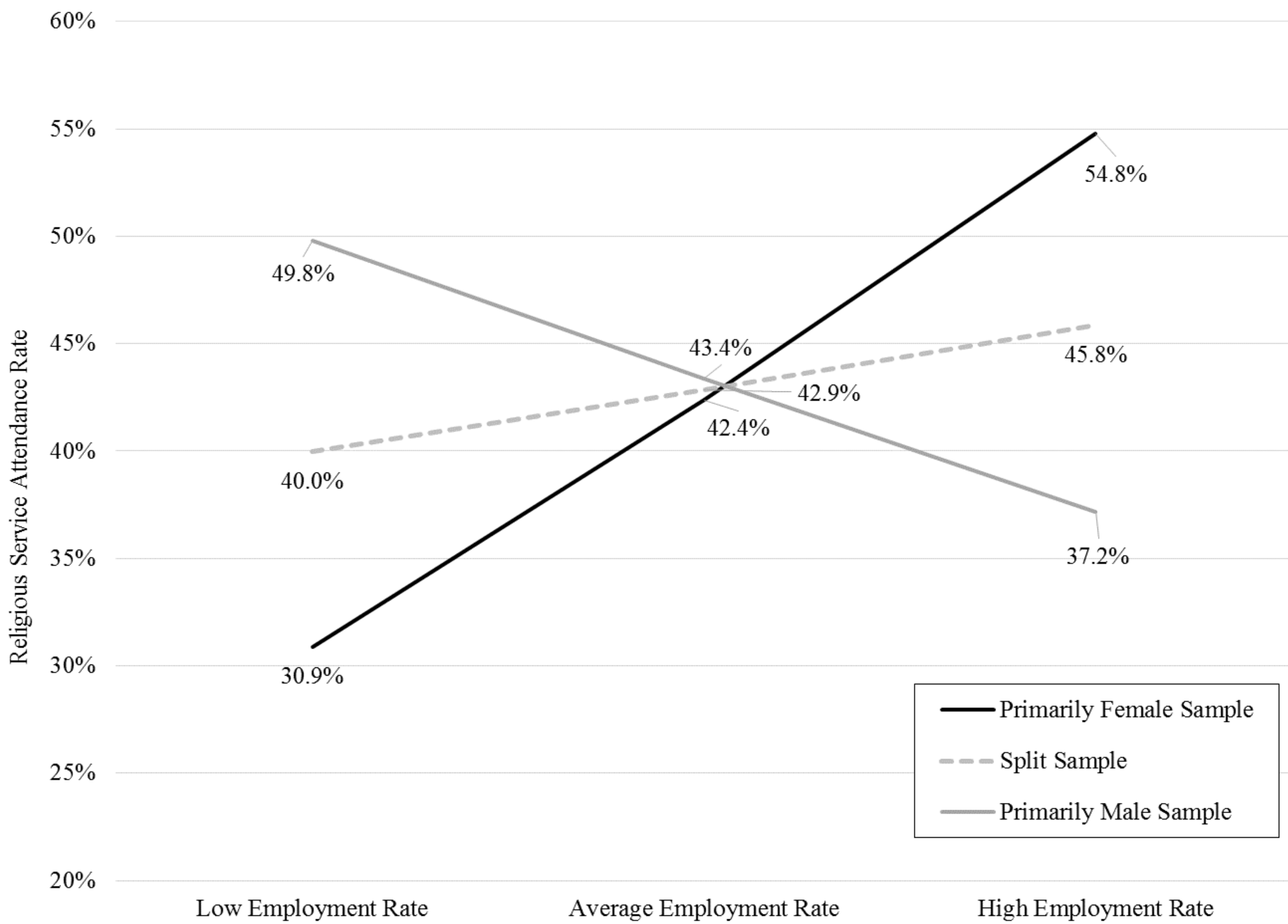

Figure 9. The interaction effect of gender and employment status on religious service attendance.

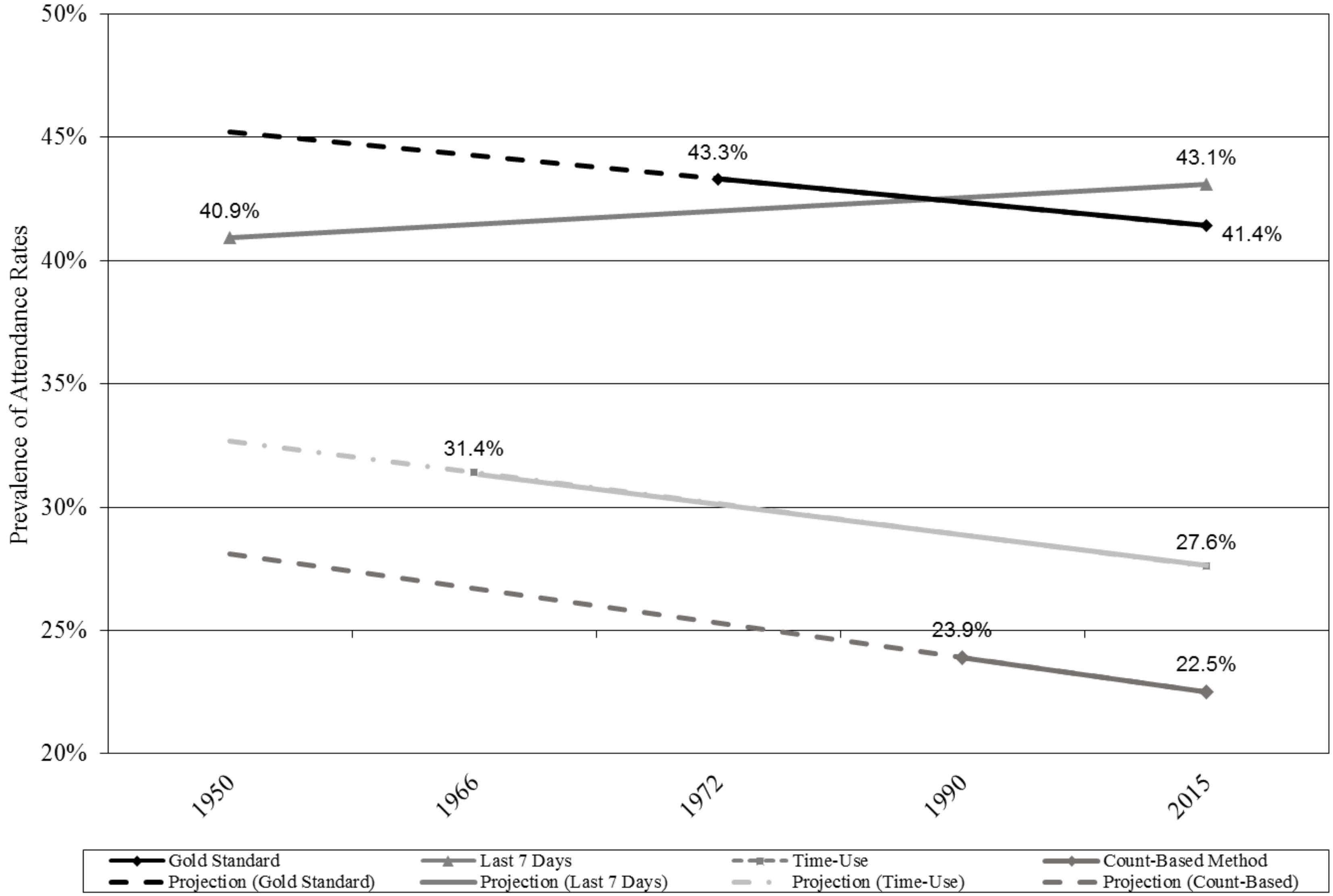

Figure 10. Prevalence of Attendance for Four Item-Types and Methodologies (With Projections From 1950 to 2015).

REFERENCES

*Aarts, O., Need, A., Te Grotenhuis, M., & de Graaf, N. D. (2008). Does belonging accompany believing? Correlations and trends in Western Europe and North America between 1981 and 2000. *Review of Religious Research*, 50 (1), 16 – 34.

Adlaf, E. M., & Smart, R. G. (1985). Drug use and religious affiliation, feelings and behavior. *British Journal of Addiction, 80*, 163 – 171.

Albrecht, S. L., Cornwall, M., & Cunningham, P. H. (1988). Religious leave-taking: Disengagement and disaffiliation among Mormons. In D. G. Bromley (Ed.), *Falling from the Faith: Causes and Consequences of Religious Apostasy* (pp. 62 – 80). Thousand Oaks, CA: Sage Publications.

Allport, G. W. (1950). The *Individual and His Religion*. NY: Macmillan.

Allport, G. W. (1987). The religious context of prejudice. *Journal for the Scientific Study of Religion, 5* (3), 447 – 457.

Allport, G. W., & Ross, J. M. (1967). Personal religious orientation and prejudice. *Journal of Personality and Social Psychology, 5*, 432 – 443.

Altemeyer, B. (2003). Why do religious fundamentalists tend to be prejudiced? *The International Journal for the Psychology of Religion, 13* (1), 17 – 28.

The American Heritage College Dictionary, 3rd Edition (2000). Definition of *Few* and *Several.* Houghton Mifflin Company: Boston, MA.

American National Election Study (ANES) (2007). *2004 American National Election Study*. Retrieved July 18, 2007, from http://sda.berkeley.edu/cgi-bin/hsda?harcsda+nes2004p.

Anders, S. F. (1955). Religious behavior of church families. *Journal of Marriage and Family Living, 17*, 54 – 57.

Argue, A., Johnson, D. R., & White, L. K. (1999). Age and religiosity: Evidence from a three-wave panel analysis. *Journal for the Scientific Study of Religion*, *38* (3), 423 – 435.

Argyle, M. (2000). *Psychology and Religion: An Introduction*. New York, NY: Routledge.

Argyle, M., & Beit-Hallahmi, B. (1975). *The Social Psychology of Religion*. London, England: Routledge & Kegan Paul.

Association of Religion Data Archives (2016). *Religion Family Trees.* Retrieved November 22, 2016, from http://www.thearda.com/denoms/families/trees/index.asp.

Auger, C. P. (1998). *Information Sources in Grey Literature. Guides to Information Sources* (4th Ed.). London: Bowker Saur.

Azzi, C., & Ehrenberg, R. (1975). Household allocation of time and church attendance. *Journal of Political Economy*, *83* (1), 27 – 56.

*Bachman, J. G., Johnston, L. D., & O'Malley, P. M. (2008). *Monitoring the future: A continuing study of the lifestyles and values of youth, 1976 – 1992: Concatenated core file* [Data file and code book]. ICPSR06227-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor], 2008-11-24. http://doi.org/10.3886/ICPSR06227.v2

Back, C. W., & Bourque, L. B. (1970). Can feelings be enumerated? *Behavioral Science, 15*, 487 – 496.

