Journal of Affective Disorders 257 (2019) 428–438

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Review article

Religion, spirituality and depression in prospective studies: A systematic review

Arjan W. Braam[a,b,*], Harold G. Koenig[c,d,e]

[a] Department of Humanist Chaplaincy Studies for a Plural Society, University of Humanistic Studies, Utrecht, The Netherlands
[b] Department of Emergency Psychiatry, Department of Residency Training, Altrecht Mental Health Care, Lange Nieuwstraat 119, 3512 PG Utrecht, The Netherlands
[c] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Box 3400, Durham, NC 27710, USA
[d] Department of Medicine, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[e] School of Public Health, Ningxia Medical University, Yinchuan 750000, PR China

ARTICLE INFO

Keywords:
Religion
Spirituality
Depression
Longitudinal
Systematic review

ABSTRACT

Background: Many empirical studies have shown inverse associations between measures of religiousness and spirituality (R/S) and depression. Although the majority of these studies is cross-sectional, a considerable number of prospective studies have also appeared.

Methods: The current systematic review offers an overview of the major pattern of associations between the measures of R/S and depression / depressive symptoms in 152 prospective studies (until 2017).

Results: With on average two R/S measures per study (excluding measures of religious struggle, treated separately), 49% reported at least one significant association between R/S and better course of depression, 41% showed a non-significant association, and 10% indicated an association with more depression or mixed results. The estimated strength of these associations was modest (d = -0.18). Of the studies that included religious struggle, 59% reported a significant association with more depression (d = +0.30). Especially among persons identified with psychiatric symptoms, R/S was significantly more often protective (d = -0.37). In younger samples and in samples of patients with medical illness, R/S was less often protective. Studies with more extensive adjustment for confounding variables showed significantly more often associations with less depression. Geographical differences in the findings were not present.

Limitations: Given the huge heterogeneity of studies (samples size, duration of follow-up), the current synthesis of evidence is only exploratory.

Conclusion: In about half of studies, R/S predicted a significant but modest decrease in depression over time. Further inquiry into bi-directional associations between religious struggle and (clinical) depression over time seems warranted.

1. Introduction

For decades, quantitative empirical studies have appeared on associations between religiousness/spirituality (R/S) and depression. Depression is often selected as the phenomenon of interest in relationship to R/S because it is a common mental disorder and is often associated with loss of hope and meaning (Dein, 2006). Koenig et al. (2012) concluded from their extensive review of the literature on R/S and depression that by its ability to neutralize life stress, R/S might help both to prevent the onset of depression, and if depression develops, shorten the time it takes to resolve. They emphasize the importance of long-term *prospective* studies that use multidimensional measures of R/S, assessed at multiple time points, and include assessment of parental religiosity, personality, and genetic traits. Apart from possible protective aspects of R/S, the possibility of reverse causation has also been entertained (Li et al., 2016; Maselko et al., 2012; VanderWeele et al., 2016). For example, those who become depressed may subsequently stop participating in religious/spiritual activities. Without longitudinal data, this could explain the seemingly protective association.

Bonelli et al. (2012) provided an overview of studies on R/S and depression related to systematic reviews reported in two successive editions of the Handbook of Religion and Health (Koenig et al., 2001, 2012). Up through 2010, 70 prospective studies were conducted on R/S and depression. Of those studies, 56% reported at least one significant

* Corresponding author at: Department of Emergency Psychiatry, Department of Residency Training, Altrecht Mental Health Care, Lange Nieuwstraat 119, 3512 PG Utrecht, The Netherlands.
E-mail addresses: a.braam@altrecht.nl (A.W. Braam), Harold.Koenig@duke.edu (H.G. Koenig).

https://doi.org/10.1016/j.jad.2019.06.063
Received 4 December 2018; Received in revised form 29 April 2019; Accepted 30 June 2019
Available online 02 July 2019

association between a measure of R/S with *lower* levels of depression at follow-up, 10% reported a significant association with *higher* levels of depression at follow-up, and 24% found either no association or mixed results. The number of R/S measures used in each study, however, was not addressed.

One may question what types of samples the findings in the reviews pertain to. For example, a meta-analysis by Smith et al. (2003), based mostly on cross-sectional studies, reported a weak but consistent association between R/S and lower levels of depression, particularly among those with higher stress levels. A question left unanswered, though, is whether R/S (as a coping response to stressful life situations) might be especially helpful among those with physical or mental illness. Similarly, it remains to be elucidated whether the impact of R/S on depression differs for different age-groups, as suggested by Blazer (2012), since the importance of R/S may vary across age cohorts.

Another concern with respect to the relationship between R/S and depression has to do with the considerable variation in findings between regions, countries and continents (Dein, 2006). Much of this research has been conducted in the US, especially in the Southeastern US, where religion has a more profound impact on social and cultural life. Western Europe, as a second source of empirical research, is more secular than the US (Sahgal, 2018). In contrast with those in the US, studies in Western Europe report less consistent findings with respect to the relationship between R/S and depression (King et al., 2013).

As a construct, R/S has (generally) to do with a worldview that includes beliefs in a transcendent reality that provides meaning and purpose to life, and occurs in a context of religious traditions and communities (religion), a more individual centered belief/activity (spirituality), or both. R/S should be understood as a multidimensional concept (Bergin, 1983) involving beliefs (that may vary widely), public practices, private practices, cognitive processes (intrinsic religious motivation or importance), and various other psychological aspects such as religious coping and attachment styles. In addition, some aspects of R/S reflect a troubled relationship with the deity or religious community, called 'religious struggle', which is often operationalized as "negative religious coping" (Pargament et al., 1998). For example, the latter may involve pessimistic interpretations about punishment and being abandoned by God. The associations between R/S and depression, then, are likely to depend on the particular aspect (and measure) of R/S that is being studied.

Almost twenty years ago, Sloan et al. (1999) expressed concerns about the methodological rigor of studies on R/S and health. They recommended caution when interpreting the results of studies that failed to control for confounding variables (e.g. age, or physical health status). They also emphasized the need to control for multiple comparisons: many studies included multiple measures of R/S and/or multiple outcome measures. Furthermore, Sloan et al. recommended that greater description of the aspects of R/S being measured might improve the research.

The current systematic review focuses on R/S and the course of depressive symptoms over time. The review sought to exhaustively identify quantitative prospective empirical studies that examine the relationship between R/S and depression. The overall research aim was to identify patterns in the relationship between R/S and depression over time that could be identified from these prospective studies. More specific research questions were:

- Which particular aspects (and measures) of R/S seem to be the most prominent or relevant with respect to the association with depression over time?
- Which other factors, related to the types of samples, may be important in understanding these associations (e.g. stage of life, physical or mental health problems, geographic region)?
- Do findings depend on how depression has been operationalized (continuous / categorical)?
- Do findings depend on the quality of the methods and statistical approach?

