

# **Modellbasierte Meta-Analyse veranschaulicht am Zusammenhang von Religiosität und Lebenszufriedenheit**

Von der Fakultät für Wirtschafts- und Sozialwissenschaften der Universität Stuttgart zur Erlangung der Würde eines Doktors der Wirtschafts- und Sozialwissenschaften (Dr. rer. pol.) genehmigte Abhandlung

Vorgelegt von  
Marius Reinhold Wuketich  
aus Heilbronn am Neckar

Hauptberichter: Prof. Dr. Dieter Urban  
Mitberichter: Prof. Dr. Jochen Mayerl

Tag der mündlichen Prüfung: 30.01.2020

Institut für Sozialwissenschaften der Universität Stuttgart

2020

"All models are wrong but some are useful".  
(George Box)

## **Danksagung**

Die Anfertigung einer Dissertation ist gewiss das Werk eines Einzelnen, aber ohne die Unterstützung durch andere Menschen kaum denkbar. Zunächst möchte ich meinem „Doktorvater“ Prof. Dr. Dieter Urban danken, der mir die Gelegenheit dazu gab, die vorgelegte Dissertationschrift unter seiner Obhut anzufertigen und mir dabei die Freiheit ließ, meinen eigenen Forschungsneigungen nachzugehen. In besonderer Weise möchte ich mich bei meinen ehemaligen Arbeitskollegen und Mitdoktoranden Thomas Krause und Andreas Wahl bedanken, da wir uns solange gegenseitig unterstützt, miteinander gelitten und uns gegenseitig immer wieder aufs Neue für die Arbeit an unseren Projekten motiviert haben. Ich danke den ehemaligen wissenschaftlichen Hilfskräften Adina Simonovic und Alexander Geisler für deren Unterstützung in der Frühphase dieser Arbeit. Darüber hinaus danke ich Mike Cheung (National University of Singapore), Suzanne Jak (University of Amsterdam) und Wolfgang Viechtbauer (Maastricht University) für die hilfreichen methodischen Hinweise zu meiner Arbeit. Mein besonderer Dank gilt Christine Kicherer, Eva Sonnenmoser, Andreas Wahl, Andrea Wöhr und Heike Wuketich für die umsichtige Durchsicht und Korrektur des Manuskripts.

Danken möchte ich meiner Familie und meinen Freunden, die mich zu dem gemacht haben, der ich heute bin und die stets viel Geduld mit mir haben. Abschließend danke ich meiner Lebensgefährtin Eva Sonnenmoser, ohne deren Unterstützung sowie Verständnis das vorliegende Werk nicht entstanden wäre. Die unzähligen Diskussionen und Gespräche über das Thema meiner Dissertation sowie die vielen kritischen und unterstützenden Anmerkungen mit vielen sehr unterschiedlichen Menschen haben zum Gelingen des Werkes maßgeblich beigetragen. Diese Gespräche haben meine wissenschaftliche Neugier stets aufs Neue befeuert.

Ich widme diese Arbeit meinen Eltern Rudolf (22.11.1943 – 01.02.2017) und Margarete Elfriede Wuketich (12.11.1956 – 26.12.2018), die beide während der Zeit, in der ich diese Arbeit angefertigt habe, verstorben sind.

Stuttgart, im Frühjahr 2019

Marius Reinhold Wuketich

# Inhaltsverzeichnis

<b>Abbildungsverzeichnis</b> .....	<b>5</b>
<b>Tabellenverzeichnis</b> .....	<b>7</b>
<b>Kurzzusammenfassung</b> .....	<b>9</b>
<b>Summary</b> .....	<b>11</b>
<b>1 Einleitende Worte</b> .....	<b>13</b>
1.1 Relevanzbegründung .....	15
1.2 Ziele dieser Arbeit und forschungsleitende Fragen.....	17
1.3 Aufbau der Arbeit.....	18
<b>2 Einführende Anmerkungen zur Meta-Analyse</b> .....	<b>21</b>
2.1 Leistung, Ablauf und Einordnung .....	21
2.2 Dokumentation von Meta-Analysen .....	28
2.3 Exkurs: Kritische Würdigung des Verfahrens .....	31
<b>3 Forschungsstand und konzeptionelle Überlegungen</b> .....	<b>37</b>
3.1 Religiosität und Spiritualität.....	38
3.2 Lebenszufriedenheit .....	43
3.3 Zusammenhangsbegründung.....	48
3.4 Ableitung des Erklärungsmodells .....	50
<b>4 Forschungsdesign</b> .....	<b>61</b>
4.1 Methodischer Review der meta-analytischen Praxis .....	61
4.1.1 Literatursuche.....	65
4.1.2 Screening.....	71
4.1.3 Anzahl der Studien .....	73
4.1.4 Kodierverfahren .....	74
4.1.5 Statistische Umsetzung .....	76
4.1.6 Zusammenfassung der methodologischen Praxis .....	81
4.2 Selektionskriterien: Inklusion und Exklusion .....	82
4.3 Literatursuchstrategie .....	84
4.4 Screeningverfahren.....	87
4.5 Kodierverfahren .....	88
<b>5 Meta-analytische (statistische) Methodik</b> .....	<b>91</b>
5.1 Grundlegende statistische Konzepte und Verfahren .....	91
5.1.1 Grundlegende Modellierungsformen .....	91
5.1.2 Zentrale Aspekte .....	96
5.2 SEM-basierte Meta-Analyse .....	100
5.3 Meta-Analytic-Structural-Equation-Modelling-Ansätze (MASEM) .....	101
5.3.1 Univariater r/z-Ansatz .....	103
5.3.2 GLS-Ansatz.....	105
5.3.3 Systematisierung der Probleme von univariatem und GLS-Ansatz.....	107
5.4 Two-Stage-Structural-Equation-Modelling-Ansatz (TSSEM) .....	110
5.4.1 Erweiterungen und Spezifika des TSSEM.....	116
5.4.2 Desiderata Two-Stage-Structural-Equation-Modelling (TSSEM).....	124
5.5 Publikationsbias und Sensitivitätsanalyse.....	128
5.5.1 Publikationsbias .....	128
5.5.2 Sensitivitätsanalyse .....	131

<b>6</b>	<b>Literatursuche und Vor-Screening</b> .....	<b>135</b>
6.1	Erhebung der Studien .....	135
6.2	Vor-Screening .....	136
<b>7</b>	<b>Kodierung</b> .....	<b>139</b>
7.1	Vorgehen bei der Kodierung .....	139
7.2	Kodierungsergebnisse .....	141
<b>8</b>	<b>Erwartungen</b> .....	<b>153</b>
<b>9</b>	<b>Durchführung der Meta-Analyse und Ergebnisdarstellung</b> .....	<b>157</b>
9.1	Programmtechnische Umsetzung .....	157
9.2	Übersicht über die statistischen Modellierungsschritte.....	159
9.3	Analyseergebnisse Mittelwert-TSSEM-Ansatz .....	163
9.4	Untersuchung der Moderatorvariablen.....	170
9.4.1	One-Stage-MASEM (OSMASEM).....	171
9.4.2	TSSEM mit Subgruppenanalyse .....	177
9.5	Verfahren zur Berücksichtigung abhängiger Effektstärken .....	189
9.5.1	Wilson-Ansatz und robuster Ansatz.....	189
9.5.2	Vergleich der Ansätze .....	196
9.6	Überprüfung des Publikationsbias .....	199
<b>10</b>	<b>Schlussfolgerung</b> .....	<b>209</b>
10.1	Systematisierung der Ergebnisse.....	209
10.2	Grenzen der vorliegenden modellbasierten Meta-Analyse .....	213
10.3	Ausblick .....	216
<b>Anhang A: Kodierschema</b> .....		<b>220</b>
<b>Anhang B: Studienübersicht</b> .....		<b>232</b>
<b>Anhang C: Studien in der Meta-Analyse</b> .....		<b>235</b>
<b>Anhang D: R-Session-Info</b> .....		<b>244</b>
<b>Anhang E: Forest-Plots</b> .....		<b>245</b>
<b>Literaturverzeichnis</b> .....		<b>251</b>