*Bader, C. D., Froese, P., Johnson, B., Mencken, F. C., & Stark, R. (2005). *Baylor Religion Survey, Wave I (2005)*. Waco, TX: Baylor Institute for Studies of Religion [producer]. Retrieved from www.thearda.com

*Bader, C. D., Froese, P., Johnson, B., Mencken, F. C., & Stark, R. (2007). *Baylor Religion Survey, Wave II (2007)*. Waco, TX: Baylor Institute for Studies of Religion [producer]. Retrieved from www.thearda.com

*Bader, C. D., Froese, P., Johnson, B., Mencken, F. C., & Stark, R. (2005). *Baylor Religion Survey, Wave III (2010)*. Waco, TX: Baylor Institute for Studies of Religion [producer]. Retrieved from www.thearda.com

*Bader, C. D., Mencken, F. C., & Froese, P. (2007). American piety 2005: Content and methods of the Baylor Religion Survey. *Journal for the Scientific Study of Religion, 46* (4), 447 – 463.

Bahr, H. M. (1970). Aging and religious disaffiliation. *Social Forces, 49* (1), 59 – 71.

Bainbridge, W. S. (2005). Atheism. *Interdisciplinary Journal of Research on Religion*, *1*, 2 – 26.

Baldacchino, D., & Draper, P. (2001). Spiritual coping strategies: A review of the nursing research literature. Journal *of Advanced Nursing, 34* (6), 833 – 841.

Baldwin, J. D., & Baldwin, J. I. (1988). Factors affecting AIDS-related sexual risk-taking behavior among college students. *The Journal of Sex Research*, *25* (2), 181 – 196.

Bangert-Drowns, R. L. (1986). Review of developments in meta-analysis method. *Psychological Bulletin, 99*, 388-399.

Barna, G. (1991). *What Americans Believe: An Annual Survey of Values and Religious Views in the United States*. Ventura, CA: Regal Books.

Barna, G. (2002). *Grow Your Church From the Outside: Understanding the Unchurched and how to Reach Them*. Ventura, CA: Regal.

Barna Research Group (2000). *Women are the Backbone of the Christian Congregations in America*. Retrieved July 21, 2007, from www.barna.org/FlexPage.aspx?Page=BarnaUpdate&BarnaUpdateID=47.

Barna Research Group (2007). *Church Attendance*. Retrieved July 21, 2007, from www.barna.org/FlexPage.aspx?Page=Topic&TopicID=10.

Batson, C. D., Schoenrade, P., & Ventis, W. L. (1993). *Religion and the Individual: A Social-Psychological Perspective*. NY: Oxford University Press.

Baylor Religion Survey (2005). *The Baylor Religion Survey*. Baylor Institute for Studies of Religion, Waco, TX.

Bax, L., Ikeda, N., Fukui, N., Yaju, Y., Tsuruta, H., & Moons, K. G. M. (2008). More than numbers: The power of graphs in meta-analysis. *American Journal of Epidemiology, 169* (2), 249 – 255. doi 10.1093/aje/kwn340

Beeghley, L., Van Velsor, E., & Bock, E. W. (1981). The correlates of religiosity among Black and White Americans. *The Sociological Quarterly, 22*, 403 – 412.

Bem, S. L. (1981a). Gender schema theory: A cognitive account of sex typing. *Psychological Review, 88* (4), 354 – 364.

Bem, S. L. (1981b). *Bem Sex Role Inventory: Professional Manual*. Palo Alto, CA: Consulting Psychological Press.

Bem, S. L. (2010). *Bem Sex-Role Inventory: A Measure of Androgyny and Gender Role*. Retrieved July 27, 2010, from http://www.mindgarden.com/products/bemss.htm.

Benjamins, M. R., Musick, M. A., Gold, D. T., & George, L. K. (2003). Age-related declines in activity level: The relationship between chronic illness and religious activities. *Journal of Gerontology, 58B* (6), S377 – S385.

Benson, P., & Spilka, B. (1973). God image as a function of self-esteem and locus of control. *Journal for the Scientific Study of Religion, 12* (3), 297 – 310.

Bergan, A. (2000). Religiosity and life satisfaction. *Activities, Adaptation & Aging, 24* (3), 23 – 34.

Berger, P. L. (1967). *The Sacred Canopy*. Garden City, NY: Doubleday.

Booth, A., Johnson, D. R., Branaman, A., & Sica, A. (1995). Belief and behavior: Does religion matter in today's marriage? *Journal of Marriage and the Family, 57*, 661 – 671.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive Meta-Analysis, Version 3.3* [computer software]. Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis.* West Sussex, United Kingdom: John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2015). *Regression in Meta-Analysis.* Draft report retrieved from: https://www.meta-analysis.com/pages/cma_manual.php

Bouma, G. D. (1979). The real reason one conservative church grew. *Review of Religious Research, 20*, 127 – 137.

Bourjolly, J. N. (1998). Differences in religiousness among black and white women with breast cancer. *Social Work in Health Care, 28* (1), 21 – 39.

Briggs, C. S., Malek, S., Davis, A., Davis, E., & Lorentz, K. (2009). *Evaluation of the SALUKI Advantage freshman interest group program (2001 – 2005).* Carbondale, IL: Southern Illinois University Carbondale, Office of Academic Initiatives.

Brinkerhoff, M. B., & Mackie, M. M. (1993). Casting off the bonds of organized religion: A religious-careers approach to the study of apostasy. *Review of Religious Research, 34* (3), 235 – 258.

Brody, G. H., Stoneman, Z., & Flor, D. (1996). Parental religiosity, family processes and youth competence in rural, two-parent African American families. *Developmental Psychology, 32* (4), 696 – 706.

Brown, L. B. (1994). *The Human Side of Prayer: The Psychology of Praying*. Birmingham, AL: Religious Education Press.

Burchinal, L. G. (1959). Some social status criteria and church membership and church attendance. *Journal of Social Psychology, 49*, 53 – 64.

Bureau of Labor Statistics (2017). *Databases, Tables & Calculators by Subject: CPI Inflation Calculator*. Retrieved from https://www.bls.gov/data/inflation_calculator.htm.

*Burns, N., Kinder, D. R., Rosenstone, S. J., & Sapiro, V. (2000). *American National Election Studies, 2000*. Waco, TX: Baylor Institute for Studies of Religion [producer]. Retrieved from www.thearda.com

*Burns, N., Kinder, D. R., Rosenstone, S. J., & Sapiro, V. (2002). *American National Election Studies, 2002*. Waco, TX: Baylor Institute for Studies of Religion [producer]. Retrieved from www.thearda.com

*Burns, N., Kinder, D. R., Rosenstone, S. J., & Sapiro, V. (2004). *American National Election Studies, 2004*. Waco, TX: Baylor Institute for Studies of Religion [producer]. Retrieved from www.thearda.com

*Burns, N., Kinder, D. R., Rosenstone, S. J., & Sapiro, V. (2008). *American National Election Studies, 2008*. Waco, TX: Baylor Institute for Studies of Religion [producer]. Retrieved from www.thearda.com

*Cable News Network and USA Today (1995a). *Gallup/CNN/USA Today Poll: Budget Deficit, Medicare, Religion, May, 1995* [Data file and code book]. USAIPOCNUS1995-5001020. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Cable News Network and USA Today (1995b). *Gallup/CNN/USA Today Poll: Clinton vs. Congress, Aug, 1995* [Data file and code book]. USAIPOCNUS1995-9508010. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Cable News Network and USA Today (1996). *Gallup/CNN/USA Today Poll: 1996 Election, Dec, 1995* [Data file and code book]. USAIPOCNUS1995-9512023. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Cable News Network and USA Today (1997). *Gallup/CNN/USA Today Poll: UPS Strike/Religion, Aug, 1997* [Data file and code book]. USAIPOCNUS1997-9708018. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

Cadge, W., & Ecklund, E. H. (2006). Religious service attendance among immigrants: Evidence from the New Immigrant Survey—Pilot. *American Behavioral Scientist, 49* (11), 1574 – 1595.

Call, V. R. A., & Heaton, T. B. (1997). Religious influence on marital stability. *Journal for the Scientific Study of Religion, 36* (3), 382 – 392.

Callahan, S. (2001). Spiritual symptoms: Does religion make you crazy? *Commonweal, April 20*, 7 – 8.

Callender, J. C., & Osburn, H. G. (1980). Development and test of a new model for validity generalization. *Journal of Applied Psychology, 65*, 543 – 558.

Cameron, P. (1969). Valued aspects of religion to Negroes and Whites. *Proceedings of the 77th Annual Convention of the American Psychological Association, 4*, 741 – 742.

Campbell, R. A., & Curtis, J. E. (1994). Religious involvement across societies: Analysis for alternative measures in national surveys. *Journal for the Scientific Study of Religion, 33* (3), 219 – 229.

Campbell, T. C., & Fukuyama, Y. (1970). *The Fragmented Layman: An Empirical Study of Lay Attitudes*. Philadelphia, PA: Pilgrim Press.

Carroll, J. W., & Roozen, D. A. (1975). *Religious Participation in American Society: An Analysis of Social and Religious Trends and Their Interaction*. Hartford, CT: Hartford Seminary Foundation.

Cavendish, J. C., Welch, M. R., & Leege, D. C. (1998). Social network theory and predictors of religiosity for Black and White Catholics: Evidence of a "Black sacred cosmos?" *Journal for the Scientific Study of Religion, 37* (3), 397 – 410.

Centers for Disease Control (CDC) (2008). *American Women Waiting to Begin Families: Average Age at First Birth up More Than Three Years From 1970 to 2000*. Retrieved February 4, 2008, from http://www.cdc.gov/OD/OC/MEDIA/pressrel/r021211.htm.

Census Bureau (2008). *Marriage and Divorce*. Retrieved February 4, 2008, from http://www.census.gov/population/www/socdemo/marr-div.html.

Census Bureau (2009). *Population Finder*. Retrieved December 14, 2009, from factfinder.census.gov/servlet/SAFFPopulation?_submenuId=population_0&_sse=on.

Center for Spirituality, Theology & Health (2017). Religious Involvement and Mortality in the Black Women's Health Study. *Crossroads: Exploring Research on Religion, Spirituality and Health, 6* (12), p. 2.

Chatters, L. M., Levin, J. S., & Taylor, R. J. (1992). Antecedents and dimensions of religious involvement among older Black adults. *Journal of Gerontology, 47* (6), S269 – S278.

Chatters, L. M., & Taylor, R. J. (1994). Religious involvement among older African Americans. In J. S. Levin (Ed.), *Religion in aging and health: Theoretical foundations and methodological frontiers* (pp. 196 – 230). Thousand Oaks, CA: Sage Publications.

*Chaves, M., & Cavendish, J. C. (1994). More evidence on U.S. Catholic church attendance. *Journal for the Scientific Study of Religion, 33* (4), 376 – 381.

Chaves, M., Konieczny, M. E., Beyerlein, K., & Barman, E. (1999). The national congregations study: Background, methods and selected results. *Journal for the Scientific Study of Religion, 38* (4), 458 – 476.

Child Trends Databank (2007). *Religious Service Attendance*. Retrieved August 13, 2007, from www.childtrendsdatabank.org/indicators/32ReligiousServices.cfm.