Table 1
Inclusion criteria.

emotional distress; psychological distress; mental distress; depressive symptoms; depressed mood; depressive mood; depression; depressive disorder
AND
religion; religious; religiosity; religiousness; spiritual; spirituality; god; prayer; mosque church; church attendance; religious attendance; synagogue; synagogue attendance
AND
prospective; longitudinal; follow-up; waves; baseline; course; trajectories; predictive; prognostic; recovery; recovering; multiwave; over time

2. Methods

2.1. Search strategy and selection criteria

The current systematic review followed the AMSTAR ('A MeaSurement Tool to Assess systematic Reviews') guidelines, 2007 version (Shea et al., 2007; **Supplementary Table A**). The search strategy included the criteria as shown in **Table 1**. Mesh terms were not utilized as these did not identify a sufficient number of relevant studies. Studies were selected when they were published in English, and when an abstract was available. Studies or trials evaluating a religious or spiritual intervention or therapy - ranging from R/S counselling (pastoral care), prayer or meditation techniques, religion-adapted psychotherapy, palliative care intervention, to psilocybin-induced mystical experience - were excluded when the intervention was the only aspect of R/S being studied. Similarly, studies that focused on 'forgiveness' and 'gratitude' were excluded. These constructs (similar as altruism and moral reasoning) imply interpersonal consequences of R/S. Levels of religiousness of parents or partners may represent an important environmental aspect of one's religious life, but were not included because this would require a complex analysis of the specific interaction between parental or spousal religiousness and the religiousness of the individual. Measures of 'meaning' were only included when they had a direct connection with R/S. When two or more studies reported about the same data, either the first study or the study with the most accurate details was selected.

Depressive symptoms, depression according to a criterion (cut-off score on a screening scale), and Major Depressive Disorder represented the primary outcome variable. Mental distress served as a search variable because it is frequently assessed with depressive symptom scales or similar measures, such as those assessing negative affect. Associations of R/S with other common mental disorders, such as anxiety disorders, prolonged or complicated grief, or alcohol abuse, were excluded. Furthermore, quality of life or positive psychological outcomes such as well-being, posttraumatic growth or resilience were not included in this review.

All studies included in this review had depression data collected at two time points and included control for baseline depression. Studies were excluded from selection if they used successive cross-sectional analyses for each follow-up assessment, or did not describe associations between measures of R/S at baseline and course of depression or depressive symptoms over time. A minority of the studies also examined whether depression predicted R/S – such as in cross-lagged analyses.

2.2. Procedure

The search strategy in PubMed and PsycInfo was carried out on 30th June 2015 and was repeated on 17th July 2017, with an update conducted on 1st September 2018 (including publications until 2017). As shown in Fig. 1, there was substantial overlap between PubMed and Psycinfo: about two-thirds of all unique papers were simultaneously identified in both systems. Furthermore, 41 papers were added from other formal sources (Koenig et al., 2001; Koenig et al., 2012; Crossroads 2007–2018; and an MBASE search on 28th January 2019,

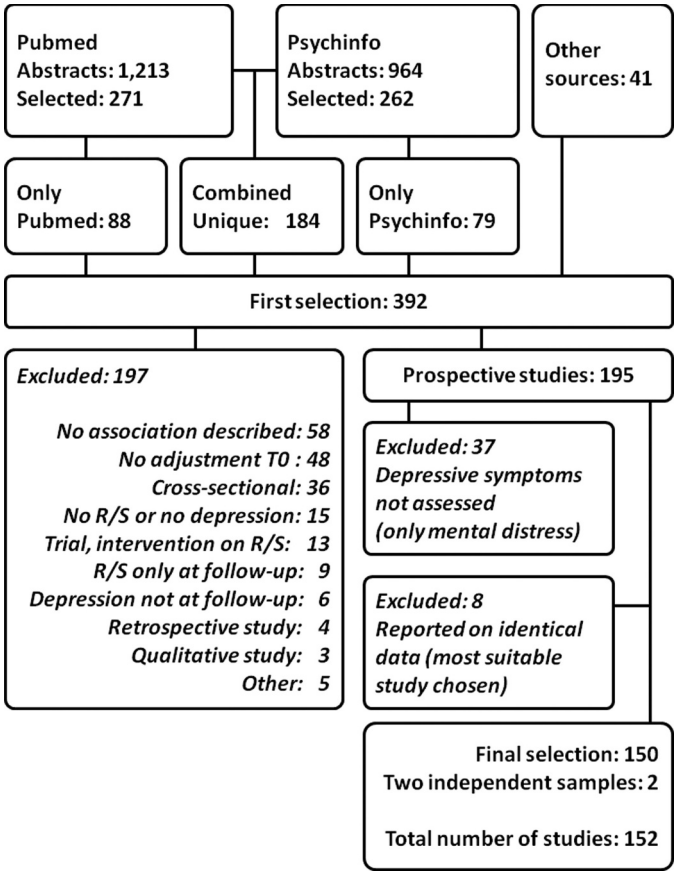

Fig. 1. Flowchart on the inclusion of papers for the current systematic review (note: 'Other sources' also included an EMBASE search [534 abstracts, 11 selected] with exclusion of papers that had also been identified in Pubmed and Psychinfo).

excluding papers that had already been identified in Psychinfo and Pubmed) and from the personal collections from both authors. In total, 195 papers contained results from prospective studies on R/S and course of depression, depressive symptoms or mental distress that met search criteria. Thirty-seven papers focused on mental distress, but not specifically on depressive symptoms or depressive disorder and were excluded, leaving 152 studies for the current systematic review summarized in **Table 2** (full reference list is provided in **Supplementary Table B**).

2.3. Data extraction

Utilizing a scoring form to organize the characteristics and results of studies, the papers were organized by author, year of publication, country of origin, sample size, gender, mean age, type of sample, duration of follow-up (in weeks), and number of follow-up assessments.

The types of R/S variables assessed in studies were also determined. The following categories were used for religiousness: religious attendance (organizational religious activities); private religious behavior (non-organizational, e.g. frequency of prayer); importance of religion (intrinsic religious motivation, salience of religion, centrality); religious denomination; positive religious coping; religious struggle (negative religious coping or R/S distress); other measures of religiousness, such as daily spiritual experiences, positive attachment to God, self-rated religiousness, and religious beliefs; and, finally, measures of religiousness that combined several distinct aspects ("composite measures"). For spirituality, the types of spirituality scales were specified: the FACIT-Sp (Functional Assessment of Chronic Illness Therapy - Spiritual Well-being), the Spiritual Well-being Scale, and a range of other measures, such as importance of spirituality.

The type of depression assessment was recorded, such as whether it was a scale assessing depressive symptoms on a continuum, depression based on a cut-off score, or depressive disorder based on a diagnostic algorithm. The type of statistical analysis was categorized as linear regression, logistic regression or Cox proportional hazards regression, advanced longitudinal modeling, and other (basic) statistical models. Adjustment for possible confounding or explanatory variables was categorized as: none, demographic only, or other variables.

The results related to the effects of R/S on depression were categorized as *non-significant*, a significant *decrease* of depressive symptoms/depression, or significant *increase* of depressive symptoms/depression. Although comparison of effect sizes would have been feasible, a provisional, exploratory approach in this systematic review was chosen for several reasons, including the extreme variation in duration of follow-up, the heterogeneity of the analytical models (e.g., advanced methods such as latent growth modeling that provide a different way of showing associations, hardly comparable to regular coefficients), number of R/S variables in each of the studies, and the distinction between models in which one R/S variable was examined, versus models in which more than one R/S variable was entered simultaneously.

2.4. Quality assessment

The quality of the studies was determined by applying the NIH *Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies* (National Institute of Health (2018), retrieved from: www.nhlbi.nih.gov/health-topics/study-quality-assessmenttools). This is a 14-item (no / yes) scale covering a range of quality criteria for observational cohort studies. Furthermore, it was determined whether identifying if R/S would predict the course of depressive symptoms was stated as a specific aim of the paper. Finally, a general judgment on the quality of the study was made by the first author, based on the Quality Assessment Tool score, but also considering the organization of the paper and the types of variables on R/S assessed (with a lower ranking in case of composite measures, 'spiritual wellbeing' variables, inclusion of several R/S variables in one statistical model, absence of adjustment for multicollinearity).