## Abbildungsverzeichnis

Abbildung 1: Fünf Phasen einer Meta-Analyse .....	23
Abbildung 2: Detaillierte Übersicht meta-analytischer APD-Verfahren .....	24
Abbildung 3: Fortgeschrittene Verfahren der Meta-Analyse .....	25
Abbildung 4: Definition von Religiosität und Spiritualität .....	42
Abbildung 5: Konzeptionelle Einordnung subjektives Wohlbefinden .....	43
Abbildung 6: Definition subjektives Wohlbefinden .....	47
Abbildung 7: Erklärungsmodell mit Mechanismusclustern .....	56
Abbildung 8: Sozialwissenschaftliches Erklärungsmodell .....	57
Abbildung 9: Vollständiges Erklärungsmodell .....	59
Abbildung 10: Anzahl der Zitationsdatenbanken .....	68
Abbildung 11: MASEM-Verfahren .....	102
Abbildung 12: Schematische Darstellung TSSEM .....	110
Abbildung 13: Hierarchische Struktur des Three-Level-Designs .....	123
Abbildung 14: Ablauf der Stichprobenkodierung .....	140
Abbildung 15: Zusammenfassung von Screening und Kodierung .....	144
Abbildung 16: Anzahl der Effektstärken .....	146
Abbildung 17: Stichprobenübersicht nach Art der Studie .....	147
Abbildung 18: Stichprobenübersicht nach Publikationsjahr .....	147
Abbildung 19: Anzahl von R/S Messinstrumenten .....	148
Abbildung 20: Anzahl von Lebenszufriedenheitsmessinstrumenten .....	148
Abbildung 21: Nicht-repräsentative Stichproben nach Stichprobenart .....	149
Abbildung 22: Gesamtmodell .....	163
Abbildung 23: Schätzergebnis des Gesamtmodells mit $k_2$ .....	167
Abbildung 24: Schätzergebnis des Gesamtmodells ohne $k_2$ .....	168
Abbildung 25: OSMASEM Untersuchungsmodell .....	171
Abbildung 26: Reduziertes OSMASEM Untersuchungsmodell .....	175
Abbildung 27: Ausgangsmodell für die Moderatoranalyse .....	177
Abbildung 28: Pfadmodell reduziertes Untersuchungsmodell (HDI-Datensatz) .....	179
Abbildung 29: Low- und High-HDI-Gruppe .....	181
Abbildung 30: Pfadmodell reduziertes Untersuchungsmodell (WR-Datensatz) .....	184
Abbildung 31: Zweite Phase Low- und High-WR-Gruppe .....	186
Abbildung 32: Zweite Phase Schätzergebnis des Wilson-Ansatzes (mit $k_2$ ) .....	194
Abbildung 33: Zweite Phase Schätzergebnis des robusten Ansatzes (mit $k_2$ ) .....	194
Abbildung 34: Zweite Phase Schätzergebnis des Wilson-Ansatzes ohne $k_2$ .....	195
Abbildung 35: Zweite Phase Schätzergebnis des robusten Ansatzes (ohne $k_2$ ) .....	195
Abbildung 36: Funnel-Plot Zelle 1 .....	200
Abbildung 37: Funnel-Plot Zelle 2 .....	200
Abbildung 38: Funnel-Plot Zelle 3 .....	201
Abbildung 39: Funnel-Plot Zelle 4 .....	201
Abbildung 40: Funnel-Plot Zelle 5 .....	202
Abbildung 41: Funnel-Plot Zelle 6 .....	202
Abbildung 42: Funnel-Plot Zelle 7 .....	203
Abbildung 43: Funnel-Plot Zelle 8 .....	204
Abbildung 44: Funnel-Plot Zelle 9 .....	204
Abbildung 45: Funnel-Plot Zelle 10 .....	205
Abbildung 46: Forest-Plot Zelle 1 .....	245
Abbildung 47: Forest-Plot Zelle 2 .....	245
Abbildung 48: Forest-Plot Zelle 3 .....	246
Abbildung 49: Forest-Plot Zelle 4 .....	247

Abbildung 50: Forest-Plot Zelle 5.....	248
Abbildung 51: Forest-Plot Zelle 6.....	248
Abbildung 52: Forest-Plot Zelle 7.....	249
Abbildung 53: Forest-Plot Zelle 8.....	249
Abbildung 54: Forest-Plot Zelle 9.....	250
Abbildung 55: Forest-Plot Zelle 10.....	250