Choi, N., Fuqua, D. R., & Newman, J. L. (2009). Exploratory and confirmatory studies of the structure of the Bem Sex Role Inventory Short Form with two divergent samples. *Educational and Psychological Measurement, 69* (4), 696 – 705.

*CNN/USA Today (2000). *Gallup/CNN/USA Today Poll: 2000 Elections/Yugoslavia/Violence/Star Wars/Abortion, Apr, 1999* [Data file and code book]. USAIPOCNUS1999-9904025, Version 1. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155 – 159.

Comstock, G. W., & Partridge, K. B. (1972). Church attendance and health. *Journal of Chronic Disease, 25*, 665 – 672.

Cooper, H. (2009). Hypotheses and problems in research synthesis. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd Ed., pp. 19 – 35). New York, NY: Russell Sage Foundation.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.) (2009). *The Handbook of Research Synthesis and Meta-Analysis* (2nd Ed.). New York, NY: Russell Sage Foundation.

Cornwall, M. (1989). The determinants of religious behavior: A theoretical model and empirical test. *Social Forces, 68*, 572 – 592.

Curtis, K. T., & Ellison, C. G. (2002). Religious heterogamy and marital conflict: Findings from the National Survey of Households. *Journal of Family Issues, 23* (4), 551 – 576.

Davis, J. A., & Smith, T. W. (1996). *General Social Surveys, 1972 – 1996: Cumulative Codebook*. Storrs, CT: Roper Center.

Davis, J. A., & Smith, T. W. (2004). *General Social Survey, 2004*. Storrs, CT: National Opinion Research Center.

Davis, J. A., Smith, T. W., & Marsden, P. V. (2007). *General Social Surveys, 1972 – 2006* [Cumulative File] [Data File]. ICPSR 04295-v1. Chicago, IL: National Opinion Research Center.

Davis, J. H. (1962). *The Outsider and the Urban Church*. Urban Pamphlet No. 19 Philadelphia, PA: Board of Missions of the Methodist Church.

de Vaus, D. A. (1984). Workforce participation and sex differences in church attendance. *Review of Religious Research, 25* (3), 247 – 256.

Demerath, N. J. (1965). *Social Class in American Protestantism*. Chicago, IL: Rand McNally.

Diener, E., & Clifton, D. (2002). Life satisfaction and religiosity in broad probability samples. *Psychological Inquiry, 13* (3), 206 – 209.

DiIulio, J. J. (2002). The three faith factors. *The Public Interest, 149*, 50 – 64.

Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method*. New York, NY: Wiley.

Dunn, K. S., & Horgas, A. L. (2000). The prevalence of prayer as a spiritual self-care modality in elders. *Journal of Holistic Nursing, 18* (4), 337 – 351.

Durkheim, E. (1897/1951). *Suicide*. New York: Free Press.

Durkheim, E. (1915/1961). *The Elementary Forms of the Religious Life*. New York: Collier Books.

Eagly, A. H. (2009). The his and hers of prosocial behavior: An examination of the social psychology of gender. *American Psychologist, 64* (8), 644 – 658.

Earl, J. R., & Perricone, P. J. (1986). Premarital sexuality: A ten-year study of attitudes and behavior on a small university campus. *The Journal of Sex Research, 22* (3), 304 – 310.

Edgell Becker, P., & Hofmeister, H. (2001). Work, family and religious involvement for men and women. *Journal for the Scientific Study of Religion, 40* (4), 707 - 722.

Eliassen, A. H., Taylor, J., & Lloyd, D. A. (2005). Subjective religiosity and depression in the transition to adulthood. *Journal for the Scientific Study of Religion, 44* (2), 187 – 199.

Ell, K. O., & Haywood, L. J. (1985). Sociocultural factors in MI recovery: An exploratory study. *International Journal of Psychiatry in Medicine, 15*, 157 – 175.

Ellison, C. W. (1983). Spiritual well-being: Conceptualization and measurement. *Journal of Psychology and Theology, 11* (4), 330 – 340.

Ellison, C. G., Boardman, J. D., Williams, D. R., & Jackson, J. S. (2001). Religious involvement, stress and mental health: Findings from the 1995 Detroit area study. *Social Forces, 80* (1), 215 – 249.

Evans, T. D., Cullen, F. T., Dunaway, R. G., & Burton, V. S. Jr. (1995). Religion and crime reexamined: The impact of religion, secular controls and social ecology on adult criminality. *Criminology, 33* (2), 195 – 225.

Evans, W., Oates, W., & Schwab, R. (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy, 100* (5), 966 – 991.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE, 4* (5): e5738, 1 – 10. doi: 10.1371/journal.pone.0005738

Fergusson, D. M., Horwood, L. J., & Shannon, F. T. (1984). A proportional hazards model of family breakdown. *Journal of Marriage and the Family, 46* (3), 539 – 549.

Ferraro, K. F., & Kelley-Moore, J. A. (2001). Religious seeking among affiliates and non-affiliates: Do mental and physical health problems spur religious coping? *Review of Religious Research, 42* (3), 229 – 251.

Fetzer Institute / National Institute on Aging Working Group (1999). *Multidimensional measurement of religiousness/spirituality for use in health research: A report of the Fetzer Institute/National Institute on Aging Working Group.* Kalamazoo, MI: Fetzer Institute.

Finke, R. (2003). *Spiritual capital: Definitions, applications and new frontiers.* Paper presented at the Spiritual Capital Planning Meeting, Philadelphia, PA.

*Fisher, K., & Gershuny, J. (2015). *American Heritage Time Use Study, release 7*. Oxford, UK: Centre for Time Use Research. Retrieved from: http://www.timeuse.org/ahtus

Fisler, J., Agati, H. A., Chance, S. M., Donahue, A. E., Donahue, G. A., Eickhoff, E. J. et al. (2009). Keeping (or losing) the faith: Reflections on spiritual struggles and their resolution by college seniors. *College Student Affairs Journal, 27* (2), 257 – 274.

Fowler, J. W. (1981). *Stages of Faith*. San Francisco, CA: Harper & Row.

Francis, L. J., & Wilcox, C. (1996). Religion and gender orientation. *Personality and Individual Differences, 20* (1), 119 – 121.

Francis, L. J., & Wilcox, C. (1998). Religiosity and femininity: Do women really hold a more positive attitude toward Christianity? *Journal for the Scientific Study of Religion, 37* (3), 462 – 469.

Frederickson, B. L. (2002). How does religion benefit health and well-being? Are positive emotions active ingredients? *Psychological Inquiry, 13* (3), 209 – 213.

Freud, S. (1907/1962). *Obsessive Actions and Religious Practices*. In J. Strachey (Ed.), *The standard edition of the complete psychological works of Sigmund Freud*. London, England: The Hogarth Press.

Freud, S. (1927/1961). *The Future of an Illusion*. In J. Strachey (Ed.), *The future of an illusion: translated and edited by James Strachey*. New York: W. W. Norton & Company.

Gallup, G. H. (1990). *Religion in America: 1990*. Princeton, NJ: The Gallup Organization.

Gallup, G. H., & Castelli, J. (1989). *The People's Religion: American Faith in the 90's*. New York: Macmillan Publishing.

Gallup, G. H., & Jones, S. (1989). *100 Questions and Answers: Religion in America*. Princeton, NJ: Princeton Religion Research Center.

Gallup, G., Jr., & Lindsay, M. (1999). *Surveying the Religious Landscape: Trends in U.S. Beliefs.* Harrisburg, PA: Morehouse Publishing.

Gallup, A. M., & Newport, F. (2008). *The Gallup Poll Cumulative Index: Public Opinion, 1998 – 2007*. New York, NY: Rowman & Littlefield Publishers.

Gallup Organization (1985). 50 years of Gallup surveys on religion. *The Gallup Report: Report No. 236*.

Gallup Organization (2007). *The Gallup Poll: Religion*. Retrieved March 1, 2007, from www.galluppoll.com/content/default.aspx?ci=1690.

*Gallup Organization (1950). *Gallup Poll #452, Jan, 1950* [Data file and code book]. USAIPO1950-0452, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May 27, 2017.

*Gallup Organization (1954a). *Gallup Poll # 1954-0530: Wages and Gender/Communism/Religion/Vaccines/McCarthy, May, 1954* [Data file and code book]. USAIPO1954-0530, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May 28, 2017.

*Gallup Organization (1954b). *Gallup Poll # 532, Jun, 1954* [Data file and code book]. USAIPO1954-0532, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-28-2017.

*Gallup Organization (1955a). *Gallup Poll # 1955-0543: Politics/China/Alcoholism, Feb, 1955* [Data file and code book]. USAIPO1955-0543, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-28-2017.

*Gallup Organization (1955b). *Gallup Poll # 1955-0547: Polio Vaccine/Religion/1956 Presidential Election, May, 1955* [Data file and code book]. USAIPO1955-0547, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-28-2017.

*Gallup Organization (1955c). *Gallup Poll # 1955-0549: Television/Diet and Health/Boxing/Safety Belts, Jun, 1955* [Data file and code book]. USAIPO1955-0549, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-28-2017.

*Gallup Organization (1955d). *Gallup Poll # 1955-0553: Politics/Defense/Education, Sep, 1955* [Data file and code book]. USAIPO1955-0553, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-28-2017.

*Gallup Organization (1956a). *Gallup Poll # 1956-0565: Presidential Election, May, 1956* [Data file and code book]. USAIPO1956-0565, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-28-2017.