2.5. Synthesis of results

First, an overview is provided with respect to the characteristics of the studies. Next, the main pattern of results is reported for each category of R/S measure, and whether the findings were non-significant, or reported a decrease or increase in depression. The level of statistical significance was defined by the standards used in each of the studies. When there were more measures of religiousness than one in a study, the following, provisional, exploratory 'vote-counting' strategy was used: (a) when there were no significant associations with depression for each of the religiousness measures included in a study, the (general) result was classified as *"non-significant";* (b) when there was at least one association of a religious variable with less depression over time, and when there were no significant associations of religious variables with more depression, the (general) result of the study was classified as *"less depression"*; and (c) when there was at least one significant association of a religious variable with more depression (irrespective of whether other R/S variables were associated with less depression), the (general) result of the study was classified as *'more depression or mixed results'*. Measures of religious struggle were considered separately as this construct is quite different from other R/S measures, and is often unrelated or inversely related to religious involvement (Koenig, 2018). Measures of spirituality were considered separately as well, because many of these assess positive emotions (peacefulness, social connectedness, meaning/purpose in life), leading to tautology in associations with depression. Finally, the distribution of results from each study was

Table 2
Studies included in the current systematic review of studies describing associations between religiousness/spirituality (R/S) and depression or depressive symptoms over time (N = 152): outline of main features and results.

Author(s)	Year	Type of sample[a]	N	Duration of follow-up (weeks)	Depression assessment[b]	R/S: Any significant association with depression[c]	Religious struggle: significant association with depression[d]	Quality of the paper[e]
Ahrenfeldt et al.	2017	comm	10,151	468	d	1		2 (8)
Ai et al.	2010	somat	262	140	d	3	0	3 (11)
Balbuena et al.	2013	comm	12,583	728	MDD	1		4 (10)
Barton et al.	2013	young	173	520	MDD	1		2 (7)
Bekke - Hansen et al.	2014	somat	85	26	d	0		2 (10)
Blalock et al.	1995	somat	265	26	d	0		3 (12)
Bosworth et al.	2003	psych	114	52	d	1	0	4 (10)
Braam et al.	1997	comm	177	52	D	1		2 (8)
Braam et al.	2004	comm	1840	312	d	3		4 (11)
Braam et al.	2007	comm	1346	156	d	0		2 (11)
Bradshaw et al.	2015	comm	1024	150	d	1		4 (12)
Brown et al.	2004	other	103	104	d	0		2 (10)
Carpenter et al.	2012	young	111	7	d	0		1 (6)
Carr and Sharp	2014	other	210	78	d	1		2 (10)
Chan, et al.	2015	young	584	208	d	0		3 (9)
Chan, et al.	2001	somat	53	156	d	0		1 (6)
Chang et al.	2003	other	1227	104	d	1		3 (10)
Cheadle et al.	2015	other	702	76	d	1		3 (9)
Chen et al.	2007	psych	1610	26	d	1		2 (7)
Choi et al.	2012	other	31	9	d	1		2 (6)
Cohen et al.	2006	young	608	520	d	0		2 (6)
Coleman et al.	2011	comm	58	52	D	1		1 (7)
Cotton et al.	2013	somat	132	52	d		2	4 (12)
Davis, et al.	2017	somat	241	52	d	1		2 (10)
Davis, R.F. 3[d], & Kiang	2016	young	180	208	d	0		2 (10)
Dew et al.	2009	psych	104	25	d	0	2	2 (8)
Ellison and Flannely	2009	comm	607	130	MDD	1		3 (8)
Ensminger and Juon	2001	other	530	1100	d	1		2 (8)
Fenix et al.	2006	other	175	56	MDD	1		3 (11)
Fisch et al	1997	other	327	9	d	0		1 (8)
Fitchett et al.	1999	somat	96	16	d	0	0	2 (11)
Ghesquiere et al.	2013	other	65	52	d	1		3 (10)
Gitlin et al.[f]	2007	comm	129	52	d	1		3 (10)
Gitlin et al.[f]	2007	comm	151	52	d	0		3 (10)
Goeke - Morey et al.	2014	young	667	52	d	3[‡]		2 (8)
Graham et al.	2002	other	163	7	d	0		2 (8)
Greenfield and Marks	2007	comm	4646	260	d	1		1 (8)
Greeson et al.	2015	other	213	9	d	1		2 (8)
Hayward and Krause	2014	other	206	314	d		2	2 (6)
Hayward et al.	2012	psych	386	12	d	1		2 (7)
Hebert et al.	2009	somat	284	40	d	0	2	4 (10)
Helms et al.	2015	young	313	52	d	0		3 (9)
Hickman et al.	2013	somat	98	52	d	0	0	3 (10)
Holt et al.	2017–8	comm	756	260	d	1	2	3 (9)
Horowitz and Garber	2003	young	196	312	MDD	0[‡]		3 (9)
Hsu	2014	comm	3537	208	d	3		2 (7)
Hu et al.	2017–8	comm	1270	260	d	0		2 (8)
Hui et al.	2017	other	230	156	d	1		2 (5)
Hunsberger et al.	2002	young	336	104	d	0	0	1 (6)
Huta and Hawley	2010	psych	38	12	d	1		2 (10)
Huynh et al.	2017	somat	234	8	d	0		2 (10)
Idler and Kasl	1992	comm	1447	156	d	1		3 (11)
Impett et al.	2011	young	587	104	d	0		2 (8)
Jung	2017–8	comm	1635	520	d	0		3 (8)
Kasen et al.	2012	young	185	520	MDD	1		2 (8)
Kennedy et al.	1996	comm	1855	104	D	2		3 (9)
Kim et al.	2015	psych	232	26	D	1		4 (10)
King et al.	2007	other	422	52	d	1		3 (9)
King et al.	2013	somat	113	10	d	0[o]		2 (8)
Kivelä et al.	1996	comm	679	260	MDD	1		2 (7)
Koenig et al.	1992	somat	202	26	d	1		2 (8)
Koenig et al.	1998	psych	86	47	MDD	1		4 (10)
Koenig	2007	psych	865	15	MDD	1		4 (10)
Korenromp et al.	2009	other	147	5	d	0		2 (7)
Krause	2009	comm	818	104	d	1		3 (6)
Krause	2012a	comm	718	52	d	0		2 (10)
Krause	2012b	comm	501	52	d	1		2 (8)
Latkin and Curry	2003	other	818	38	d	1		3 (8)
Law and Sbarra	2009	comm	791	416	d	1		3 (9)
Le et al.	2007	young	13,317	52	d	2		2 (10)
Leeson et al.	2015	somat	220	52	d	1		3 (11)

(*continued on next page*)

Table 2 (*continued*)