## Tabellenverzeichnis

Tabelle 1: Meta-Analysis Reporting Standards (MARS) .....	29
Tabelle 2: Übersicht Erklärungsmechanismen.....	54
Tabelle 3: Basisinformationen der gefundenen Meta-Studien.....	63
Tabelle 4: Verwendete Zitationsdatenbanken .....	67
Tabelle 5: Inhaltliche Suchstrings .....	70
Tabelle 6: Auffälligkeiten Inklusions- und Exklusionskriterien .....	72
Tabelle 7: Anzahl Primärstudien sowie Effekt- und Samplegröße .....	73
Tabelle 8: Kodierverfahren und Reliabilitätsüberprüfung .....	75
Tabelle 9: Ausgewählte methodologisch-statistische Aspekte .....	78
Tabelle 10: Moderatoruntersuchung .....	80
Tabelle 11: Selektionskriterien.....	84
Tabelle 12: Suchstrategie .....	84
Tabelle 13: Bekannte Literaturreviews und Meta-Studien .....	85
Tabelle 14: Literaturdatenbanken nach Disziplin .....	86
Tabelle 15: Suchstrings .....	87
Tabelle 16: Bereiche des Kodierschema .....	89
Tabelle 17: Schritte einer MASEM nach Viswesvaran/Ones 1995 .....	104
Tabelle 18: Schwächen der herkömmlichen MASEM-Verfahren .....	108
Tabelle 19: Ansätze zum Umgang mit Heterogenität .....	121
Tabelle 20: Umgang mit statistischer Abhängigkeit der Korrelationskoeffizienten.....	124
Tabelle 21: Meta-Studien und Literaturreviews.....	135
Tabelle 22: Datenbanksuche .....	135
Tabelle 23: Literatursuche und Vor-Screening .....	137
Tabelle 24: Tautologische Messinstrumente.....	142
Tabelle 25: Empirisch gefundene Mediatoren .....	145
Tabelle 26: Stichprobenübersicht nach Ländern.....	146
Tabelle 27: Berücksichtigte Messinstrumente .....	150
Tabelle 28: Erwartungen .....	154
Tabelle 29: Korrelationsmatrix .....	159
Tabelle 30: Ausgeschlossene Studien (K-IDs).....	160
Tabelle 31: Stichproben- und Fallzahl .....	161
Tabelle 32: Anzahl der Effektstärken.....	161
Tabelle 33: Deskriptive Übersicht über die Effektstärken .....	162
Tabelle 34: (Random-Effects-) Ergebnisse erste Phase Mittelwert-TSSEM-Ansatz.....	164
Tabelle 35: Fallzahlen nach Beziehungen.....	165
Tabelle 36: Zweite Phase Mittelwert-TSSEM-Ansatz (Gesamtmodell mit k2 und ohne k2) .....	166
Tabelle 37: Vergleich TSSEM und OSMASEM (HDI-Datensatz) .....	172
Tabelle 38: Vergleich TSSEM und OSMASEM (WR-Datensatz) .....	174
Tabelle 39: Reduziertes OSMASEM Untersuchungsmodell .....	176
Tabelle 40: Mediansplit der Makrovariablen für die Moderatoranalyse.....	177
Tabelle 41: Erste Phase reduziertes Untersuchungsmodell (HDI-Datensatz).....	178
Tabelle 42: Zweite Phase reduziertes Untersuchungsmodell (HDI-Datensatz).....	178
Tabelle 43: Erste Phase für beide Gruppen der HDI-Moderatoranalyse ( $I^2$ -Werte) .....	179
Tabelle 44: Erste Phase der HDI-Moderatoranalyse (getrennt) .....	180
Tabelle 45: Zweite Phase in beiden Gruppen (HDI-Datensatz).....	180
Tabelle 46: Vergleich freies Modell mit Constraint-Modell (HDI-Datensatz).....	182
Tabelle 47: Erste Phase reduziertes Untersuchungsmodell (WR-Datensatz) .....	183
Tabelle 48: Zweite Phase reduziertes Untersuchungsmodell (WR-Datensatz) .....	183
Tabelle 49: Erste Phase der WR-Moderatoranalyse ( $I^2$ -Werte) .....	184

Tabelle 50: Erste Phase der WR-Moderatoranalyse (getrennt).....	185
Tabelle 51: Schätzergebnis der zweiten Phase in beiden Gruppen (WR-Datensatz).....	186
Tabelle 52: Vergleich freies Modell und Constraint Modell (WR-Moderator).....	187
Tabelle 53: (Random-Effects-) Ergebnis der ersten Phase des Wilson-Ansatzes.....	190
Tabelle 54: Erste Phase robuster Ansatz .....	190
Tabelle 55: Vergleich erste Phase Wilson-Ansatz und Mittelwert-TSSEM-Ansatz.....	191
Tabelle 56: Gesamtmodell mit k2 und ohne k2 (Wilson-Ansatz).....	192
Tabelle 57: Gesamtmodell mit k2 und ohne k2 (robuster Ansatz) .....	193
Tabelle 58: Vergleich der Ansätze (Zweite Phase mit k2).....	196
Tabelle 59: Bivariat kumulierte Korrelationskoeffizienten .....	199
Tabelle 60: Rosenbergs fail-safe Ns und Begg/Mazumdar Rangkorrelation.....	205
Tabelle 61: Egger-Testergebnisse .....	206
Tabelle 62: Erwartungen und Ergebnis .....	210

## Kurzzusammenfassung

Die modellbasierte Meta-Analyse ist eine bislang nur wenig bekannte multivariate Form der Meta-Analyse. Sie ermöglicht die Überprüfung multivariater Erklärungsmodelle mit manifesten und latenten Variablen sowie mehreren abhängigen und unabhängigen Variablen. Dieses Verfahren geht somit über die Leistungsfähigkeit vieler vorliegender meta-analytischer Verfahren hinaus und ermöglicht etwa die Spezifizierung von Erklärungsmechanismen in Form von Mediatoren unter Verwendung von Strukturgleichungsmodellierung. Dies wird Meta-Analytic-Structural-Equation-Modelling genannt (MASEM). Zur Umsetzung hat sich bislang keiner der vorgeschlagenen MASEM-Ansätze (univariater  $r/z$ -, GLS- und TSSEM-Ansatz) durchgesetzt. Der bislang vielversprechendste Ansatz, das TSSEM-Verfahren, ist bislang nur wenig erprobt (Cheung 2015; Jak 2015; Jak/Cheung 2018).

In der vorliegenden Studie wird anhand eines empirischen Beispiels, den in vielen wissenschaftlichen Disziplinen relevanten Zusammenhang von Religiosität und Lebenszufriedenheit, der Two-Stage-Structural-Equation-Modelling-Ansatz (TSSEM) angewandt. Dabei handelt es sich um einen gut untersuchten Zusammenhang zweier metrisch gemessener sozialer Konstrukte und damit um einen häufig auftretenden sozialwissenschaftlichen Anwendungsfall. Die vorliegenden empirischen Studien zeigen einen schwach positiven Zusammenhang. Allerdings gibt es eine erhebliche Befundstreuung und auch negative Zusammenhänge. Eine Reihe von unterschiedlichen Erklärungsmechanismen wird für diesen Effekt diskutiert.

Eine aktuelle meta-analytische Untersuchung des Zusammenhangs liegt derzeit nicht vor und eine modellbasierte Meta-Analyse, mit Einbezug von Erklärungsmechanismen, ist zu dieser Forschungsfrage noch nicht angefertigt worden. Für die Untersuchung wird ein entsprechendes Erklärungsmodell empirisch-analytisch aus der Literatur abgeleitet. Diese explorative Vorgehensweise steht im Kontrast zur bisherigen Anwendung modellbasierter Verfahren, die in der Regel auf etablierte Erklärungsmodelle zurückgreifen. Somit evaluiert diese Arbeit, ob der TSSEM-Ansatz zur explorativen modellbasierten Meta-Analyse genutzt werden kann. Das Untersuchungsmodell besteht aus den beiden genannten Variablen sowie den drei am häufigsten in der vorliegenden Literatur untersuchten Erklärungsmechanismen: soziale Unterstützung, Optimismus und Selbstwert. Eingebettet ist das Erklärungsmodell in den kulturellen Kontext. Dies wird unter Zuhilfenahme zweier ausgewählter Kultur-Proxy-Moderatorvariablen kontrolliert. Weitere methodische Moderatoren werden ebenfalls überprüft.

Neben der inhaltlichen Untersuchung des Erklärungsmodells werden die methodischen Probleme des TSSEM-Ansatzes systematisiert und Ansätze für weitergehende Forschung präsentiert. Zwei dieser methodischen Fragen werden empirisch beleuchtet. Zunächst werden drei Verfahren zur Berücksichtigung abhängiger Effektstärken (Mittelwertbildung, Wilson-Ansatz und robuster Ansatz) und deren Auswirkung auf das Schätzergebnis miteinander verglichen. Darüber hinaus werden zwei Verfahren zur Untersuchung von Moderatorvariablen in der modellbasierten Meta-Analyse verglichen. Zum einen wird im herkömmlichen TSSEM-Ansatz eine Subgruppenanalyse durchgeführt und zum anderen werden die Möglichkeiten des neuen One-Stage-MASEM (OSMASEM) zur Untersuchung von Moderatorvariablen angewandt.