*Gallup Organization (1956b). *Gallup Poll # 1956-0568: Politics, Aug, 1956* [Data file and code book]. USAIPO1956-0568, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1956c). *Gallup Poll # 1956-0573: Presidential Election/Political Parties, Oct, 1956* [Data file and code book]. USAIPO1956-0573, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1956d). *Gallup Poll # 1956-0575: Television/Politics, Nov, 1956* [Data file and code book]. USAIPO1956-0575, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1957). *Gallup Poll # 1957-0580, Mar, 1957* [Data file and code book]. USAIPO1957-0580, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1958). *Gallup Poll # 1958-0601: Politics, Jul, 1958* [Data file and code book]. USAIPO1958-0601, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1959). *Gallup Poll # 622, Dec, 1959* [Data file and code book]. USAIPO1959-0622, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1960). *Gallup Poll # 625, Mar, 1960* [Data file and code book]. USAIPO1960-0625, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1961). *Gallup Poll # 1961-0653: 1964 Presidential Election, Dec, 1961* [Data file and code book]. USAIPO1961-0653, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1962). *Gallup Poll # 1962-0658: Race Relations, May, 1962* [Data file and code book]. USAIPO1962-0658, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1963). *Gallup Poll # 1963-0678: Kennedy Administration/Russia/Race Relations, Oct, 1963* [Data file and code book]. USAIPO1963-0678, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1968a). *Gallup Poll # 1968-0765: 1968 Presidential Election, Jul, 1968* [Data file and code book]. USAIPO1968-0765, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1968b). *Gallup Poll # 1968-0766: 1968 Presidential Election/Birth Control, Aug, 1968* [Data file and code book]. USAIPO1968-0766, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1971). *Gallup Poll # 839, Oct, 1971* [Data file and code book]. USAIPO1971-0839, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1972). *Gallup Poll # 861, Dec, 1972* [Data file and code book]. USAIPO1972-0861, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1973). *Gallup Poll # 873, Jun, 1973* [Data file and code book]. USAIPO1973-0873, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1974). *Gallup Poll # 911, Jul, 1974* [Data file and code book]. USAIPO1974-0911, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1975). *Gallup Poll # 933, Aug, 1975* [Data file and code book]. USAIPO1975-0933, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1976). *Gallup Poll # 960, Oct, 1976* [Data file and code book]. USAIPO1976-0960, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1977). *Gallup Poll # 989, Nov, 1977* [Data file and code book]. USAIPO1977-0989, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1978). *Gallup Poll # 1106G, Jul, 1978* [Data file and code book]. USAIPO1978-1106G, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1979). *Gallup Poll # 1139G, Sep, 1979* [Data file and code book]. USAIPO1979-1139G, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1980). *Gallup Poll # 1153G, Apr, 1980* [Data file and code book]. USAIPO1980-1153G, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1981). *Gallup Poll # 1173G, May, 1981* [Data file and code book]. USAIPO1981-1173G, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1982). *Gallup Poll # 1196G, Jun, 1982* [Data file and code book]. USAIPO1982-1196G, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1983). *Gallup Poll # 1220G, Aug, 1983* [Data file and code book]. USAIPO1983-1220G, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1984). *Gallup Poll # 1242G, Sep, 1984* [Data file and code book]. USAIPO1984-1242G, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1985). *Gallup Poll # 1255G, Jul, 1985* [Data file and code book]. USAIPO1985-1255G, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1986). *Gallup Poll # 1269G, Oct, 1986* [Data file and code book]. USAIPO1986-1269G, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1987). *Gallup Poll # 1987-1272G: Reagan/1988 Presidential Election/Political Party/Finances/Federal Spending, Jan, 1987* [Data file and code book]. USAIPO1987-1272G, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1990). *Gallup News Service Survey # 1990-922012: Paranormal Phenomena, Jun, 1990* [Data file and code book]. USAIPOGNS1990-922012, Version 1. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1991). *Gallup News Service Poll: 50th Anniversary of Pearl Harbor, Nov, 1991* [Data file and code book]. USAIPOGNS1991-222025, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1992). *Gallup News Service Poll # 1992-222049: Presidential Election/Former Soviet Union/Population Growth/Religion, Apr, 1992* [Data file and code book]. USAIPOGNS1992-222049, Version 3. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1993). *Gallup News Service Poll # 1993-322061: The Supreme Court/Jurassic Park, Jun, 1993* [Data file and code book]. USAIPOGNS1993-322061, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1996). *Gallup News Service Poll: Gambling/Alcohol, Jun, 1996* [Data file and code book]. USAIPOGNS1996-9606015. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1997). *A CNN/USA Today/Gallup Poll: March Wave 1, Mar, 1997* [Data file and code book]. USAIPOCNUS1997-9703007. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (1998). *Gallup News Service Poll # 9812046: Economy/Religion, Dec, 1998* [Data file and code book]. USAIPOGNS1998-9812046, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2000). *Gallup News Service Poll: August Wave 1 Poll, Aug, 2000* [Data file and code book]. USAIPOGNS2000-36, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2001). *Gallup News Service Poll: Death Penalty/Persian Gulf War/Religion, Feb, 2001* [Data file and code book]. USAIPOGNS2001-06, Version 1. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2004a). *Gallup News Service Poll: 2004 Presidential Election/Prescription Drugs/JFK/Terrorism/Religion, Nov, 2003* [Data file and code book]. USAIPOGNS2003-50, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2004b). *Gallup News Service Poll: Economy/Employment/War in Iraq/Price of Gasoline/Religion, Jun, 2004* [Data file and code book]. USAIPOGNS2004-19, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2005a). *Gallup News Service Poll # 2005-19: Economy/Finances, Apr, 2005* [Data file and code book]. USAIPOGNS2005-19, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2005b). *Gallup News Service Poll # 2005-56: Economy/Finances/Religion, Nov, 2005* [Data file and code book]. USAIPOGNS2005-56, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2009). *Gallup News Service Poll: Social Series--Values and Beliefs, May, 2009* [Data file and code book]. USAIPOGNS2009-09, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2012a). *Gallup News Service Poll: November Wave 1--Personal Satisfaction/2012 Presidential Election/Religion, Nov, 2011* [Data file and code book]. USAIPOGNS2011-21, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2012b). *Gallup News Service Poll: June Wave 1--2012 Presidential Election/Religion, Jun, 2012* [Data file and code book]. USAIPOGNS2012-08, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*Gallup Organization (2015). *Gallup News Service Poll: Values and Beliefs, May, 2015* [Data file and code book]. USAIPOGNS2015-05, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

Gallup Poll and Princeton Religion Research Center (1978). *The Unchurched Americans*. Princeton, NJ: Gallup International.

Galton, F. (1872). Statistical inquiries into the efficacy of prayer. *Fortnightly Review, 68*, 125 – 135.

Garroutte, E. M., Beals, J., Anderson, H. O., Henderson, J. A., Nez-Henderson, P., Thomas, J., et al. (2014). Religio-spiritual participation in two American Indian populations. *Journal for the Scientific Study of Religion, 53*(1), 17 – 37.

Gartner, J., Larson, D. B., & Allen, G. D. (1991). Religious commitment and mental health: A review of the empirical literature. *Journal of Psychology and Theology, 19* (1), 6 – 25.

Glenn, N. D. (1964). Negro religion and Negro status in the United States. In L. Schneider (Ed.), *Religion, culture and society*. New York, NY: Wiley.

Glenn, N. D., & Supancic, M. (1984). The social and demographic correlates of divorce and separation in the United States: An update and reconsideration. *Journal of Marriage and the Family, 46*, 563 – 576.

Glenn, N. D., & Weaver, C. N. (1978). A multivariate, multisurvey study of marital happiness. *Journal of Marriage and the Family, 40*, 269 – 282.

Glock, C. Y. (2009). Early religious research association history reflecting on the fifties. *Review of Religious Research, 50* (3), 257 – 260.

Glock, C. Y., Ringer, b. R., & Babbie, E. E. (1967). *To Comfort and to Challenge*. Berkeley, CA: University of California Press.

Glock, C. Y., & Stark, R. (1965). *Religion and Society in Tension*. Chicago: Rand McNally.

Goodman, M. A., & Dollahite, D. C. (2006). How religious couples perceive the influence of God in their marriage. *Review of Religious Research, 48* (2), 141 – 155.

Gorsuch, R. L., & McFarland, S. G. (1972). Single vs. multiple item scales for measuring religious commitment. *Journal for the Scientific Study of Religion, 11*, 53 – 64.

Gray, A., Berlin, J. A., McKinlay, J. B., & Longcope, C. (1991). An examination of research design effects on the association of testosterone and male aging: Results of a meta-analysis. *Journal of Clinical Epidemiology, 44* (7), 671 – 684.

Greeley, A. M. (1963). A note on the origins of religious differences. *Journal for the Scientific Study of Religion, 3*, 21 – 31.

Greeley, A. M. (1972). *The Denominational Society*. Glenview, IL: Scott, Foresman.

Greeley, A. M. (1976). *Ethnicity, Denomination and Inequality*. Beverly Hills, CA: Sage Publications.

Greeley, A. M. (1979). The sociology of American Catholics. *Annual Review of Sociology, 5*, 91 – 111.

Green, J. C. (2007). *The Faith Factor: How Religion Influences American Elections*. Westport, CT: Praeger Publishers.

Gunnoe, M. L., & Moore, K. A. (2002). Predictors of religiosity among youth aged 17 – 22: A longitudinal study of the national survey of children. *Journal for the Scientific Study of Religion, 41* (4), 613 – 622.

Hadaway, C. K., & Marler, P. L. (1997). *Do Canadians over-report church membership and attendance?: A case study of religion in a Canadian County*. Paper presented at the annual meeting of the Association for the Sociology of Religion, Toronto, Canada.

*Hadaway, C. K., & Marler, P. L. (2005). How many Americans attend worship each week? An alternative approach to measurement. *Journal for the Scientific Study of Religion, 44* (3), 307 – 322.

*Hadaway, C. K., Marler, P. L., & Chaves, M. (1993). What the polls don't show: A closer look at U.S. church attendance. *American Sociological Review, 58*, 741 – 752.

Hadaway, C. K., Marler, P. L., & Chaves, M. (1998). Overreporting church attendance in America: Evidence that demands the same verdict. *American Sociological Review, 63* (1), 122 – 130.

Hadden, J. K. (1987). Toward desacralizing secularization theory. *Social Forces, 65* (3), 587 – 611.

Hartford Institute for Religion Research (2016). *Official Denominational Web Sites*. Retrieved November 22, 2016, from http://hirr.hartsem.edu/denom/homepages.html.

Hays, J. C., Landerman, L. R., Blazer, D. G., Koenig, H. G., Carroll, J. W., & Musick, M. A. (1998). Aging, health and the "electronic church." *Journal of Aging and Health, 10* (4), 458 – 482.

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin*, *92*, 490 – 499.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin, 88* (2), 359 – 369.

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. New York: Academic Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analysis. *British Medical Journal*, *327*, 557 – 560. doi 10.1136/bmj.327.7414.557

Hill, P. C., & Hood, R. W. Jr. (1999). *Measures of Religiosity*. Birmingham, AL: Religious Education Press.

Hill, P. C., & Pargament, K. I. (2003). Advances in the conceptualization and measurement of religion and spirituality: Implications for physical and mental health research. *American Psychologist*, *58* (1), 64 – 74.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied Statistics for the Behavioral Sciences.* Boston, MA: Houghton Mifflin Company.

Hirschi, T., & Stark, R. (1969). Hellfire and delinquency. *Social Problems*, *17* (2), 202 – 213.

Hoge, D. R., & Carroll, J. W. (1978). Determinants of commitment and participation in suburban Protestant churches. *Journal for the Scientific Study of Religion*, *17*, 107 – 127.

Hoge, D. R., Johnson, B., & Luidens, D. A. (1993). Determinants of church involvement of young adults who grew up in Presbyterian Churches. *Journal for the Scientific Study of Religion*, *32* (3), 242 – 255.

Hoge, D. R., & Petrillo, G. H. (1978a). Development of religious thinking in adolescence: A test of Goldman's theories. *Journal for the Scientific Study of Religion, 17* (2), 139 – 154.

Hoge, D. R., & Petrillo, G. H. (1978b). Determinants of church participation and attitudes among high school youth. *Journal for the Scientific Study of Religion, 17* (4), 359 – 379.

Hoge, D. R., & Roozen, D. A. (1979). *Understanding Church Growth and Decline: 1950 – 1978*. NY: The Pilgrim Press.

Hood, R. W. Jr. (1975). The construction and preliminary validation of a measure of reported mystical experience. *Journal for the Scientific Study of Religion, 14*, 29 – 41.