Author(s)	Year	Type of sample[a]	*N*	Duration of follow-up (weeks)	Depression assessment[b]	R/S: Any significant association with depression[c]	Religious struggle: significant association with depression[d]	Quality of the paper[e]
Lefevor et al.	2017	young	12,825	18	d	0		2 (7)
Leurent et al.	2013	other	8318	52	MDD	2		3 (11)
Levin et al.	1996	comm	624	572	d	1[o]		3 (9)
Li, Okereke, et al.	2016	other	48,984	416	D	1[‡]		3 (11)
Lieberman and Winzelberg	2009	somat	91	26	d	0		2 (8)
Lo et al.	2010	somat	239	8	d	1		3 (10)
Lowell et al.	2017	other	83	156	MDD	0		2 (10)
Magyar-Russell et al.	2013	somat	70	8	d	0	2	3 (9)
Mann et al.	2008a	other	307	36	d	1		4 (12)
Mann et al.	2008b	other	16	43	d	0		1 (12)
Manne et al.	2003	other	207	26	d	1		3 (10)
McFarland	2010	comm	1024	156	d	1		2 (9)
McIntosh et al.	2011	other	890	156	D	1		1 (5)
Mihaljevic et al.	2016	psych	99	52	d	1		2 (12)
Miller et al.	1997	psych	60	520	MDD	1		1 (4)
Miller et al.	2012	young	114	520	MDD	1		1 (6)
Miller and Saunders	2011	psych	55	12	d	0		2 (7)
Min et al.	2016	comm	4098	208	d	2		3 (9)
Monserud and Markides	2017	other	385	364	d	1		4 (12)
Morgan et al.	2017	other	263	70	d	0[o]		2 (8)
Mosqueiro et al.	2015	psych	143	4	d	1		1 (7)
Murphy and Fitchett	2009	psych	136	8	d	1		3 (9)
Musick et al.	1998	somat	103	156	d	0		2 (10)
Musick et al.	2000	comm	1897	156	d	3		2 (9)
Musick and Wilson 65-[f]	2003	comm	2043	416	d	0		2 (10)
Musick and Wilson 65 +[f]	2003	comm	305	416	d	1		2 (10)
Nasser and Overholser	2005	psych	62	12	d	0		3 (9)
Norton et al.	2008	comm	2989	156	MDD	3		3 (9)
Oates and Goode	2013	comm	2780	156	d	1		2 (7)
Pakenham	2008	other	232	52	d	1		2 (8)
Pargament	2004	somat	239	91	d	0	2	3 (7)
Park et al.	1990	young	83	8	d	1		1(6)
Park et al.	2011	somat	101	12	d		0	2 (9)
Park and Dornelas	2012	somat	56	4	d	2	2	2 (7)
Park et al.	2014	somat	111	12	d	0		2 (8)
Park et al.	2017	comm	937	156	d	1	2	2 (8)
Paunesku et al.	2008	young	4791	52	MDD	0		3 (9)
Payman and Ryburn	2010	psych	94	104	d	1		3 (11)
Pérez et al.	2009a	somat	180	26	d	1		2 (8)
Pérez et al.	2009b	young	1096	52	d	0		2 (8)
Peselow et al.	2014	psych	84	8	d	1		2 (6)
Petts	2014	young	5736	46	d	1		2 (8)
Pirutinsky et al.	2011	psych	80	2	d		2[o]	3 (6)
Pössel et al.	2011	young	273	16	d	1[o]		3 (9)
Rasic et al.	2011	comm	1005	520	MDD	0		3 (9)
Rasic et al.	2013	young	976	104	D	1[‡]		4 (10)
Reynolds et al.	2014	somat	128	104	d	1[o]	0[#]	4 (13)
Riley et al.	2016	young	1777	12	d	0		2 (8)
Roh et al.	2015	comm	6647	156	d	1		3 (11)
Ronneberg et al.	2016	comm	7732	104	D	3		3 (9)
Rose et al.	2009	somat	142	52	d	1		1 (9)
Rosmarin et al.	2013a	psych	47	1	d	1	0	3 (10)
Rosmarin et al.	2013b	psych	159	2	D	1		2 (9)
Rush et al.	2016	somat	58	16	d	1		2 (9)
Sallquist et al.	2010	young	136	68	d	0[o]		2 (8)
Schettino et al.	2011	psych	148	7	d	1[g]		2 (8)
Schnittker	2001	comm	2836	156	d	3[o]		4 (10)
Sherman et al.	2009	somat	94	13	d	0[o]	2[o]	3 (12)
Smokowski et al.	2014	young	4036	52	d	0		3 (10)
Strawbridge et al.	2001	comm	2676	1500	d	1		3 (9)
Subramaney et al.	2015	other	102	12	d	0		1 (6)
Sun et al.	2012	comm	75	208	d	1		4 (11)
Szczesniak et al.	2017	other	112	104	d	1	2[#]	3 (11)
Teel et al.	2001	other	83	21	d	0		1 (9)
Toussaint et al.	2012	comm	966	26	MDD	0		2 (7)
Trevino et al.	2010	somat	329	78	d	0	2	3 (9)
Trevino et al.	2016	somat	111	78	d	2		2 (10)
Vander Ploeg-Booth et al.	2008	young	4791	52	MDD	1		3 (10)
Van Voorhees et al.	2008	young	4791	52	D	1		2 (8)
Wadsworth et al.	2009	other	76	26	D	0	0	2 (6)
Weissman et al.	1978	psych	94	208	D	0		1 (4)
Wijngaards-de Meij et al.	2005	other	219	61	d	0		3 (8)
Wilcox et al.	2015	other	253	104	d	1		3 (9)

(*continued on next page*)

Table 2 (*continued*)

Author(s)	Year	Type of sample[a]	*N*	Duration of follow-up (weeks)	Depression assessment[b]	R/S: Any significant association with depression[c]	Religious struggle: significant association with depression[d]	Quality of the paper[e]
Wink et al.	2005	comm	184	1560	d	0		1 (4)
Yanez et al.	2009	somat	399	52	d	1		4 (11)
Yang et al.	2017	young	2239	26	d	3		3 (9)
Yeager et al.	2006	comm	2930	208	d	0		3 (10)
Ysseldyk et al.	2013	comm	7021	312	d	1		1 (4)
Zhang et al.	2013	other	128	135	MDD	0		2 (8)
Zou et al.	2014	comm	754	1040	d	1		3 (8)
Zunzunegei et al.	2002	other	119	52	D	0		2 (9)

[a] young = youth / adolescents / students; somat = patients with a somatic condition; psych = persons identified with psychiatric symptoms (mostly depression); other = other groups (see Table 2); comm = community studies.

[b] depression assessment: *d* = at symptom level; *D* = *a* single criterion was used (e.g. cut-off score, clinical diagnosis); MDD (Major Depressive Disorder) = after application of a diagnostic algorithm.

[c] 0 = non-significant; 1 = significant association of at least one R/S measure (religious struggle excluded) with less depression over time; 2 = significant association with more depression over time; 3 = mixed results.

[d] 0 = non-significant; 1 = significant association of religious struggle with less depression over time; 2 = significant association of religious struggle with more depression over time; 3 = mixed results.

[e] Based on the score on the NIH *Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies* (between brackets, range 0–13), as well as the organisation of the paper and the accuracy of analyzing the aspects of R/S under study: 1 = poor, 2 = fair, 3 = good, 4 = excellent.

[f] Two independent samples with valid, separate results, described in one publication.

[g] Compared to lower ánd higher levels, intermediate level of religious behaviour was associated to less depression at follow-up.

[‡] cross-lagged analyses: significant association of depression predicting lower levels of R/S.

[#] cross-lagged analyses: significant association of depression predicting higher levels of R/S.