Die Anwendung des TSSEM-Ansatzes zeigt, dass dieser auch explorativ meta-analytisch genutzt werden kann. Dabei kommt es jedoch zu spezifischen Problemen. Die ungleich besetzten Zellen der Korrelationsmatrix führen zu Konvergenzproblemen, insbesondere bei der Moderatoruntersuchung. Die Ergebnisse der modellbasierten Meta-Studie können den positiven Effekt von Religiosität auf Lebenszufriedenheit nicht falsifizieren. Die drei überprüften Erklärungsmechanismen (soziale Unterstützung, Optimismus und Selbstwert) wirken als partielle Mediatoren. Durch Berücksichtigung der Mediatorvariablen wird der direkte Effekt zwischen Religiosität und Lebenszufriedenheit schwächer. Es ist davon auszugehen, dass sich der Effekt beim Einbezug weiterer, in der Literatur diskutierter, Erklärungsmechanismen weiter verringert. Dies kann als Entzauberung oder Verweltlichung der individuellen Wirkungen von Religiosität gedeutet werden, da diese auf profane psychosoziale Erklärungsmechanismen zurückgeführt werden. Der Vergleich der Verfahren zur Berücksichtigung abhängiger Effektstärken zeigt, dass die anspruchsvolleren Verfahren (Wilson-Ansatz und robuster Ansatz) die Präzision der Schätzung bei einer geringen Anzahl von Effektstärken erhöhen können. Liegt jedoch ausreichend empirisches Material (hier: mehr als 10 Effektstärken für eine Beziehung) und keine auf die Messinstrumente bezogene Forschungsfrage vor, kann die einfache Mittelwertbildung verwendet werden. Die Überprüfung der Moderatorvariablen belegt die Kulturbedingtheit des Erklärungsmodells. Die methodischen Moderatoren können im Gegensatz dazu kaum weitere Erklärungsleistung erbringen. In Bezug auf die zwei Verfahren zur Moderatoranalyse zeigt sich ein deutlicher Vorteil des OSMASEM-Ansatzes gegenüber des Subgruppenvergleichs. Mit OSMASEM sind präzisere Aussagen über die Wirkung der postulierten Moderatorvariablen möglich. Abschließend werden die systematisierten Desiderata des TSSEM-Verfahrens mit den empirischen Ergebnissen der Anwendung verknüpft, um Ansatzpunkte zur Weiterentwicklung der Methodik darzulegen.

## Summary

Model-based meta-analysis is a multivariate form of meta-analysis that has so far not been widely used. It enables the examination of multivariate models with manifest and latent variables as well as several dependent and independent variables. This method goes beyond the capabilities of existing meta-analytical methods and enables, for example, the specification of explanatory mechanisms in the form of mediators by using structural equation modelling. This is called Meta-Analytic-Structural-Equation-Modelling (MASEM). So far, none of the existing MASEM approaches (univariate  $r/z$ -, GLS and the TSSEM approach) has been implemented. The most promising approach so far, the TSSEM approach, has been rarely tested (Cheung 2015; Jak 2015; Jak/Cheung 2018).

In this study, the TSSEM procedure is being applied on an empirical example, the association between religiousness and life satisfaction, which is relevant in many scientific disciplines. This is a well-studied correlation of two metrically measured social constructs and therefore a commonly occurring social science application. The empirical studies at hand show a weak positive correlation. However, there is a considerable variation in findings and there are also negative correlations. A number of different explanatory mechanisms are discussed for this correlation. There is no up-to-date meta-analytical investigation of the relationship and a model-based meta-analysis with consideration of explanatory mechanisms has not yet been carried out. An empirical-analytical explanatory model will be derived from the literature. This explorative approach contrasts the previous application of model-based approaches, which usually rely on existing established models. Thus, this work evaluates whether the TSSEM approach can be used for explorative model-based meta-analysis. The model consists of the two variables mentioned above and the three most common explanatory mechanisms studied in the literature: social support, optimism and self-esteem. The explanatory model is embedded in the cultural context. This is examined with the help of two selected culture proxy moderator variables. Further methodological moderators will also be examined.

In addition to the examination of the conceptual content of the explanatory model, the methodological problems of the TSSEM approach are systematized and approaches for further research are presented. Two of these methodological questions will be empirically examined. First, three methods to take dependent effect sizes into account (averaging, Wilson approach and robust approach) and their effect on the estimated result are compared. In addition, two methods for the examination of moderator variables are compared in the model-based meta-analysis. On the one hand, a subgroup analysis is performed in the conventional TSSEM approach and on

the other hand, the possibilities of the new OSMASEM for the investigation of moderator variables are being used.

The application of the TSSEM approach shows that it can also be used for explorative meta-analysis. However, this leads to specific problems. The unequally populated cells of the correlation matrix lead to convergence problems, especially in the moderator analysis. Findings from the model-based meta-study cannot falsify the positive effect of religiousness on life satisfaction. The three explanation mechanisms examined (social support, optimism and self-esteem) are partial mediators. By considering the mediator variables, the direct effect between religiousness and life satisfaction becomes weaker. It can be assumed that the effect will be even smaller if further explanation mechanisms that are discussed in the literature are included. This can be interpreted as demystification or secularization of the individual effects of religiosity, since these are attributed to psychosocial explanatory mechanisms.

The implementation techniques for the consideration of dependent effect sizes show that the more sophisticated techniques (Wilson approach and robust approach) can increase the precision of estimations with a small number of effect sizes. If, however, there is sufficient empirical material - somewhat over 10 effect sizes - for a given relationship and no research question related to the measuring instruments, a simple mean can be used. The examination of the moderator variables confirms the cultural dependence of the explanatory model. The methodical moderators, on the other hand, can hardly provide further clarification. Regarding the two methods for moderator analysis, the OSMASEM approach has a clear advantage over subgroup analysis. With OSMASEM, more precise conclusions about the effect of the postulated moderator variables are possible. Finally, the systematized desiderata of the TSSEM method are linked with the empirical results of the application to present starting points for further improvements of the methodology.

# 1 Einleitende Worte

Wissenschaft besteht maßgeblich aus der kumulativen Produktion von Wissen (Card 2012: 3; Cooper/Hedges 2009b: 4; Schmidt/Hunter 2015: 17). Wenn jedoch zu einem bestimmten Forschungsfeld bereits zahlreiche empirische Studien vorliegen, so liefern, trotz des wissenschaftlichen Replikationsgedankens, weitere (identische) Studien kaum mehr relevanten Wissenszuwachs (Koenig/King/Carson 2012: 143). Von größerer Relevanz sind dann oftmals (empirische) Studien, die die Frage der dahinterliegenden (kausalen) Mechanismen beantworten (experimentelle Arbeiten), Entwicklungen aufzeigen (Längsschnittuntersuchungen) oder die versuchen, die vorliegenden empirischen Studien zu systematisieren und die Variabilität der Ergebnisse statistisch zu erklären (Meta-Analysen). Experimentelle Arbeiten sind in vielen sozialwissenschaftlichen Bereichen, etwa aufgrund ethischer Überlegungen, nur schwer durchführbar und Längsschnitterhebungen sind in der Regel sehr teuer. Meta-Analysen haben im Gegensatz zu den beiden genannten Verfahrensgruppen eine Reihe von Vorzügen.