*House, J. S. (2014). *Americans' Changing Lives: Waves I, II, III, IV, and V, 1986, 1989, 1994, 2002, and 2011* [Data file and code book]. ICPSR04690-V7. Ann Arbor, MI: Bibliographic Citation: Inter-university Consortium for Political and Social Research [Distributor], 2014-09-09. http://doi.org/10.3886/ICPSR04690.v7

Hout, M., & Fischer, C. S. (2002). Why more Americans have no religious preference: Politics and generations. *American Sociological Review, 67*, 165 – 190.

Hout, M., & Greeley, A. (1998). What church officials' reports don't show: Another look at church attendance data. *American Sociological Review, 63* (1), 113 – 119.

Hummer, R. A., Rogers, R. G., Nam, C. B., & Ellison, C. G. (1999). Religious involvement and U.S. adult mortality. *Demography, 36* (2), 273 – 285.

Hunsberger, B. (2000). Swimming against the current: Exceptional cases of apostates and converts. In L. J. Francis & Y. J. Katz (Eds.), *Joining and Leaving Religion: Research Perspectives*. Trowbridge, England: Redwood Books.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings* (1st Ed.). Newbury Park, CA: Sage Publications.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-Analysis: Cumulating Research Findings Across Studies*. Beverly Hills, CA: Sage Publications.

Iannaccone, L. R. (1984). *Consumption capital and habit formation with an application to religious participation.* Unpublished doctoral dissertation, The University of Chicago, Chicago, IL.

Iannaccone, L. R. (1988). A formal model of church and sect. *American Journal of Sociology, 94,* 241 – 268.

Iannaccone, L. R. (1990). Religious practice: A human capital approach. *Journal for the Scientific Study of Religion, 29* (3), 297 – 314.

Iannaccone, L. R. (1994). Why strict churches are strong. *American Journal of Sociology, 99* (5), 1180 – 1211.

Iannaccone, L. R. (1995a). Risk, rationality and religious portfolios. *Economic Inquiry, 33*, 285 – 295.

Iannaccone, L. R. (1995b). Household production, human capital and the economics of religion. In M. Tommasi & K. Ierulli (Eds.), *The New Economics of Human Behavior* (pp. 172 – 187). Cambridge, MA: Cambridge University Press.

Iannaccone, L. R. (1997). Toward an economic theory of "fundamentalism." *Journal of Institutional and Theoretical Economics, 153*, 100 – 116.

Iannaccone, L. R. (1998). Introduction to the economics of religion. *Journal of Economic Literature, 36*, 1465 – 1496.

Iannaccone, L. R. (July, 2003). *Looking backward: A cross-national study of religious trends*. Retrieved March 7, 2011, from http://religionomics.com/iannaccone/papers/Iannaccone%20-%20Looking%20Backward.pdf.

Iannaccone, L. R., & Everton, S. F. (2002). Never on sunny days: Lessons from weekly attendance counts. *Journal for the Scientific Study of Religion*, *43* (2), 191 – 207.

Idler, E. L., Boulifard, D. A., Labouvie, E., Chen, Y. Y., Krause, T. J., & Contrada, R. J. (2009). Looking inside the black box of “attendance at services:” New measures for exploring an old dimension in religion and health research. *The International Journal for the Psychology of Religion, 19*, 1 – 20.

Idler, E. L., Kasl, S. V., & Hays, J. C. (2001). Patterns of religious practice in belief in the last year of life. *Journals of Gerontology, 56B* (6), 326 – 334.

Ilies, R., Hauserman, N., Schwochau, S., & Stibal, J. (2003). Reported incidence rates of work-related sexual harassment in the United States: Using meta-analysis to explain reported rate disparities. *Personnel Psychology, 56*, 607 – 631.

*Jackson, J. S., & Gurin, G. (1999). *National Survey of Black Americans, 1979-1980* [Data file and code book]. ICPSR08512-v1. Ann Arbor, MI: Inter-university Bibliographic Citation: Consortium for Political and Social Research [distributor], 1999. http://doi.org/10.3886/ICPSR08512.v1

Jackson, R. S., & Reddick, B. (1999). The African American Church and university partnerships: Establishing lasting collaborations. *Health Education Behavior, 26*, 663 – 674.

Janda, K. (1970). Data quality control and library research on political parties. In R. Narall & R. Cohen (Eds.), *A Handbook of Method in Cultural Anthropology*. New York: Doubleday.

James, W. (1902). *The Varieties of Religious Experience*. New York, NY: Penguin Books.

Jasso, G., Massey, D. S., Rosenzweig, M. R., & Smith, J. P. (2002). *Exploring the religious preference of recent immigrants to the United States: Evidence from the New Immigrant Survey Pilot.* Retrieved August 17, 2009, from nis.princeton.edu/downloads/papers/jmrsrel.pdf.

Johnson, B. R., Larson, D. B., Li, S., & Jang, S. J. (2000). Escaping from the crime of inner cities: Church attendance and religious salience among disadvantaged youth. *Justice Quarterly, 17* (2), 377 – 391.

*Johnston, L. D., Bachman, J. G., & O'Malley, P. M. (2006). *Monitoring the future: A continuing study of the lifestyles and values of youth, 1993* [Data file and code book]. ICPSR06367-v3. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor], 2006-08-21. http://doi.org/10.3886/ICPSR06367.v3

*Johnston, L. D., Bachman, J. G., O'Malley, P. M., & Schulenberg, J. (1999a). *Monitoring the future: A continuing study of American youth (8th and 10th Grade Surveys), 1993* [Data file and code book]. ICPSR02523-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor], 1999. http://doi.org/10.3886/ICPSR02523.v1

*Johnston, L. D., Bachman, J. G., O'Malley, P. M., & Schulenberg, J. (1999b). *Monitoring the future: A continuing study of American youth (8th and 10th Grade Surveys), 1994* [Data file and code book]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and Distributor], 1999.

*Johnston, L. D., Bachman, J. G., O'Malley, P. M., & Schulenberg, J. (1999c). *Monitoring the future: A continuing study of American youth (8th and 10th Grade Surveys), 1995* [Data file and code book]. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [producer and Distributor], 1999.

*Johnston, L. D., Bachman, J. G., O'Malley, P. M., & Schulenberg, J. (2006). *Monitoring the future: A continuing study of American youth (8th and 10th Grade Surveys), 1996* [Data file and code book]. ICPSR02350-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor], 2006-03-06. https://doi.org/10.3886/ICPSR02350.v2

*Johnston, L. D., Bachman, J. G., O'Malley, P. M., & Schulenberg, J. (2008). *Monitoring the future: A continuing study of American youth (8th and 10th Grade Surveys), 1992* [Data file and code book]. ICPSR02522-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [Distributor], 2008-07-16. http://doi.org/10.3886/ICPSR02522.v2

Jones, A. P., Lee, A. J., Butler, M. C., & Main, D. S. (1983). Apples and oranges: An empirical comparison of commonly used indices of interrater agreement. *Academy of Management Journal*, *26* (3), 507 – 519.

Judson University (December 8, 2016). Benjamin P. Browne Library. Retrieved from http://www.judsonu.edu/Library/

Kelley, D. M. (1972). *Why Conservative Churches are Growing*. New York: Harper & Row.

King, V., Elder, G. H., & Whitbeck, L. B. (1997). Religious involvement among rural youth: An ecological and life course perspective. *Journal of Research on Adolescence, 7* (4), 431 – 456.

King, M., Speck, P., & Thomas, A. (2001). The royal free interview for spiritual and religious beliefs: development and validation of a self-report version. *Psychological Medicine, 31* (6), 1015 – 1023.

Koenig, H. G., Hays, J. C., Larson, D. B., George, L. K., Cohen, H. J., McCullough, M. et al. (1999). Does religious attendance prolong survival? A six-year follow-up study of 3,968 older adults. *Journal of Gerontology, 54A,* M370 – M377.

Koenig, H. G., & Larson, D. B. (1998). Use of hospital services, religious attendance and religious affiliation. *Southern Medical Journal, 91* (10), 925 – 932.

Koenig, H. G., King, D. E., & Carson, V. B. (2012). *Handbook of Religion and Health* (2nd Ed.). New York: Oxford University Press.

Koenig, H. G., McCullough, M. E., & Larson, D. B. (2001). *Handbook of Religion and Health*. New York: Oxford University Press.

Ladd, K. L., & Spilka, B. (2006). Inward, outward and upward: Scale reliability and validation. *Journal for the Scientific Study of Religion, 45* (2), 233 – 251.

Larson, D. B. (1985). Religious involvement: Its association with marital status, marital well-being and morality. In G. Rekers (Ed.), *Family building* (pp. 121 – 147). Ventura, CA: Regal Books.

Larson, D. B., Pattison, E. M., Blazer, D. G., Omran, A. R., & Kaplan, B. H. (1986). Systematic analysis of research on religious variables in four major psychiatric journals, 1978 – 1982. *American Journal of Psychiatry, 143*, 329 – 334.

Larson, D. B., Swyers, J. P., & McCullough, M. E. (1998). *Scientific Research on Spirituality and Health: A Consensus Report*. National Institute for Healthcare Research.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994a). Nonlinearity of range corrections in meta-analysis: A test of an improved procedure. *Journal of Applied Psychology, 79*, 425 – 438.

Law, K. S., Schmidt, F. L., & Hunter, J. E. (1994b). A test of two refinements in meta-analysis procedures. *Journal of Applied Psychology, 79*, 978 – 986.

Lazar, A. (2006). Fear of personal death as a predictor of motivation for religious behavior. *Review of Religious Research, 48* (2), 179 – 189.

Lazerwitz, B. (1961). Some factors associated with variations in church attendance. *Social Forces, 39* (4), 301 – 309.

Lazerwitz, B. (1970). Contrasting the effects of generation, class, sex and age on group identification in the Jewish and Protestant communities. *Social Forces, 49,* 50 – 59.

Lenski, G. (1963). *The Religious Factor*. Garden City, NY: Doubleday.

Levin, J. S. (1996). How religion influences morbidity and health: Reflections on natural history, salutogenesis and host resistance. *Social Science & Medicine, 43* (5), 849 – 864.

Levin, J. S., Taylor, R. J., & Chatters, L. M. (1994). Race and gender differences in religiosity among older adults: Findings from four national surveys. *Journal of Gerontology: Social Sciences, 49,* S137 – S145.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis* (Applied Social Research Method Series, Vol. 49). Thousand Oaks, CA: Sage Publications.

Loury, L. D. (2004). Does church attendance really increase schooling? *Journal for the Scientific Study of Religion, 43* (1), 119 – 127.

Luckmann, T. (1967). *The Invisible Religion*. New York, NY: Macmillan.

Lummis, A. T. (2004). A research note: Real men and church participation. Review of *Religious Research, 45* (4), 404 – 414.

Lundahl, B. W., Taylor, M. J., Stevenson, R., & Roberts, K. D. (2008). Process-based forgiveness interventions: A meta-analytic review. *Research on Social Work Practice, 18,* 465 – 478.

McCarthy, J. (1979). Religious commitment, affiliation and marriage dissolution. In R. Wuthnow (Ed.), *The religious dimension: New directions in quantitative research* (pp. 179 – 197). New York, NY: Academic Press.