[o] cross-lagged analyses: depression did *not* significantly predict R/S.

Table 3
Types of samples in 152 prospective studies on Religiousness / Spirituality and depressive symptoms / depression.

	N	%
Patients with a physical condition	29	19
Youth / adolescents / students	26	17
Community / general population, ≥ 60 years	26	17
Psychiatric patients or those identified with serious psychiatric symptoms	21	14
Community / general population, 25–60 years	15	10
Persons with grief	8	5
Caregivers	8	5
Pregnancy	5	3
General practice	3	2
Disaster survivors	2	1
Other	8	5

analyzed for subgroups with respect to type of sample, type of analytical method, and quality of the study. Possible differences between the subgroups were tested using chi-square statistic.

As indicated above, the current comprehensive review included too many heterogeneous constituents (types of variables, research designs, types of coefficients, analytical models) to conduct at a straightforward meta-analysis. Nevertheless, and in spite of violating assumptions about sufficient homogeneity necessary for a meta-analysis (Kuijpers, 2016), effect sizes (Cohen's d) were calculated for the majority of the studies (see footnote under **Supplementary Table C**) to arrive at a general impression of the strength of associations, both for studies with general measures on R/S (excluding those on spiritual well-being) and for the studies that examined religious struggle. Mean effect sizes were computed for each of these variable types. Next, mean effect-sizes were compared between relevant subgroups, using analysis of variance (F-statistic), both unweighted as well as weighted for sample size (logarithmic).

3. Results

3.1. Study characteristics

3.1.1. Country of origin

The first study, by Weissman and colleagues, appeared in 1978. Half of the studies appeared after 2010. Most studies were conducted in North-America: 114 in the US (75%) and 5 in Canada. Of the studies in the US, 19 originated from states matching the region described as the 'Bible Belt.' Sixteen studies were conducted in Europe, 10 in Eastern Asia, three in Australia, two in Africa, one in South America (Brazil) and one in Israel. No studies were conducted in North Africa or in Islamic countries in the Middle East.

3.1.2. Samples

Sample sizes varied widely (range 16 to 48,984), with a mean size of 1533 (SD = 4554) and median sample size of 204. **Table 3** summarizes the main types of samples. Community-based studies were the most frequent type, often involving adults in later life, followed by samples of patients with medical problems, young persons (youth, adolescents, students), and psychiatric patients or those identified with serious psychiatric symptoms.

3.1.3. Assessments

The majority of studies (60%) had only one follow-up after baseline, whereas 18% had two follow-up assessments, 11% three, and 11% four or more (up to nine assessments). With respect to the duration of follow-up, the studies also demonstrated wide variation ranging from one week to 1560 weeks, with an average of 153 weeks (SD = 242) and median of 52 weeks.

Depression was assessed as level of depressive symptoms in 120 studies (79%) and as a clinically relevant syndrome (based on cut-off scores or diagnostic measure) in 32 studies. Among assessment tools, the Center for Epidemiological Studies-Depression scale (CES-D) was used in 67 studies (44%), followed by the Beck Depression Inventory

(BDI) in 13 studies (9%). Nineteen studies (13%) examined depression as Major Depressive Disorder, most employing a semi-structured diagnostic interview such as the Diagnostic Interview Schedule or Composite International Diagnostic Interview.

3.1.4. Analytical approach

The most common method of analyzing the data was linear regression (61 studies; 40%). Regression techniques used to analyze a dichotomized outcome (logistic regression, Cox proportional hazard regression) were used in 29 studies. In 43 studies, the researchers chose an advanced longitudinal modelling technique such as multilevel analysis, structural equation modeling, general estimated equations, or growth curve modeling. More basic analytical procedures, such as partial correlations, especially in older studies and in smaller samples, were utilized in the remaining 19 studies.

In one-sixth of studies ($N = 22$), no adjustment was made for confounding or explanatory variables. In 33 studies, adjustment was made for demographic variables only. In the other studies, additional adjustments were made for physical health ($N = 54$), social support ($N = 50$), or other variables (such as psychological resources, treatment, stress, cognitive ability, life events, or substance abuse; $N = 62$).

3.1.5. Quality of studies

Scores on the NIH *Quality Assessment Tool* (QAT) were normally distributed with a range from 4 to 13, a mean score of 8.7 and a median of 9.0. Overall study quality was judged as 'poor' for 18 studies (mean QAT score = 6.5), 'fair' for 67 studies (QAT = 8.3), 'good' for 51 studies (QAT = 9.5), and 'excellent' for 16 studies (QAT = 10.9; $F = 32.9$, df 3 / 148, $P < .001$). Although most studies formulated a specific research question related to R/S as predictor of depression, a substantial minority (27%; $N = 41$) described R/S as one of a range of other predictors.

3.2. Patterns of results

3.2.1. Main findings

Table 4 presents an overview of the main findings with respect to the ability of R/S to predict changes in depression over time. The most common measure of R/S was religious attendance employed in 45% of studies. Importance of religion, positive religious coping, private religious behavior, and religious denomination were used less often, each in about one fifth of studies. Most studies utilized more than one measure of R/S, with a range of between one and six measures (excluding religious struggle and spiritual well-being measures: mean = 1.9; median = 2.0; $N = 138$).

Religious attendance and importance of R/S did not predict change in depression in about 60% of the studies, and were associated with less depression in about 40% of studies.

Positive religious coping, private religious behavior, and religious denomination were less likely to predict lower rates of depression over time. 'Composite' religious variables (combining measures of religious attendance, motivation, and contents of beliefs) were more likely to predict less depression over time. Religious struggle never predicted less depression, but in 59% of studies predicted an increase in depression.

Among studies assessing spirituality ($N = 23$), the FACIT-Sp was the most commonly used measure (7 studies), followed by the Spiritual Well-Being Scale (4 studies). Both scales explicitly examined aspects of spiritual well-being such as meaning and purpose, peace, or existential well-being, potentially confounding the association with depression. Indeed, most studies (8 out of 11 studies) using spiritual well-being scales showed significant associations with less depression over time. The results in studies ($N = 12$) with other spirituality measures were less pronounced (Table 4).

The major pattern of findings appeared from the "general result" per study (lowest line in Table 4), excluding results for religious struggle and spiritual well-being measures: 49% of the studies reported at least one association with R/S predicting less depression, 41% showed a non-significant association, and 10% indicated an association with more depression or mixed results.

3.2.2. Regions and types of samples

With respect to region, there was no clear evidence that studies conducted in the US 'Bible Belt' were more likely to find that R/S predicted less depression over time (**Table 5**). Instead, this seemed to occur more often in the rest of the US and Canada, and less often in studies from Europe and the Far East.

With respect to age, non-significant results were more often found in samples with a mean age below 25 ($p = .018$). In the comparison between the *types of samples*, a highly significant difference was found ($p < .001$). In samples of younger groups and in patients with medical illness, R/S at baseline was less likely to predict depression at follow-up. However, in three-quarters of studies reporting on samples with persons identified as having psychiatric symptoms, R/S was more likely to predict significantly less depression.

Table 4
Main pattern of statistically significant results in 152 prospective studies on Religiousness / Spirituality (R/S) and depressive symptoms / depression.