Meta-analytische Verfahren können, mit vergleichsweise geringen finanziellen und personellen Mitteln, eine Vielzahl durchgeführter empirischer Studien zu einer bestimmten Forschungsfrage zusammenfassen, sowie Unstimmigkeiten und Muster aufdecken. Einer klassischen Definition nach bezeichnet der Begriff Meta-Analyse „the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings“ (Glass 1976: 3). Es geht also um die (statistische) Analyse von Analyseergebnissen. Es handelt sich im Gegensatz zum narrativen Review um eine Form der systematischen Literatursynthese, die sich explizit einen quantitativen Ansatz zu eigen macht, der es erlaubt, statistisch abgesicherte Schlüsse auf Basis bereits bestehender empirischer Literatur zu ziehen und deshalb einen geringeren Grad an Subjektivität aufweist (Card 2012: 4; Eisend 2014: 2; Kopp/Lois 2014: 210). Diese Form der quantitativen Meta-Analyse hat sich inzwischen zur Standardmethode für die systematische Zusammenführung von empirischen Forschungsergebnissen in vielen wissenschaftlichen Disziplinen entwickelt (Cheung 2008: 183; Cheung/Vijayakumar 2016: 127; Klein/Kopp/Rapp 2013: 234). Meta-Analysen gehören zu den am stärksten zitierten Studien überhaupt und ihnen wird sowohl in der Wissenschaft als auch in der Praxis großes Vertrauen entgegengebracht (Eisend 2017: 12).

Die am häufigsten eingesetzten meta-analytischen Verfahren untersuchen die Beziehung zweier Variablen, obwohl der überwiegende Anteil quantitativer empirischer Forschung inzwischen mithilfe multivariater Modelle durchgeführt wird (Card 2017: 302). Die modellbasierten Meta-Analyse-Verfahren (Becker 2009) hingegen erlauben die Untersuchung von komplexen mul-

tivariaten Modellen etwa unter Berücksichtigung von Erklärungsmechanismen (Mediatorvariablen) oder die Verwendung latenter Konstrukte. Dabei werden meta-analytische Techniken mit den Vorteilen der Strukturgleichungsmodellierung verbunden (Cheung 2015; Jak 2015).<sup>1</sup> Dies wird Meta-Analytic-Structural-Equation-Modelling (MASEM) genannt und dazu liegen verschiedene Umsetzungsvorschläge vor. In dieser Arbeit wird der Two-Stage-Structural-Equation-Modelling-Ansatz (TSSEM) verwendet.

Diese Verfahrensgruppe gilt generell als eine der vielversprechendsten im meta-analytischen Methodenkanon (Cooper/Hedges 2009a: 564ff.). Sie sind jedoch weit weniger bekannt, als die herkömmlichen meta-analytischen Methoden, werden aber aufgrund der Möglichkeit, multivariate Modelle testen zu können, in Zukunft an Bedeutung gewinnen (Cheung 2018: 5). Diese Arbeit macht die Verfahren einem breiten praxisorientierten Rezipientenkreis zugänglich. Das Thema, an dem diese Methode beispielhaft durchgeführt wird, ist bewusst als Querschnittsthema ausgewählt. Es handelt sich um die in unterschiedlichen wissenschaftlichen Disziplinen diskutierte Frage nach Art und Stärke des Zusammenhangs von Religiosität/Spiritualität und (subjektivem) Wohlbefinden bzw. Lebenszufriedenheit und welche Erklärungsmechanismen für die postulierte Wirkrichtung verantwortlich sind.<sup>2</sup> Diese Frage wird in den Geistes- und Sozialwissenschaften, aber auch in verschiedenen Gesundheitswissenschaften<sup>3</sup> ausführlich diskutiert.

---

<sup>1</sup> Zu den Vorteilen der Strukturgleichungsmodellierung zählen unter anderem die simultane Schätzung von mehreren abhängigen und unabhängigen Variablen, die Möglichkeit, latente Konstrukte zu berücksichtigen, die Messfehlerbereinigung sowie die Möglichkeit, die Qualität der Schätzung mittels Fit-Indizes zu testen (Urban/Mayerl 2014: 15f.).

<sup>2</sup> Die Grundlage einer wissenschaftlichen Untersuchung, die das Phänomen der Religiosität bzw. der Spiritualität im Zentrum hat, kann nur auf Grundlage des Prinzips des „Transzendenzausschlusses“ nach Flournoy durchgeführt werden (Klein/Albani 2011b: 227). Diesem Prinzip zufolge ist es für die wissenschaftliche Untersuchung der Auswirkungen von Religiosität unerheblich, ob es übernatürliche oder göttliche Kräfte gibt oder nicht, da der Glauben daran konkrete Folgen für das menschliche Leben hat. Die Diskussion über Sinn und Unsinn von Religion ist eine individuelle oder theologische Frage und sollte deshalb kein Gegenstand sozialwissenschaftlicher Forschung sein.

<sup>3</sup> Wird in der vorliegenden Arbeit von Gesundheitswissenschaften gesprochen, sind damit die Disziplinen Medizin und Psychologie sowie alle angrenzenden (Unter-)Disziplinen (bspw. Pflegewissenschaften) gemeint.

## 1.1 Relevanzbegründung

Religion und deren Folgen für das menschliche Leben als Untersuchungsgegenstand ist bereits seit den frühen Anfängen der Soziologie ein zentrales Thema. Dies zeigt sich darin, welchen zentralen Stellenwert das Phänomen Religion bei den klassischen theoretischen und empirischen Arbeiten von Comte, Durkheim, Marx und Weber einnimmt (Hill et al. 2000: 53). Nach einer zwischenzeitlichen Abkehr vom Untersuchungsgegenstand Religion erfreut sich das Phänomen wieder größerer Beliebtheit in den Sozialwissenschaften (Knoblauch 2008: 42ff.; Koenig/Wolf 2013: 2ff.; Sherkat/Ellison 1999: 386).<sup>4</sup>

Dem Selbstverständnis vieler Religionen nach führt Religiosität zu individuellem Wohlergehen und Gesundheit (Klein/Albani 2011b: 215). In den USA zeigt sich seit den 1980er Jahren eine starke Entwicklung innerhalb der Gesundheitswissenschaften, die die individuelle Religiosität und Spiritualität sowie deren Auswirkungen auf die (mentale und körperliche) Gesundheit in den Fokus nimmt und der Religiosität verschiedene positive Auswirkungen zuschreibt (Angel 2006: 151; Koenig 2008: 349; Koenig/King/Carson 2012: 33ff.; Oman 2013: 23).