McConahay, J. B. (1970). Attitudes of Negroes toward the Church following the Los Angeles riot. *Sociological Analysis, 31*, 12 – 22.

McCullough, M. E., Pargament, K. I., & Thoresen, C. E. (2000). *Forgiveness: Theory, Research and Practice*. NY: The Guilford Press.

McGuire, M. B. (1981). The individual's religion. In M. B. McGuire (Ed.), *Religion: The social context*. Belmont, CA: Wadsworth Publishing Company.

Maccoby, E. E., & Maccoby, N. (1954). The interview: A tool of social science. In L. Gardiner (Ed.), *Handbook of social psychology*. Cambridge, MA: Addison – Wesley.

Mahoney, A., Pargament, K. I., Jewell, T., Swank, A. B., Scott, E., Emery, E. et al. (1999). Marriage and the spiritual realm: The role of proximal and distal religious constructs in marital functioning. *Journal of Family Psychology, 13* (3), 321 – 338.

Mahoney, A., Pargament, K. I., Tarakeshwar, N., & Swank, A. B. (2001). Religion in the home in the 1980s and 1990s: A meta-analytic review and conceptual analysis of links between religion, marriage and parenting. *Journal of Family Psychology, 15* (4), 559 – 596.

Marks, L. (2005). Religion and bio-psycho-social health: A review and conceptual model. *Journal of Religion and Health, 44* (2), 173 – 186.

Marler, P. L., & Hadaway, C. K. (1999). Testing the attendance gap in a conservative church. *Sociology of Religion, 60* (2), 175 - 186.

Martin, D. (1967). *A Sociology of English Religion*. London, England: SCM Press.

Martins, S. S., Tavares, H., Lobo, D. S., Galetti, A. M., & Gentil, V. (2004). Pathological gambling, gender and risk-taking behaviors. *Addictive Behaviors, 29*, 1231 – 1235.

Marty, M. E., Rosenberg, S. E., & Greeley, A. M. (1968). *What Do We Believe?* New York, NY: Meredith.

McKimmie, T., & Szurmak, J. (2002). Beyond grey literature: How grey questions can drive research. *Journal of Agricultural and Food Information, 4* (2), 71 – 79.

Meisenhelder, J. B., & Chandler, E. N. (2000). Faith, prayer and health outcomes in elderly Native Americans. *Clinical Nursing Research, 9* (2), 191 – 203.

Merriam-Webster (2011). *Merriam-Webster's On-Line Dictionary*. Retrieved January 16, 2011, from http://www.merriam-webster.com/dictionary/sacrament.

Metz, D. L. (1965). *The Invisible Member: A Report on Inactive Methodists*. Berkeley, CA: California Bureau of Community Research.

Mickley, J. R., & Soeken, K. L. (1993). Religiousness and hope in Hispanic- and Anglo-American women with breast cancer. *Oncology Nursing Forum, 20*, 1171 – 1177.

Mickley, J. R., Carson, V., & Soeken, K. L. (1995). Religion and adult mental health: State of the science in nursing. *Issues in Mental Health Nursing, 16*, 345 – 360.

Miller, A. S., & Hoffmann, J. P. (1995). Risk and religion: An explanation of gender differences in religiosity. *Journal for the Scientific Study of Religion, 34* (1), 63 – 75.

Miller, A. S., & Stark, R. (2002). Gender and religiousness: Can socialization explanations be saved? *American Journal of Sociology, 107* (6), 1399 – 1423.

Mindel, C., & Vaughan, C. E. (1978). A multidimensional approach to religiosity and disengagement. *Journal of Gerontology, 33*, 103 – 108.

Mitchell, J., & Weatherly, D. (2000). Beyond church attendance: Religiosity and mental health among rural older adults. *Journal of Cross-Cultural Gerontology, 15,* 37 – 54.

Moberg, D. O. (1961). Potential uses of the church-sect typology in comparative religious research. *International Journal of Comparative Sociology, 2,* 47 – 58.

Mockabee, S. T., Monson, J. Q., & Grant, J. T. (2001). Measuring religious commitment among Catholics and Protestants: A new approach. Journal *for the Scientific Study of Religion, 40* (4), 675 – 690.

Mueller, C. W., & Johnson, W. T. (1975). Socioeconomic status and religious participation. *American Sociological Review, 40*, 785 – 800.

Muller, C., & Ellison, C. G. (2001). Religious involvement, social capital and academic achievement: Evidence from the National Education Longitudinal Study of 1988. *Sociological Focus, 34* (2), 155 – 183.

Musick, M. A. (1996). Religion and subjective health among Black and White elders. *Journal of Health and Social Behavior, 37*, 221 – 237.

Musick, M. A., Blazer, D. G., & Hays, J. C. (2000). Religious activity, alcohol use and depression in a sample of elderly Baptists. *Research on Aging, 22* (2), 91 – 116.

Musick, M. A., House, J. S., & Williams, D. R. (2004). Attendance at religious services and mortality in a national sample. *Journal of Health and Social Behavior, 45*, 198 – 213.

Musick, M. A., Traphagan, J. W., Koenig, H. G., & Larson, D. B. (2000). Spirituality in physical health and aging. *Journal of Adult Development, 7* (2), 73 – 86.

Nash, D., & Berger, P. L. (1962). The child, the family and the religious revival in suburbia. *Journal for the Scientific Study of Religion, 2* (1), 85 – 93.

Newport, F. (2005). *Church attendance and party identification: Frequency of church attendance strongly related to partisanship. A Gallup news services report*. Retrieved September 8, 2009, from www.gallup.com/poll/16381/Church-Attendance-Party-Identification.aspx.

Newport, F. (2006). *Estimating Americans' worship behavior, part II.* Retrieved from www.gallup.com/poll/20824/Estimating-Americans-Worship-Behavior-Part.aspx.

Newport, F. (2007). *Highly religious republicans less likely to support Giuliani. A Gallup news service report*. Retrieved March 27, 2007, from www.galluppoll.com/content/?ci=26884.

Nisbet, P. A., Duberstein, P. R., Yeates, C., & Seidlitz, L. (2000). The effect of participation in religious activities on suicide versus natural death in adults 50 and older. *Journal of Nervous and Mental Disease, 188*, 543 – 546.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd Ed.). New York: McGraw-Hill.

O'Connor, T. P., Hoge, D. R., & Alexander, E. (2002). The relative influence of youth and adult experiences on personal spirituality and church involvement. *Journal for the Scientific Study of Religion, 41* (4), 723 – 732.

Oman, D., & Reed, D. (1998). Religion and mortality among the community-dwelling elderly. *American Journal of Public Health, 88* (10), 1469 – 1475.

Orwin, R., & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychological Bulletin, 97* (1), 134 – 147.

Pargament, K. I. (1997). *The Psychology of Religion and Coping: Theory, Research and Practice*. NY: Guilford Press.

Pargament, K. I., Silverman, W., Johnson, S., Echemendia, R., & Snyder, S. (1983). The psychosocial climate of religious congregations. *American Journal of Community Psychology, 11* (4), 351 – 381.

*Pearce, L. D., Hardie, J. H., & Foster, E. M. (2013). A person-centered examination of adolescent religiosity using latent class analysis. *Journal for the Scientific Study of Religion, 52* (1), 57 – 79.

Petts, R. J. (2009). Trajectories of religious participation from adolescence to young adulthood. *Journal for the Scientific Study of Religion, 48* (3), 552 – 571.

*Pew Forum on Religion & Public Life (2007). *Religious Landscape Survey - Continental Dataset, 2007* [Codebook]. Waco, TX: Baylor Institute for Studies of Religion [Producer]. Retrieved from www.thearda.com

Pew Forum on Religion & Public Life (2012). *The Pew Forum on Religion and Public Life: U.S. Religious Landscape Survey*. Retrieved June 12, 2012, from religions.pewforum.org/maps.

Pew Research Center (2016). *The Gender Gap in Religion Around the World: Women are Generally More Religious than Men Particularly Among Christians.* Retrieved June 16, 2016, from http://www.pewforum.org/2016/03/22/the-gender-gap-in-religion-around-the-world.

Piedmont, R. L., Kennedy, M. C., Sherman, M. F., Sherman, N. C., & Williams, J. E. G. (2008). A psychometric evaluation of the assessment of spirituality and religious sentiments (ASPIRES) scale: Short form. *Research in the Social Scientific Study of Religion, 19,* 163 – 181.

Ploch, D. R., & Hastings, D. W. (1998). Effects of parental church attendance, current family status and religious salience on church attendance. *Review of Religious Research, 39* (4), 309 – 320.

Poloma, M. M., & Pendleton, B. F. (1989). Exploring types of prayer and quality of life: A research note. *Review of Religious Research, 31* (1), 46 – 53.

Poppen, P. J. (1995). Gender and patterns of sexual risk taking in college students. *Sex Roles, 32* (7/8), 545 – 555.

Powell, L. H., Shahabi, L., & Thoresen, C. E. (2003). Religion and spirituality: Linkages to physical health. *American Psychologist*, *58* (1), 36 – 52.

Powell, M., & Ansic, D. (1997). Gender differences in risk behavior in financial decision-making: An experimental analysis. *Journal of Economic Psychology, 18*, 605 – 628.

*Presser, S., & Chaves, M. (2007). Is religious service attendance declining? *Journal for the Scientific Study of Religion, 46* (3), 417 – 423.

*Presser, S., & Stinson, L. (1998). Data collection mode and social desirability bias in self-reported religious attendance. *American Sociological Review, 63*, 137 – 145.

Protestant Episcopal Church of the United States (1992). *Church Annual*. New York: General Convention Office.

Raju, N. S., & Burke, M. J. (1983). Two new procedures for studying validity and generalization. *Journal of Applied Psychology, 68*, 382 – 395.

Regnerus, M. D. (2003). Linked lives, faith and behavior: Intergenerational religious influence on adolescent delinquency. *Journal for the Scientific Study of Religion, 42* (2), 189 – 203.

Regnerus, M. D., Elder, G. H. (2003). Staying on track in school: Religious influences in high- and low-risk settings. *Journal for the Scientific Study of Religion*, *42* (4), 633 – 649.

Regnerus, M. D., & Uecker, J. E. (2006). Finding faith, losing faith: The prevalence and context of religious transformations during adolescence. *Review of Religious Research*, *47* (3), 217 – 237.

Religious Congregations and Membership Study (2016). *U.S. Religion Census: 1952 to 2010.* Retrieved November 22, 2016, from www.rcms2010.org/index.php.

Ribar, D. (1994). Teenage fertility and high school completion. *Review of Economics and Statistics*, *76*, 413 – 424.

Richard, A. J., Bell, D. C., Carlson, J. W. (2000). Individual religiosity, moral community and drug user treatment. *Journal for the Scientific Study of Religion, 39* (2), 240 - 246.

Roof, W. C., & Roof, J. L. (1984). Review of the polls: Images of God among Americans. *Journal for the Scientific Study of Religion*, *23*, 201 – 205.