Type of variable	Predictor included *N* (%)[a]	Non-significant%	Significantly less depression%	Significantly more depression %
Religious attendance	69 (45)	55	44	1
Importance / salience of religion	32 (21)	63	34	3
Positive religious coping	28 (18)	71	21	7
Private religious behavior	28 (18)	75	21	4
Religious denomination	23 (15)	61	22	17
Religious struggle	22 (14)	41	0	59
Other measures (beliefs / God)	47 (31)	62	32	6[c]
Composite measures	17 (11)	41	47	12
Spiritual well-being	11 (7)	27	73	0
Spirituality - other measures	12 (8)	58	25	17
'General result' per study, based on all measures on R/S *except* religious distress and spiritual well-being	138 (91)[b]	41	49	10[d]

[a] between brackets: percentage of total number of studies in the review ($N = 152$); the cumulative percentage is above 100% because most studies had more than one measure on R/S.

[b] four studies were not included as these studies had data on religious struggle only, and ten other studies were not included as these studies had data on spirituality measures only (often about spiritual well-being).

[c] for one study, the results were mixed: one variable was associated with more depression, one other with less.

[d] for 5%, there were mixed results - some variables were associated with more depression, others with less.

Table 5
Main patterns of results of prospective associations between Religiousness / spirituality (R/S) and depression: distribution of the 'general result' per study across regions, age-groups, types of samples, level of depression, types of analytical modelling, and quality of studies.

Type of variable		*n*	Non-significant%	Less depression%	More depression or mixed results%	χ^2 (P)
Region	US Bible belt	19	47	42	11	9.9 (0.130) df 6
	other US (and Canada)	87	36	58	7	
	Europe	16	50	31	19	
	Far East	10	40	30	30	
Age	< 25	27	**52**	**37**	**11**	**12.0 (0.018) df 4**
	25–60	64	**45**	**52**	**3**	
	≥60	47	**26**	**55**	**19**	
Sample	Youth / adolescents / students	27	**48**	**41**	**11**	**32.9 (< 0.001) df 10**
	Patients with a physical condition	23	**70**	**17**	**13**	
	Persons with psychiatric symptoms	19	**26**	**74**	**0**	
	Other types of samples	30	**40**	**57**	**3**	
	Community 25–60 years	15	**40**	**60**	**0**	
	Community ≥ 60 years	24	**13**	**58**	**29**	
Level of depression	Depressive symptoms (scales)	106	43	47	9	2.4 (0.301) df 2
	Criterion-based depression	32	28	59	13	
Duration of follow-up	Less than 26 weeks (6 months)	30	53	43	3	4.9 (0.300) df 4
	26–104 weeks (6–24months)	57	40	47	12	
	> 24 months	51	31	57	12	
Number of follow-ups	One	82	43	48	10	5.3 (0.254) df 4
	Two	25	36	44	20	
	Three or more	31	36	61	3	
Adjustment	No adjustment	20	**60**	**40**	**0**	**11.3 (0.023) df 4**
	Adjustment for demographics	29	**45**	**55**	**0**	
	Adjustment for other confounders	89	**34**	**51**	**16**	
Number of participants	*N*: < 100	26	**54**	**42**	**4**	**19.4 (0.004) df 6**
	N: 100–249	38	**50**	**47**	**3**	
	N: 250–999	36	**33**	**61**	**6**	
	N: ≥ 1000	38	**26**	**47**	**26**	
Number of R/S variables	one	68	**46**	**50**	**4**	**11.9 (0.018) df 4**
	two	29	**48**	**45**	**7**	
	three or more	41	**24**	**54**	**22**	
Analytical model	Linear regression	54	50	41	9	5.3 (0.501) df 6
	Logistic regression	29	28	62	10	
	Advanced longitudinal modelling	38	34	53	13	
	Other / basic	17	41	53	6	
Research question R/S	No	35	54	37	3	4.1 (0.128) df 2
	Explicit	103	35	54	11	
Quality	Low quality assessment score ≤ 7	31	42	52	7	1.1 (0.900) df 4
	intermediate (8–10)	87	40	49	10	
	high (11–13)	20	35	50	15	
Judgement	Poor (mean quality score 6.5)[b]	16	**50**	**50**	**0**	**12.9 (0.045) df 6**
	Fair (mean quality score 8.3)	62	**48**	**42**	**10**	
	Good (mean quality score 9.5)	46	**35**	**50**	**15**	
	Excellent (mean quality score 10.9)	14	**7**	**86**	**7**	

N = 138 (four studies with 'religious struggle' as single predictor were excluded, as 'religious struggle' was recognized as a critically different aspect of R/S; ten studies with measures of spiritual well-being were excluded to prevent possible tautological associations with depression. Significant results are shown in **bold**.
[b] Analysis of variance $F = 33$, df 3 / 148, $P < .001$.

3.2.3. Design

Studies that assessed depression as a categorical variable instead of a continuous variable were somewhat more likely (but not statistically more likely) to report that R/S predicted a decrease in depression over time. Studies with shorter follow-up periods (< 6 months) were less likely (but not statistically less likely) to report R/S predicting less depression compared with studies having longer follow-up periods. Number of follow-up assessments did not make a difference, although studies that assessed depression at three or more follow-ups were more likely to report that R/S predicted less depression over time. Studies without statistical adjustments were more likely to report non-significant results compared to those with adjustment for demographics or other confounding variables ($p = .023$). Studies with larger samples (> 250) were more likely to report significant results than studies with smaller samples ($p = .004$). Studies with three or more R/S variables were less likely to report non-significant results and more likely to report mixed results ($p = .018$).

With respect to statistical approach, studies utilizing linear regression analysis or more basic types of analyses were more likely to report non-significant associations. In contrast, studies applying logistic regression or advanced longitudinal models were more likely (but not statistically more likely) to find that R/S predicted less depression over time.

3.2.4. Quality

Studies that stated among their specific aims the goal of determining whether R/S predicted depression over time did not more often show significant associations. Studies with the lowest quality assessment scores were less likely (but not statistically less likely) to yield significant results, compared to studies with the highest quality scores where significant results were present in about 75% of studies. With respect to differences between the studies based on overall judgment of quality, a significant difference ($p = .045$) was found with 'excellent' studies more often producing significant results. In those studies, over 80% reported that R/S predicted less depression over time. These studies were also more likely to report mixed results.

3.2.5. Estimated effect-sizes

Effect-sizes for all studies (as far as possible, one effect-size for the most prominent association per study) are shown in **Supplementary**

Table 6
Distribution of estimated effect-sizes (Cohen's d) for associations between general measures on R/S and depression over time (excluding measures of spiritual well-being) and for associations between religious struggle/distress and depression over time.

	d	Qualification of effect-size	General measures of R/S *N* = 131 %	Religious struggle *N* = 18 %
	< 0.8	Large	3	0
Negative effect	−0.8 to −0.5	Intermediate	7	0
	−0.5 to −0.2	Small	32	0
	−0.2 to 0.0	No	29	4
Zero (no details)[a]	[0.0]		16	30
	0.0 to 0.2	No	7	11
Positive effect	0.2 to 0.5	Small	5	35
	0.5 to 0.8	Intermediate	1	16
	> 0.8	Large	0	4

[a] Only the absence of a significant association had been reported in the study, therefore, value '0.0' was assigned as effect-size.

Table C. The mean Cohen's d effect size ($N = 130$ studies, excluding those measuring only religious struggle or spiritual well-being) was −0.18 (median −0.18; SD 0.28; range −1.15 to 0.61), indicating an absent up to small effect, with considerable variation. For studies assessing religious struggle ($N = 22$) the mean effect size was 0.30 (median 0.23; SD 0.36; range −0.04 to 1.50), corresponding to a small to moderate effect. The distribution of the effect-sizes is shown in Table 6.