Zu den als positiv bewerteten Auswirkungen auf der Individualebene gehören etwa bessere geistige und körperliche Gesundheit, weniger Drogenmissbrauch, niedrigere Sterblichkeit, niedrigere Suizidraten, niedrigere Kriminalitätsraten, besserer Umgang mit Stress sowie eine bessere generelle Lebensbewältigung. Eine der am meisten beachteten positiven Auswirkungen von Religiosität ist jedoch das gesteigerte individuelle (subjektive) Wohlbefinden (Argyle 2003: 365ff.). Viele Autoren in der Tradition der positiven Psychologie halten Wohlbefinden für den wichtigsten Aspekt des menschlichen Lebens (Diener/Oishi/Lucas 2011: 191; Koenig/King/Carson 2012: 123; Larsen/Eid 2008: 8). Es ist ein zentraler Indikator zur Bewertung des erfolgreichen Alterns und kann Variablen wie Gesundheit, Lebensdauer und Krankheit vorhersagen (Capaldi/Dopko/Zelenski 2014: 2; Pinquart/Sörensen 2000: 187).<sup>5</sup>

---

<sup>4</sup> Ein ähnlicher Trend zeigt sich in Psychologie und Medizin. Die psychologischen Klassiker aus dem 19. und 20. Jahrhundert wie Sigmund Freud oder die Behavioristen um Watson und Skinner waren offen ablehnend gegenüber Religion („kollektive Zwangsneurose“ und „irrationale Weltsicht“). Religion und Religiosität galt keineswegs als legitimer wissenschaftlicher Untersuchungsgegenstand (Klein/Berth/Balck 2011: 11).

<sup>5</sup> Neben den individuellen Wirkungen werden auch die Auswirkungen des aggregierten individuellen Wohlbefindens auf der gesellschaftlichen Ebene und die Nutzung von subjektivem Wohlbefinden als Sozialindikator diskutiert (Diener et al. 2017). Dabei rückt das individuelle Wohlbefinden der Bürger immer stärker als übergeordnetes Ziel allen staatlichen Handelns in den Fokus (Gabriel 2012: 99). In Bhutan etwa gehört inzwischen ein nationaler Wohlbefindens-Index fest zur staatlichen Erfolgskontrolle politischen Handelns (Kahneman/Krueger 2006: 22; Ura 2012). Wohlbefinden ist auch außerhalb der Psychologie und Medizin ein wichtiger Sozialindikator und unterstützt Ökonomen, Politiker und Sozialwissenschaftler in ihrer Entscheidungsfindung (Crespo/Mesurado 2015; Diener/Seligman 2004; Diener/Tay 2014; Fischer/Boer 2011; Greiffenhagen/Greiffenhagen 1988; Mayerl 2001; Veenhoven 2008).

Wie deutlich wird, ist das inhaltliche Thema in verschiedenen wissenschaftlichen Disziplinen relevant. Mit dieser Studie können eine Vielzahl von unterschiedlichen Perspektiven zu einem besseren Problemverständnis integriert werden. Die Erkenntnisse einer solchen Meta-Studie sind deshalb in vielen Bereichen direkt für weitere Forschung und für die Praxis anschlussfähig. Neben diesen inhaltlichen Argumenten gibt es einige auf die Analysemethode bezogene Gründe, die für die inhaltliche Wahl des Untersuchungsgegenstandes sprechen. Die Studienlage in diesem Forschungsbereich ist unübersichtlich, widersprüchlich und generell wenig systematisch und ein dominantes Erklärungsmodell liegt nicht vor (Klein/Berth/Balck 2011: 12). Es liegt somit ausreichend heterogenes, bezogen auf Konzeptualisierung (Definition, Operationalisierung, Messung und Forschungsdesign) und Ergebnis, empirisches Material in Form von Studien vor (Koenig/King/Carson 2012: 144). Viele dieser empirischen Studien basieren jedoch auf kleinen Fallzahlen und sollten nur vorsichtig interpretiert werden, da sie eine geringe Teststärke aufweisen (Diener/Clifton 2002: 209). Darüber hinaus werden häufig spezielle Populationen (Erkrankte, Schüler etc.) untersucht und die Ergebnisse dieser Studien sind nicht ohne weiteres auf die Allgemeinbevölkerung übertragbar. Bei der präsentierten Variablenbeziehung handelt es sich um komplizierte Beziehungen zwischen komplexen sozial konstruierten Phänomenen mit einem hohen Grad an Subjektivität. Damit liegt ein typischer sozialwissenschaftlicher Anwendungsfall vor, bei dem verschiedene theoretische Erklärungsansätze miteinander konkurrieren. Somit ist eine modellbasierte Meta-Analyse besonders für die Untersuchung geeignet, da diese Modellierungsart es erlaubt, komplexe multivariate Modelle (etwa unterschiedliche Mediatoren) meta-analytisch vergleichend zu testen (Bergh et al. 2014; Schmidt/Hunter 2015).

Zu den bislang ausgeführten Gründen kommt hinzu, dass zu diesem Zusammenhang keine der bislang bekannten Meta-Studien ein modellbasiertes Forschungsdesign verwendet und dadurch eine erhebliche Forschungslücke vorliegt. Durch die Berücksichtigung mehrerer Erklärungsmechanismen in einem meta-analytischen Modell kann relevantes Wissen gewonnen und somit die bislang vorliegenden Erkenntnisse, die zumeist auf die Untersuchung von Erklärungsmechanismen verzichten, einem konservativerem wissenschaftlichen Test unterzogen werden. Somit liegen gute Voraussetzungen für die planmäßige Durchführung einer modellbasierten Meta-Analyse vor. Die vollständige Durchführung und Dokumentation des Vorgehens ist besonders wichtig, da in den bisherigen modellbasierten MASEM-Arbeiten der Fokus vorwiegend auf die Einführung der statistischen Analysemethoden gelegt und die Besonderheiten und Probleme bezüglich der davorliegenden meta-analytischen Arbeiten (Problemformulierung, Literatursuche und Studienevaluation etc.) kaum berücksichtigt wurden (Cheung 2015; Jak 2015).

## 1.2 Ziele dieser Arbeit und forschungsleitende Fragen

Die Ziele der vorliegenden Arbeiten ergeben sich aus der Kombination von methodologischer und inhaltlicher Thematik. An dem ausgewählten Zusammenhang werden alle relevanten Schritte einer modellbasierten Meta-Analyse aus Anwendersicht nachvollziehbar dokumentiert und Desiderata gesammelt sowie gleichzeitig auf inhaltlicher Ebene relevantes Wissen für Wissenschaft und Praxis generiert.

Ein großer Teil der bislang vorliegenden Arbeiten zur Forschungsfrage, also dem Zusammenhang von Religiosität und Wohlbefinden, stammt aus den Gesundheitswissenschaften. Diese Arbeiten verwenden in der Regel Untersuchungskonzepte, die zum Großteil empirisch-analytisch abgeleitet sind. Die sozialwissenschaftlichen Studien in diesem Bereich sind vorwiegend aus dem Bereich der Sozialindikatorenforschung (Mochon/Norton/Ariely 2011; Poloma/Pendleton 1990; Sawatzky/Ratner/Chiu 2005). Diese basieren ebenfalls auf empirisch-analytisch hergeleiteten Analysekonzepten, die auf vorangegangenen empirischen Ergebnissen beruhen und kein geschlossenes theoretisches Konzept aufweisen. Sie stützen sich auf gesundheitswissenschaftliche Erklärungsmodelle und berücksichtigen zum Teil den kulturellen Kontext (bspw. Abu-Raiya/Agbaria 2016; Eichhorn 2012; Stavrova/Fetchenhauer/Schlosser 2013).