Roozen, D. A. (1979). The efficacy of demographic theories of religious change: Protestant church attendance, 1952 – 1968. In D. R. Hoge & D. A. Roozen (Eds.), *Understanding church growth and decline: 1950 – 1978*. New York, NY: Pilgrim.

Rosenthal, R. (1991). *Meta-Analytic Procedures for Social Research*. Newbury Park, CA: Sage Publications.

Rosenthal, R., & Rubin, D. B. (1982a). Comparing effect sizes of independent studies. *Psychological Bulletin*, *92*, 500 – 504.

Rosenthal, R., & Rubin, D. B. (1982b). Further meta-analytic procedures for assessing cognitive gender differences. *Journal of Educational Psychology, 74*, 708 – 712.

Rossi, M., & Scappini, E. (2014). Church attendance, problems of measurement, and interpreting indicators: A study of religious practice in the United States, 1975 – 2010. *Journal for the Scientific Study of Religion, 53* (2), 249 – 267.

Rothstein, H. R., & Hopewell, S. (2009). Grey literature. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The Handbook of Research Synthesis and Meta-Analysis* (2nd Ed., pp. 103 – 125). New York, NY: Russell Sage Foundation.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Chichester, UK: John Wiley & Sons.

Rubin, D. B. (1990). A new perspective on meta-analysis. In K. W. Wachter & M. L. Straf (Eds.), *The Future of Meta-Analysis*. New York: Russell Sage.

Sasaki, M. S. (1979). Status inconsistency and religious commitment. In R. Wuthnow (Ed.), *The religion dimension: New directions in quantitative research* (pp. 135 – 156). New York, NY: Academic Press.

Scanzoni, J., & Arnett, C. (1987). Enlarging the understanding of marital commitment via religious devoutness, gender role preferences and locus of marital control. *Journal of Family Issues, 8* (1), 136 – 156.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529 – 540.

Schnoll, R. A., Harlow, L. L., & Brower, L. (2000). Spirituality, demographic and disease factors and adjustment to cancer. *Cancer Practice, 8* (6), 298 – 304.

Schulze, R. (2004). *Meta-Analysis: A Comparison of Approaches*. Cambridge, MA: Hogrefe & Huber Publishers.

Schwarz, N. (1999). Self-Reports: How the questions shape the answers. *American Psychologist, 54* (2), 93 – 105.

Seidman, S. N. (2003). The aging male: Androgens, erectile dysfunction and depression. *Journal of Clinical Psychiatry*, *64*, 31 – 37.

Seldom. (2000). In *American Heritage College Dictionary (3rd Ed.)*. Boston, MA: Houghton Mifflin Company.

Shadish, W. R. (2011). *Meta-analysis*. Unpublished presentation, School of Social Sciences, Humanities and Arts, University of California, Merced.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston, MA: Houghton-Mifflin.

Shafranske, E. P., & Malony, H. N. (1985). *Psychologists' religious and spiritual orientations and their practice of psychotherapy*. Paper presented at the annual meeting of the American Psychological Association, Los Angeles, CA.

Shrum, W. (1980). Religion and marital instability: Change in the 1970's? *Review of Religious Research, 21*, 135 – 147.

Simon, D., Nahoul, K., & Charles, M. A. (1996). Sex hormones, aging, ethnicity and insulin sensitivity in men: An overview of the TELECOM study. In A. Vermeulen & B. J. Oddens (Eds.), *Androgens and the Aging Male* (pp. 85 – 102). New York, NY: Parthenon Publishing.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *The Educational Researcher*, *15,* 5 – 11.

Smietana, B. (2006). Statistical illusion. *Christianity Today, April*, 2006. Retrieved December 8, 2009, from http://www.christianitytoday.com/ct/2006/april/32.85.html.

Smith, T. W. (1990). Classifying protestant denominations. *Review of Religious Research, 31* (3), 225 – 245.

Smith, T. W. (1998). A review of church attendance measures. *American Sociological Review, 63*, 131 – 136.

Smith, T. W. (2006). The national spiritual transformation study. *Journal for the Scientific Study of Religion, 45* (2), 283 – 296.

*Smith, T. W., Marsden, P., Hout, M., & Kim, J. (2016). *General Social Surveys, 1972 – 2014*. [Data file and code book]. Chicago, IL: NORC at the University of Chicago. Retrieved from http://gss.norc.org/Get-The-Data

Smith, C., Lundquist Denton, M., Faris, R., & Regnerus, M. (2002). Mapping American adolescent religious participation. *Journal for the Scientific Study of Religion, 41* (4), 597 – 612.

Smith, T. B., McCullough, M. E., & Poll, J. (2003). Religiousness and depression: Evidence for a main effect and the moderating influence of stressful life events. *Psychological Bulletin, 129* (4), 614 – 636.

Southern Illinois University at Carbondale (December 8, 2016). Morris Library: Find Library Resources. Retrieved from http://lib.siu.edu/

Stace, W. T. (1960). *Mysticism and Philosophy*. New York: J. B. Lippincott Co.

Stark, R. (1972). The economics of piety: Religious commitment and social class. In G. W. Thielbar & S. D. Feldman (Eds.), *Issues in Social Inequality* (pp. 483 – 503). Boston, MA: Little, Brown and Company.

Stark, R. (1996). Religion as context: Hellfire and delinquency one more time. *Sociology of Religion, 57* (2), 163 – 173.

Stark, R. (2002). Physiology and faith: Addressing the 'universal' gender difference in religious commitment. *Journal for the Scientific Study of Religion, 41* (3), 495 – 507.

Stark, R., & Bainbridge, W. S. (1979). Of churches, sects and cults: Preliminary concepts for a theory of religious movements. *Journal for the Scientific Study of Religion, 18* (2), 117 – 133.

Stark, R., & Glock, C. Y. (1965). The "new denominationalism." *Review of Religious Research, 7,* 8 – 17.

Stark, R., & Glock, C. Y. (1968). *American Piety: The Nature of Religious Commitment*. Berkeley, CA: University of California Press.

Stark, R., & Iannaccone, L. R. (1997). Why the Jehovah's Witnesses grow so rapidly: A theoretical application. *Journal of Contemporary Religion, 12* (2), 133 – 157.

Stark, R., Iannaccone, L., & Finke, R. (1997). *Rationality and the "religious mind*." Retrieved March 23, 2006, from www.religionomics.com/archives/archive/38/rationality-and-the-religious-mind.

Steffensmeier, D. J., Allan, E. A., Harer, M. D., & Streifel, C. (1989). Age and the distribution of crime. *The American Journal of Sociology, 94* (4), 803 – 831.

Stolzenberg, R. M., Blair-Loy, M., & Waite, L. J. (1995). Religious participation in early adulthood: Age and family life cycle effects on church membership. *American Sociological Review, 60,* 84 – 103.

Strawbridge, W. J., Cohen, R. D., Shema, S. J., & Kaplan, G. A. (1997). Frequent attendance at religious services and mortality over 28 years. *American Journal of Public Health, 87* (6), 957 – 961.

Strong, W. S. (1990). *The Copyright Book: A Practical Guide* (3rd Ed.). Cambridge, MA: MIT Press.

Taylor, R. J. (1986). Religious participation among elderly Blacks. *The Gerontologist, 26*, 630 – 636.

Taylor, R. J. (1988a). Correlates of religious non-involvement among Black Americans. *Review of Religious Research, 30*, 126 – 139.

Taylor, R. J. (1988b). Structural determinants of religious participation among Black Americans. *Review of Religious Research, 30*, 114 – 125.

Taylor, R. J., & Chatters, L. (1991). Religious Life. In J. S. Jackson (Ed.), *Life in Black America* (pp. 105 – 123). Newbury Park, CA: Sage Publications.

Taylor, R. J., Chatters, L. M., Jayakody, R., & Levin, J. S. (1996). Black and White differences in religious participation: A multisample comparison. *Journal for the Scientific Study of Religion, 35*, 403 – 410.

Taylor, R. J., Chatters, L. M., Mattis, J. S., & Joe, S. (2010). Religious involvement among Caribbean blacks residing in the United States. *Review of Religious Research, 52* (2), 125 – 145.

Taylor, R. J., Ellison, C. G., Chatters, L. M., Levin, J. S., & Lincoln, K. D. (2000). Mental health services in faith communities: The role of clergy in Black churches. *Social Work, 45* (1), 73 – 87.

Thompson, E. H., & Remmes, K. R. (2002). Does masculinity thwart being religious? An examination of older men's religiousness. *Journal for the Scientific Study of Religion, 41* (3), 521 – 532.

Thornton, A., Axinn, W. G., & Hill, D. H. (1992). Reciprocal effects of religiosity, cohabitation and marriage. *American Journal of Sociology, 98* (3), 628 – 651.

Tourangeau, R., & Smith, T. (1996). Asking sensitive questions: The impact of data collection mode, question format and question context. *Public Opinion Quarterly, 60*, 275 – 304.

Toussaint, L. L., Williams, D. R., Musick, M. A., & Everson, S. A. (2001). Forgiveness and health: Age differences in a U.S. probability sample. *Journal of Adult Development, 8* (4), 249 – 257.

Trochim, W. M. K. (2001). *The Research Methods Knowledge Base* (2nd Ed.). Cincinnati, OH: Atomic Dog Publishing.

Tsoi, K. (2011). *Heterogeneity and Bias*. Hong Kong, China: Clinical Data Management and Statistical Consultation Centre (CliniStat).

*USA Today (2010). *Gallup/USA Today Poll: December Wave 1, Dec, 2010* [Data file and code book]. USAIPOUSA2010-22, Version 2. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

*USA Today and The Cable News Network (1993). *Gallup/USA Today/CNN Poll: Gambling, Sep, 1993* [Data file and code book]. USAIPOCNUS1993-422011. Cornell University, Ithaca, NY: Roper Center for Public Opinion Research, RoperExpress [Distributor], accessed May-29-2017.

U.S. Census Bureau (2017). *Statistical Abstract of the United States: 1954 (Part 2)*. Retrieved from: https://www.census.gov/library/publications/1954/compendia/statab/75ed.html

*U.S. Department of Education, National Center for Education Statistics (2001). *High School and Beyond, 1980: A Longitudinal Survey of Students in the United States* [Data file, code book]. Inter-university Consortium for Political and Social Research (ICPSR 7896). Ann Arbor, MI.

*U.S. Department of Education, National Center for Education Statistics (2002). *National Education Longitudinal Study: Base Year Through Fourth Follow-up, 1988-2000* [Data file, code book and manual]. Inter-university Consortium for Political and Social Research (ICPSR 3955). Ann Arbor, MI.

Uecker, J. E., Regnerus, M. D., & Vaaler, M. L. (2007). Losing my religion: The social sources of religious decline in early adulthood. *Social Forces, 85* (4), 1667 – 1692.

Underwood, L. G., & Teresi, J. A. (2002). The daily spiritual experience scale: Development, theoretical description, reliability, exploratory factor analysis and preliminary construct validity using health-related data. *Annals of Behavior Medicine, 24* (1), 22 – 33.