Comparison of the strength of effect sizes (**Supplementary Table D**) across relevant subgroups yielded significant differences between the types of samples and types of analytical modeling. The effect-sizes were smallest in the samples of patients with a physical condition (−0.10) and largest in samples of persons with psychiatric symptoms (−0.37) ($F = 3.1$, df 5 / 124; $P = .012$). With respect to statistical modeling, linear regression and advanced longitudinal modeling yielded lower effect-sizes than logistic regression and other models ($F = 2.9$; df 3 / 126; $P = .035$). In addition, the analyses were repeated after weighting for number (N) of participants per study, redressed to the original number of studies included (138), with weightfactor: ln(N) * 138 / 833. The same main findings emerged, with significance only for the comparison of types of sample ($F = 2.5$, df 5 / 124; $P = .037$; further results on request).

4. Discussion

Based on identification of 152 prospective studies, the current review found that about half of these reported a significant association between measures of R/S and a better course of depressive symptoms/depression over time. Forty percent of studies did not find a significant effect, and about 10% reported associations with more depression over time. The estimated strength of these associations was modest ($d = -0.18$). In addition, among studies that operationalized R/S as religious struggle (analyzed separately), 59% found that it predicted more depression over time ($d = +0.30$).

Religious struggle is likely to be closely related to aspects of personality such as neuroticism that undermine psychological well-being and contribute to vulnerability to depression (Ano and Pargament, 2013; Wilt et al., 2017) or may be a manifestation of depression itself (Koenig, 2018). There is also variation within the category of more general measures of R/S involvement. Religious attendance was the most commonly used measure and was most likely to predict a decline in depression over time. The findings for salience of religion are similar, but they are weaker for positive forms of religious coping and prayer, which are likely to be mobilized as a 'last resort' in times of distress as depression worsens. Measures of spirituality also frequently predict a decline in depression over time, although as noted previously, these measures (particularly the FACIT-Sp and SWBS) are contaminated by items assessing positive emotions (Koenig, 2008; Garssen et al., 2016). Therefore, the inverse associations identified between measures of spirituality assessed in this way and depression may be tautological in nature.

In 1983, Bergin published a critical review of religiosity and mental health. He concluded that results were often ambiguous. He suggested that this could be addressed by treating religiosity as a multidimensional construct which has a mixture of both positive and negative effects. The current review confirms this early speculation, although measures of R/S were clearly more often associated with a better course of depression than with a worse course. However, several studies identified here used composite measures on R/S and reported similar results as other studies. These measures may have been developed to avoid multiple testing, increase sensitivity, or improve the economy of presentation of results. A disadvantage is that using composite measures with heterogeneous indicators of R/S complicates the comparison of studies. A different, but similarly complex issue, is that a few studies evaluated a range of R/S measures in predicting depression all included in a single multivariate model. Here, it is not certain whether problems of multicollinearity have been sufficiently addressed, and that spurious associations or artefacts may have been reported.

An important finding of the current review is that the *type of sample* can affect the pattern of results. Studies of younger people and those with medical illness are less likely to report significant results. In contrast, studies in persons identified with psychiatric illness and in population-based samples are more likely to find that R/S predicts a decline in depression over time. The estimated effect-size in samples of persons with psychiatric symptoms was somewhat more substantial ($d = -0.37$), although this was still small to moderate. The difference in findings between those with physical health problems and those with psychiatric conditions was not expected. Previous meta-analyses and reviews have not suggested the possibility of differences between samples. One may speculate that among those with medical illness, particularly chronic medical illness, the physical condition may prompt a turning to religion and the persistence of depressive symptoms (due to failure of the medical illness to improve), therefore disguising any beneficial effects that R/S may have. Similarly, among those with psychiatric symptoms, and especially depressive symptoms, the levels of symptoms may fluctuate considerably (compared to chronic physical health problems) and be more responsive to R/S involvement. Hence, in samples of psychiatric patients, there may be more variance in depressive symptoms to explain thereby resulting in significant findings.

Although not statistically significant, studies originating in the US and Canada were more likely to report significant associations with less depression over time compared to studies in Europe or East Asia. As more studies become available from other regions in the world outside of North-America, particularly the Middle East, this difference might become more pronounced.

With respect to design, studies that adjusted for demographic variables or other confounders were more likely to report significant results, which was unexpected and contrasts with concerns about the research voiced by critics (Sloan et al., 1999; Sloan, 2006). Studies identified in this review with more adjustments for confounders often had a more rigorous design (e.g., larger samples) than studies without adjustment. Furthermore, studies utilizing linear regression analyses or basic statistical approaches tended to be less likely to report statistically significant results. Not clear is whether the level of adjustment for depressive symptoms at baseline was more stringent in these types of analyses compared to those that applied logistic regression techniques or more advanced longitudinal modelling.

4.1. Limitations and recommendations

Although the current review was systematic, exhaustive, and identified average effect sizes, it did not conduct a meta-analysis, leaving this as a next step in the future, possibly with even a larger number of studies (although the same issues that prevented us from doing this may hamper such efforts). The provisional, exploratory 'vote-counting' method of analyzing the distribution of results across subgroups of interest can be considered a concise version of a narrative discussion of very heterogeneous studies. The results of the present review indicate that several methodological aspects of studies must be taken into consideration in future reviews, such as duration of follow-up, number of follow-ups, and the analytical approach. Furthermore, consideration of sample types will almost certainly need to be taken into account.

A significant limitation of the current review is that the majority of researchers assumed a linear association between R/S and depression over time. However, some studies have identified a U-shaped association between these constructs (e.g. Schnittker, 2001; Schettino et al., 2011). For example, both high and low levels of R/S may predict a better course of depression, whereas intermediate levels may predict the opposite. One recommendation for future research is to anticipate non-linear associations in the analytical approach.

Some other limitations pertain to the language restriction in the literature search, as there may exist studies published in other languages than English, as well as to the restriction that only published research articles were included. Although this restriction to published research was considered a quality criterion, other reports may exist in the so called 'gray literature' (theses, book-chapters, etc.).

Future studies will need to employ a design which permits cross-lagged analyses in order to get a sense of causal direction (VanderWeele et al., 2016; Maselko et al., 2012). For example, Maselko et al. (2012), although partly using retrospective data, argued that those who become depressed early in life tend to be more likely to stop attending religious services. Indeed, some of the fourteen studies with a cross-lagged design in the current review suggest bidirectional causal effects (see **Table 2**, marked with '‡', '#').

Another recommendation is to include a qualitative approach (Dein, 2006), or mixed-methods design, to determine how different aspects of R/S matter for individuals in the study. For example, Lieberman and Winzelberg (2009) utilized software for linguistic analysis to derive a measure of percentage of religious expressions in written messages in online support groups of patients with breast cancer. Another example is to apply content analysis of in-depth interviews and essays to obtain codes on specific aspects of R/S coping, as has been described by Kremer and Ironson (2014) in a longitudinal study of people with HIV.