Die Verknüpfung und Integration der gesundheitswissenschaftlichen Modelle mit der sozialwissenschaftlichen Perspektive stellt den theoretisch-analytischen Rahmen des Forschungsvorhabens dar. Die Einbettung in den aktuellen sozialwissenschaftlichen Diskurs ist in der vorliegenden methodisch-statistisch orientierten Arbeit jedoch nachgeordnet.<sup>6</sup> Allerdings ist die sorgfältige analytische Konstruktion eines entsprechenden Erklärungsmodells im Hinblick auf die enthaltenen Variablen zwingend notwendig, um das modellbasierte MASEM-Verfahren korrekt durchführen und anschaulich darstellen zu können. Die Grundlage für die Konzeptionalisierung und das Forschungsdesign stellt ein systematisches Literaturreview der meta-analytischen Praxis in diesem Forschungsfeld dar.

Durch dieses Vorgehen hat die vorliegende modellbasierte Meta-Analyse einen explorativen Charakter, da sich im skizzierten Forschungsfeld bislang kein dominierendes Erklärungsmodell

---

<sup>6</sup> Ein fruchtbarer Ansatz, diese Fragestellung in den soziologischen Theoriekanon einzuordnen, wäre etwa die Bezugnahme auf die Arbeiten von Georg Simmel. Bei ihm stehen die Wechselwirkungsprozesse zwischen Individuum (Subjekt) und Gesellschaft (Kultur) im Analysefokus (Jakobs 2006). Wie sich im weiteren Verlauf dieser Arbeit zeigen wird, eignet sich jene Konzeption besonders für die Analyse der individuellen Wirkungen von Religiosität auf das subjektive Wohlbefinden unter besonderer Berücksichtigung des kulturellen Kontexts (Gabriel/Reuter/Große Kracht 2010: 71ff.; Rosa/Strecker/Kottmann 2009: 92ff.). Darüber hinaus wären sicherlich auch die soziologischen Diskussionen um die zunehmende Privatisierung, religiöse Individualisierung und die zunehmend „unsichtbare Religion“ im Gegensatz zur klassischen Säkularisierungsthese mit dieser Konzeptionierung anschlussfähig (Knoblauch 2005: 123ff.; Norris/Inglehart 2011; Pickel 2011: 59ff.; Pollack 2015).

ausgebildet hat. Im Vergleich zur aktuellen Anwendungen der Methodik ist dies ungewöhnlich (Hagger/Polet/Lintunen 2018; Lesener/Gusy/Wolter 2019; Scherer/Siddiq/Tondeur 2019). Mit dieser Arbeit kann evaluiert werden, ob das ursprünglich eher für etablierte Modelle („Theorien“) angewendete Analyseverfahren auch für explorative Fragestellungen eingesetzt werden kann.

Darüber hinaus soll diese Arbeit den Weg dafür ebnen, die (modellbasierte) meta-analytische Methodik auch im sozialwissenschaftlichen Kontext bekannt zu machen. Dazu werden alle notwendigen Schritte einer modellbasierten Meta-Analyse durchgeführt und die damit verbundenen Entscheidungspunkte nachvollziehbar dargelegt. Neben der Durchführung einer modellbasierten Meta-Analyse anhand des abgeleiteten Erklärungsmodells werden darüber hinaus ausgewählte kulturelle und methodische Moderatorvariablen getestet. Neben der Einführung in diese Form der meta-analytischen Modellierung werden in dieser Arbeit die zahlreichen noch bestehenden methodischen Lücken dieser Verfahrensgruppe gesammelt und systematisiert sowie potenzielle Erweiterungen vorgestellt. Dabei werden konkrete Ansatzpunkte für weitere Forschungsvorhaben, wie etwa Simulationsstudien, entwickelt. Im Zuge dessen werden zwei methodische Fragestellungen in der Anwendung näher beleuchtet. Es wird das Problem der abhängigen Effektstärken, also dem Problem, dass manche Studien für die gleiche Beziehung mehrere Effektstärken liefern, da sie etwa verschiedene Varianten der abhängigen Variable in ihrem Untersuchungsmodell eingeschlossen haben, näher beleuchtet und verschiedene Verfahren zur Berücksichtigung dieser Problematik miteinander verglichen. Darüber hinaus werden die Ergebnisse von zwei unterschiedlichen Varianten zur Untersuchung von Moderatorvariablen miteinander verglichen. Ziel dieses Untersuchungsschritts ist die Gewinnung von Hinweisen darüber, welchen Einfluss die verschiedenen Verfahren auf die Ergebnisse der Schätzung haben.

### **1.3 Aufbau der Arbeit**

Im ersten Schritt erfolgt eine allgemeine Einführung in die meta-analytische Methodik in Kapitel 2. Dabei wird die Leistungsfähigkeit dieser Verfahren herausgearbeitet, ein Überblick über die Vielzahl von meta-analytischen statistischen Umsetzungsformen gegeben und umrissen, was sich hinter dem Begriff der modellbasierten Meta-Analyse verbirgt (Kapitel 2.1). Daran anknüpfend wird die typische Dokumentations- und Darstellungsform von Meta-Analysen eingeführt (Kapitel 2.2). Den Abschluss des Kapitels bildet die Übersicht über die gegen Meta-Analysen vorgebrachte Kritik (siehe Kapitel 2.3). Im darauffolgenden Kapitel 3 wird der Forschungsstand zum inhaltlichen Gegenstand der Exemplifizierung eingeführt. Dazu werden zunächst die beiden grundlegenden Konzepte dieser Studie Religiosität/Spiritualität (Kapitel 3.1)

und Wohlbefinden bzw. Lebenszufriedenheit (Kapitel 3.2) definiert, bevor der Zusammenhang analytisch-empirisch begründet (Kapitel 3.3) und ein umfassendes Erklärungsmodell (Kapitel 3.4) abgeleitet wird. Danach folgt die Entwicklung des Forschungsdesigns der Studie in Kapitel 4. Dazu wird zunächst ein methodologisches Review bereits durchgeführter Meta-Studien in diesem Bereich angefertigt (Kapitel 4.1). Basierend auf den Erkenntnissen dieses Reviews werden dann die notwendigen meta-analytischen Vorarbeiten bezüglich Selektionskriterien (Kapitel 4.2), Literatursuche (Kapitel 4.3), Auswahl relevanter Publikationen (Kapitel 4.4) und Kodierung der Studien (Kapitel 4.5) festgelegt.

Anschließend (Kapitel 5) wird die statistische Methodik der modellbasierten Meta-Analyse, die im Zentrum dieser Arbeit steht, vorgestellt. Dazu werden zunächst die relevanten grundlegenden statistischen Konzepte der Meta-Analyse eingeführt (Kapitel 5.1). Zum besseren Verständnis wird die SEM-basierte Meta-Analyse von MASEM abgegrenzt (Kapitel 5.2). Danach folgt die Erläuterung in die klassischen MASEM-Ansätze (Kapitel 5.3). Sodann folgt das zentrale Verfahren der vorliegenden Arbeit – der TSSEM-Ansatz – in Kapitel 5.4. Dort werden dessen Grundlogik, sein Ablauf, vorgeschlagene Erweiterungen und Spezifika erläutert sowie Desiderate umfassend systematisiert. Den Abschluss der methodischen Ausführungen bildet die Diskussion zweier spezifischer Aspekte modellbasierter Meta-Analysen, die in der Literatur bislang kaum behandelt werden, nämlich der Publikationsbias und die Sensitivitätsanalyse (Kapitel 5.5). Danach folgen die Ergebnisse der Studierhebung (Kapitel 6). Dabei wird zunächst über die Erhebung der Studien berichtet (Kapitel 6.1) und anschließend das (Vor-)Screening der Studien dokumentiert (Kapitel 6.2). Direkt daran anknüpfend folgt die Dokumentation der Studienkodierung (Kapitel 7). Zunächst wird das Vorgehen nochmals nachvollziehbar dargestellt, bevor die Ergebnisse dieses Arbeitsschritts dokumentiert werden (Kapitel 7.2).