Van Tubergen, F. (2006). Religious affiliation and attendance among immigrants in eight western countries: Individual and contextual effects. *Journal for the Scientific Study of Religion, 45* (1), 1 – 22.

Vezzoni, C., & Biolcati-Rinaldi, F. (2015). Church attendance and religious change in Italy, 1968 – 2010: A multilevel analysis of pooled datasets. *Journal for the Scientific Study of Religion, 54* (1), 100 – 118.

Viswesvaran, C., & Schmidt, F. L. (1992). A meta-analytic comparison of the effectiveness of smoking cessation methods. *Journal of Applied Psychology, 77* (4), 554 – 561.

Vogt, W. P. (1999). *Dictionary of Statistics & Methodology: A Non-Technical Guide for the Social Sciences* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Wallace, J. M., & Forman, T. A. (1998). Religion's role in promoting health and reducing risk among American youth. *Health Education & Behavior, 25* (6), 721 – 741.

Weintraub, I. (2012). The Role of Grey Literature in the Sciences. Retrieved June 10, 2012, from http://library.brooklyn.cuny.edu/access/greyliter.htm.

Whitener, E. M. (1990). Confusion of confidence intervals and credibility intervals in meta-analysis. *Journal of Applied Psychology, 75* (3), 315 – 321.

Wiehe, S. E., Carroll, A. E., Liu, G. C., Haberkorn, K. L., Hoch, S. C., Wilson, J. S., et al. (2008). Using GPS-enabled cell phones to track the travel patterns of adolescents. *International Journal of Health Geographics, 7* (22), 1 – 11. doi:10.1186/1476-072X-7-22

Williams, L. M., & Lawler, M. G. (2001). Religious heterogamy and religiosity: A comparison of interchurch and same-church individuals. *Journal for the Scientific Study of Religion, 40* (3), 465 – 478.

Wilson, J., & Sherkat, D. E. (1994). Returning to the fold. *Journal for the Scientific Study of Religion*, *33*, 148 – 161.

Woolever, C., & Bruce, D. (2001). *U.S. Congregational Life Survey*. Louisville, KY: Research Services, Presbyterian Church (U.S.A.).

Woosley, S. A. (2005). Survey response and its relationship to educational outcomes among first-year college students. *Journal of College Student Retention, 6* (4), 413 – 423.

Wuthnow, R. (1977). A religious marketplace. *Journal of Current Social Issues*, *15*, 38 – 42.

Wuthnow, R. (1978). Peak experiences: Some empirical tests. *Journal of Humanistic Psychology*, *18*, 59 – 75.

Wuthnow, R. (Ed.) (1979). *The Religious Dimension: New Directions in Quantitative Research*. New York, NY: Academic Press.

Wuthnow, R. (1991). *Acts of Compassion: Caring for Others and Helping Ourselves*. Princeton, NJ: Princeton University Press.

Zech, C., & Gautier, M. L. (2004). Catholic parish organizational structure and parish outcomes. *Journal for the Scientific Study of Religion, 43* (1), 141 – 150.

Zhao, Y., Hu, F., Wang, F., Han, B., Li, D., Li, X. et al. (2009). Meta-analysis of *MSH6* gene mutation frequency in colorectal and endometrial cancers. *Journal of Toxicology and Environmental Health, Part A, 72*, 690 – 697.

Zinnbauer, B. J., Pargament, K. I., Cole, B., Rye, M. S., Butter, E. M., Belavich, T. G. et al. (1997). Religion and spirituality: Unfuzzying the fuzzy. *Journal for the Scientific Study of Religion, 36* (4), 549 – 564.

* Reference was included in the meta-analysis.

APPENDICES

APPENDIX A

EXAMPLE OF A DAILY TIME-USE INTERVIEWER SCRIPT AND RESPONDENT DIARY FROM THE 1965-1966 MULTINATIONAL COMPARATIVE TIME-BUDGET RESEARCH PROJECT

In this booklet we'd like to have you keep a list of all of your activities starting at midnight, running through the daytime on (DAY OF WEEK), up to midnight again (DAY OF WEEK) night. We'd also appreciate it if you could write in the times when you stopped one activity and started on a different one.

These shaded columns leave room to write things like where you were when you were doing the activity written down in this line -- like "eating -- at home," "shopping -- at the drugstore," and things like that. This other column is for noting who was with you, or if you were alone, when you were doing each activity. If you feel like filling in these shaded columns, that would be fine. If you don't want to bother with them, that's all right too. I can fill those in when I come back to go over the diary with you.

What is most important is for you to be as careful as you can to fill in <u>all</u> your activities in the order they occur, here in the column that isn't shaded. For instance, you might: get up, get dressed, eat breakfast, drive to work, work, take a coffee break, work, lunch, etc. We have a list of the kinds of activities we are interested in, here in the booklet. I have an example here, too, of the way a person might fill out the diary. (SHOW R SAMPLE DIARY)

We find that people we talk to can recall their day better if they fill in the booklet at least four times during the day. BREAKFAST, LUNCH, SUPPER and JUST BEFORE GOING TO BED are good times to try to do it.

WHAT YOU DID FROM MIDNIGHT UNTIL 9 IN THE MORNING

Time	What did you do?	Time Began	Time Ended	Where	With Whom	Doing Anything Else?	Remarks
Midnight		12:00					
1 AM							
2 AM							
3 AM							
4 AM							
5 AM							
6 AM							
7 AM							
8 AM							

WHAT YOU DID FROM 9 IN THE MORNING UNTIL 5 IN THE AFTERNOON

Time	What did you do?	Time Began	Time Ended	Where	With Whom	Doing Anything Else?	Remarks
9 AM							
10 AM							
11 AM							
Noon							
1 PM							
2 PM							
3 PM							
4 PM							

WHAT YOU DID FROM 9 IN THE MORNING UNTIL 5 IN THE AFTERNOON

Time	What did you do?	Time Began	Time Ended	Where	With Whom	Doing Anything Else?	Remarks
9 AM							
10 AM							
11 AM							
Noon							
1 PM							
2 PM							
3 PM							
4 PM							

WHAT YOU DID FROM 5 IN THE AFTERNOON UNTIL MIDNIGHT

Time	What did you do?	Time Began	Time Ended	Where	With Whom	Doing Anything Else?	Remarks
5 PM							
6 PM							
7 PM							
8 PM							
9 PM							
10 PM							
11 PM							
			Mid-night				

REVIEW OF TIME DIARY - READ CAREFULLY

ASK EVERYBODY:

Now let's go over what you wrote down in the time diary to make sure we have accounted for everything we need to know. IF R HAS NOT FILLED OUT DIARY, MARK "New diary" ON FRONT OF TIME DIARY AND GO THROUGH THE DAY'S ACTIVITIES FOR THE SCHEDULED DIARY DAY.

IF R WAS SLEEPING AT MIDNIGHT, START WITH FIRST ACTIVITY AFTER GETTING UP.

I see that at midnight on (DIARY DAY), you were _________-ing and that you stopped _________-ing around (TIME ENTERED IN DIARY).

1. Where were you when you (did this)? (The question need not be asked if location is obvious, although entry must be written in.)
ENTER "home," "work," "store," "in transit," etc. IN "WHERE" COLUMN.

2. What other people (and how many people) were involved in this activity--by involved I mean taking part in this with you, or discussing it with you? (People working at separate jobs in the same room are not involved. Typical entries for long periods of time, like work, might be "10 fellow-workers," or "4 clients").
RECORD NUMBER AND RELATION TO R e.g., "wife," "2 sons," IN "WITH WHOM" COLUMN. IF NO ONE ELSE INVOLVED, AS WHEN SLEEPING OR GETTING WASHED AND DRESSED, RECORD "0."

3. Were you doing anything else during this time, like listening to the radio while driving, or knitting while watching TV? (Certain of these simultaneous activities may qualify as main activities, e.g., going to the store while dinner is cooking, and should be recorded like the other main activities. We must leave such decisions to your judgment but review instructions in the Instruction Manual for P491.)
RECORD ENTRIES LIKE "radio" OR "knitting" IN "DOING ANYTHING ELSE" COLUMN; RECORD "0" IF R WAS NOT DOING ANYTHING ELSE.

4. USE "REMARKS" COLUMN TO RECORD
a) (FOR TRAVEL) All means of travel R used, including walks over one minute. Remember each trip is a separate activity.
b) (FOR SHOPPING TRIPS) Types of store R went to.
c) Any useful information R volunteers (or you find out) to help us interpret anything unusual, like the name of a play R goes to.

5. Now the next activity you have listed is (MENTION NEXT ACTIVITY) which began at ________ and ended at ________. (UNLESS OBVIOUS AND LOGICAL) Was there anything you did between this (ACTIVITY) and (PRECEDING ACTIVITY)?
/No/ Ask question sequence 1-5 for this activity.
/Yes/ Record this new activity, the time, and ask question sequence 1-5 for it. If no room is available, use the blank pages on the reverse of the diary sheet and clearly indicate where the activity belongs in R's time schedule.

* * * * *

ITEMS TO REMEMBER:

1) EXPERIENCE INDICATES THAT MOST PEOPLE HAVE AT LEAST 20 ACTIVITIES FOR A DAY, SOME HAVE AS MANY AS 50.

2) CHECK TO MAKE SURE R HAS NOT OMITTED ACTIVITIES (Your example diary and instruction manual contain methods of checking for omitted activities).

3) THERE SHOULD BE AN ENTRY IN EACH COLUMN (EXCEPT REMARKS) FOR EVERY ACTIVITY.

VITA

Graduate School

Southern Illinois University Carbondale

Chad S. Briggs

chad.briggs@judsonu.edu

Wartburg College
Bachelor of Arts, Psychology, May 1997

Southern Illinois University Carbondale
Master of Arts, Applied Experimental Psychology, August 2001

Outstanding Thesis Award, Department of Psychology, SIUC (2001)
Outstanding Thesis Award, Southern Illinois University Carbondale (2001)
Distinguished Master's Thesis Award, Midwestern Association of Graduate Schools (2002)

Dissertation Title:
The Prevalence of Religious Service Attendance in America: A Review and Meta-Analysis

Major Professor: Meera Komarraju, Ph.D.

Publications:
Sundt, J. L., Castellano, T. C., & Briggs, C. S. (2007). The socio-political context of prison violence and its control: A case study of supermax and its effect in Illinois. *The Prison Journal, 88* (1), 94 – 122.

Vaux, A. C., & Briggs, C. S. (2005). Conducting mail and internet surveys. In F. T. Leong & J. T. Austin (Eds.), *The Psychology Research Handbook: A Guide for Graduate Students and Research Assistants* (2nd Ed.). Thousand Oaks, CA: Sage Publications.

Briggs, C. S., Sundt, J. L., & Castellano, T. C. (2003). The effect of supermaximum security institutions on aggregate levels of institutional violence, *Criminology, 41* (4), 1341 – 1376.

Briggs, C. S., & Ribich, F. (1998). The relationship between test anxiety, sleep habits and self-perceived academic competency. *Psych-E, 1*, Sample Experimental Paper. Retrieved from http://truth.boisestate.edu/psyche/archives/vol3/jekzb.html