5. Conclusion

Over the past several decades a substantial body of research with origins in very different disciplines, ranging from the social sciences to psychiatry to clinical epidemiology, has emerged on the association between R/S and the course of depression over time. Several aspects of R/S such as church-attendance and salience of religion have now been shown to have a modest but consistent ability to predict lower levels of depression over time. Whether aspects of R/S reflect a characteristic that is inherent to mental health, or whether R/S represents an independent predictor of depression outcome remains uncertain. Carefully done life-time studies may be required to elucidate the underlying dynamics of this relationship. Nevertheless, the current evidence suggests that clinicians and therapists should pay careful attention to the R/S of their clients, evaluating its impact on the individual with depression or at risk for it. Positive aspects of R/S may at times come to the fore, while at other times religious struggle may be found to be present, and sometimes both. From an epidemiological viewpoint, the present review suggests that the major patterns of association may depend on the types of R/S measures, the types of samples, and the methodological approach.

CRediT authorship contribution statement

Arjan W. Braam: Investigation, Formal analysis, Writing - review & editing, Validation. **Harold G. Koenig:** Investigation, Writing - review & editing, Validation.

Acknowledgements

The authors would like to thank the library staff of Altrecht Science Library, Mrs. Fieke Bannink and Mrs. Marion Scheepers for their accurate assistance with collecting the first series of research papers, as well as Mrs. Rita Vos for her assistance with documentation of the papers. In addition, it was a privilege to have been assisted in collecting the papers of the first update by to the colleagues of 'GGZKenniscentrum' Mrs. Trijntje Lucassen, Mrs. Ingrid Pasker, and Mrs. Laura Linger.

Role of funding source

Not applicable

Ethics committee approval

Not applicable

Conflict of interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jad.2019.06.063.

References

Ano, G.G., Pargament, K.I., 2013. Predictors of spiritual struggles: an exploratory study. Ment. Health Relig. Cult. 16, 419–434. https://doi.org/10.1080/13674676.2012.680434.

Bergin, A.E., 1983. Religiosity and mental health: a critical reevaluation and meta-analysis. Prof. Psychol. Res. Pract. 14, 170–184.

Blazer, D., 2012. Religion/spirituality and depression: what can we learn from empirical studies? Am. J. Psychiatry 169, 10–12. https://doi.org/10.1176/appi.ajp.2012.169.8.a10.

Bonelli, R., Dew, R.E., Koenig, H.G., Rosmarin, D.H., Vasegh, S., 2012. Religious and spiritual factors in depression: review and integration of the research. Depress. Res. Treat. 2012. https://doi.org/10.1155/2012/962860.

Crossroads, 2007. A Monthly Publication of Research on Spirituality and Health. Duke University Medical Center, Center for Spirituality, Theology and Health, Durham, North Carolina. https://spiritualityandhealth.duke.edu/index.php/publications/crossroads.

Dein, S., 2006. Religion, spirituality and depression: implications for research and treatment. Prim. Care Community Psychiatry 11, 67–72. https://doi.org/10.1185/135525706X121110.

Garssen, B., Visser, A., de Jager Meezenbroek, E., 2016. Examining whether spirituality predicts subjective well-being: how to avoid tautology. Psychol. Relig. Spiritual. 8, 141–148. https://doi.org/10.1037/rel0000025.

King, M., Marston, L., McManus, S., Brugha, T., Meltzer, H., Bebbington, P., 2013. Religion, spirituality and mental health: results from a national study of English households. Br. J. Psychiatry 202, 68–73. https://doi.org/10.1192/bjp.bp.112.112003.

Koenig, H.G., McCullough, M.E., Larson, D.B., 2001. Handbook of Religion and Health. Oxford University Press, Oxford, UK / New York, NY, pp. 118–135 Chapter 7527–530.

Koenig, H.G., 2008. Concerns about measuring "spirituality" in research. J. Nerv. Ment. Dis. 196, 349–355. https://doi.org/10.1097/NMD.0b013e31816ff796.

Koenig, H.G., King, D.E., Benner Carson, V., 2012. Handbook of Religion and Health. Oxford University Press, Oxford, UK / New York, NY, pp. 145–173 Chapter 7609.

Koenig, H.G., 2018. Religion and Mental Health: Research and Clinical Applications. Academic Press (Elsevier), San Diego, CA, pp. 177–201.

Kremer, H., Ironson, G., 2014. Longitudinal spiritual coping with trauma in people with

HIV: implications for health care. AIDS Patient Care STDS 28, 144–154. https://doi.org/10.1089/apc.2013.0280.
Kuijpers, P., 2016. Meta-analyses in Mental Health Research: A Practical Guide. Vrije Universiteit, Amsterdam, The Netherlands.
Li, S., Okereke, O.I., Chang, S.C., Kawachi, I., VanderWeele, T.J., 2016. Religious service attendance and lower depression among women–a prospective cohort study. Ann. Behav. Med. 50, 876–884. https://doi.org/10.1007/s12160-016-9813-9.
Lieberman, M.A., Winzelberg, A., 2009. The relationship between religious expression and outcomes in online support groups: a partial replication. Comput. Human. Behav. 25, 690–694.
Maselko, J., Hayward, R.D., Hanlon, A., Buka, S., Meador, K., 2012. Religious service attendance and major depression: a case of reverse causality? Am. J. Epidemiol. 175, 576–583. https://doi.org/10.1093/aje/kwr349.
National Institute of Health. Quality assessment tool for observational cohort and cross-sectional studies, 2018. https://www.nhlbi.nih.gov/health-topics/study-quality-assessmenttools (accessed August 6, 2018).
Pargament, K.I., Smith, B.W., Koenig, H.G., Perez, L., 1998. Patterns of positive and negative religious coping with major life stressors. J. Sci. Study Relig. 37, 710–724.
Sahgal, N., 2018. 10 key findings about religion in Western Europe. Pew Research Center. http://www.pewresearch.org/fact-tank/2018/05/29/10-key-findings-about-religion-in-western-europe/ (accessed August 12, 2018).
Shea, B.J., Grimshaw, J.M., Wells, G.A., Boers, M., Andersson, N., Hamel, C., Porter, A.C., Tugwell, P., Moher, D., Bouter, L.M., 2007. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med. Res. Methodol. 7, 10. https://doi.org/10.1186/1471-2288-7-10. 2007 Feb 15.
Sloan, R.P., Bagiella, E., Powell, T., 1999. Religion, spirituality, and medicine. Lancet 353, 664–667.
Sloan, R.P., 2006. Blind Faith: The Unholy Alliance of Religion and Medicine. Macmillan, New York, NY.
Schettino, J.R., Olmos, N.T., Myers, H.F., Joseph, N.T., Poland, R.E., Lesser, I.M., 2011. Religiosity and treatment response to antidepressant medication: a prospective multi-site clinical trial. Ment. Health Relig. Cult. 14, 805–818. https://doi.org/10.1080/13674676.2010.527931.
Schnittker, J., 2001. When is faith enough? The effects of religious involvement on depression. J. Sci. Study Relig. 40, 393–411. https://doi.org/10.1111/0021-8294.00065.
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. Psychol. Bull. 129, 614–636.
VanderWeele, T.J., Jackson, J.W., Li, S., 2016. Causal inference and longitudinal data: a case study of religion and mental health. Soc. Psychiatry Psychiatr. Epidemiol. 51, 1457–1466. https://doi.org/10.1007/s00127-016-1281-9.
Wilt, J.A., Grubbs, J.B., Pargament, K.I., Exline, J.J., 2017. Religious and spiritual struggles, past and present: relations to the big five and well-being. Int. J. Psychol. Relig. 27, 51–64.