In Kapitel 8 werden Erwartungen bezüglich der Analyse formuliert. Den Höhepunkt bildet die statistische Analyse der gewonnenen Daten (Kapitel 9). Zunächst werden die programmtechnische Umsetzung (Kapitel 9.1) und die einzelnen Analyseschritte beschrieben (Kapitel 9.2). Danach folgen zunächst die inhaltlichen (Kapitel 9.3 und 9.4) und dann die methodischen Analysen (Kapitel 9.5). Abschließend erfolgt die Überprüfung der meta-analytischen Ergebnisse bezüglich des Vorliegens eines Publikationsbias (Kapitel 9.5).

Abschließend werden in Kapitel 10 die gewonnenen inhaltlichen und methodischen Erkenntnisse zunächst systematisiert (Kapitel 10.1). Danach werden die Grenzen der vorliegenden Meta-Analyse diskutiert (Kapitel 10.2) und Anknüpfungspunkte für weitere Forschungsvorhaben formuliert (Kapitel 10.3).



## 2 Einführende Anmerkungen zur Meta-Analyse

Eine Meta-Analyse ist die quantitative Form einer wissenschaftlichen Übersichtsarbeit (Literaturreview) mit dem Ziel einer Forschungssynthese auf Basis von numerischen Primärstudien-ergebnissen (Card 2012: 5f.).<sup>7</sup> Das Ziel einer solchen Meta-Analyse besteht dabei nicht nur in der systematischen Zusammenfassung von Primärstudien-ergebnissen, sondern darüber hinaus wird häufig die Ergebniskonsistenz und Verallgemeinerbarkeit überprüft (Eller 2015: 80). Eine Meta-Analyse ist somit ein quantitativ-statistisches Verfahren zur Zusammenfassung von Primärstudien-ergebnissen zur verbesserten Ergebnisintegration, zur Systematisierung eines Themenfelds und darüber hinaus zur Erklärung der potenziellen Befundheterogenität.

### 2.1 Leistung, Ablauf und Einordnung

Wie alle Formen des Literaturreviews systematisieren Meta-Studien den Stand der Forschung und können darüber hinaus Forschungstrends über die Zeit hinweg darstellen (Cooper/Hedges 2009a: 563). Im Gegensatz zu den anderen Formen der Literatursynthese, wie etwa einem qualitativen narrativen Review, ermöglichen Meta-Analysen jedoch statistisch abgesicherte und zuverlässige Schlussfolgerungen (Weiß/Wagner 2008: 251).

Die Leistungsfähigkeit von Meta-Analysen beruht vor allem darin, dass es ihnen gelingt die Konsistenz des gesamten zu einer konkreten Fragestellung vorliegenden empirischen Materials zu überprüfen, die unterschiedlichen Ergebnisse zu erklären, die Generalisierbarkeit von Erkenntnissen zu erhöhen und handfeste Erkenntnisse für die Praxis zu liefern (Eisend 2017: 12). Aufgrund der größeren Menge von Daten, die in einer Meta-Analyse genutzt werden, wird die Präzision und die Genauigkeit der Schätzungen im Vergleich zu einer einzelnen Studie deutlich erhöht (Jak 2015: 2). Darüber hinaus ist es möglich, widersprüchliche Studienergebnisse mittels Moderatoranalysen aufzuklären (Eisend 2014: 69). Die größeren Stichproben und die damit einhergehende höhere Teststärke sowie der verringerte Einfluss des Stichprobenfehlers hat robustere Parameterschätzungen zur Folge (Sheng et al. 2016: 187). Meta-analytische Ergebnisse werden deshalb von Entscheidungsträgern als Grundlage für wichtige Entscheidungen beispielsweise in der evidenzbasierten Politik herangezogen.

Ein weiterer Vorteil von Meta-Analysen besteht darin, dass Hypothesen getestet werden können, die in keiner Primärstudie zuvor getestet wurden (Cooper/Hedges 2009a: 563). Besonders in modellbasierten meta-analytischen Verfahren kommt dieser Vorteil zum Tragen, da dort in der Regel komplexe analytische Modelle vorliegen (Jak 2015: 12f.). Darüber hinaus können

---

<sup>7</sup> Es liegen jedoch auch qualitative Ansätze für die Durchführung von Meta-Analysen vor (Ruddat/Schulz 2010).

Meta-Analysen ausgedehntere Fragestellungen beantworten, da eine Vielzahl von unterschiedlichen Forschungsdesigns (etwa mit unterschiedlichen Messinstrumenten) berücksichtigt werden können (Borenstein et al. 2011: 358). Zusammenfassend haben sie eine höhere Präzision und Reliabilität aufgrund der systematischen Literatursuche und der Kodierung sowie den verwendeten statistischen Methoden zur Ergebnisintegration.

Sie tragen somit einen bedeutsamen Teil zur kumulativen Wissensgenerierung bei, indem sie es erlauben, den Informationsgehalt von Erklärungsmodellen empirisch zu überprüfen und die wissenschaftliche Suche nach universell gültigen Gesetzen zu unterstützen. Aufgrund des (exponentiellen) Wachstums von wissenschaftlichen Veröffentlichungen wird diese Form der Ergebnisintegration in den kommenden Jahren weiter zunehmen (Eisend 2014: 1; Jak/Cheung 2018a: 3; Kopp/Lois 2014: 208).<sup>8</sup> Allerdings ist die Methode in den Sozialwissenschaften im Gegensatz zu Disziplinen wie Medizin und Psychologie bislang wenig verbreitet (Klein/Kopp/Rapp 2013: 223; Wagner/Weiß 2003: 30; Weiß/Wagner 2008: 250).<sup>9</sup> Der hauptsächlichste Grund hierfür liegt vermutlich darin, dass dort nur wenige experimentelle Studien, sondern zumeist ein ex-post-facto Forschungsdesign, basierend auf Umfragedaten, durchgeführt wird (Brüderl 2004: 84; Kopp/Lois 2014: 211).<sup>10</sup>

Der Aufbau einer Meta-Analyse orientiert sich an fünf aufeinanderfolgenden Phasen (Cooper/Hedges 2009b: 9). Diese sind unabhängig von der statistischen Umsetzung und sind auch Grundlage der vorliegenden modellbasierten Meta-Analyse (siehe Abbildung 1).

Zur Durchführung von Meta-Analysen haben sich inzwischen eine Vielzahl von Modellierungstechniken herausgebildet. Prinzipiell lassen sich zwei Formen unterscheiden. Zum einen gibt es die individualdatenbasierte Meta-Analyse (Medizin: individual patient data (IPD); Psychologie: integrative data analysis), die die Synthese auf Grundlage der Primärdaten vornimmt, und zum anderen die stärker verbreitete aggregatdatenbasierte Meta-Analyse (APD: aggregate

---

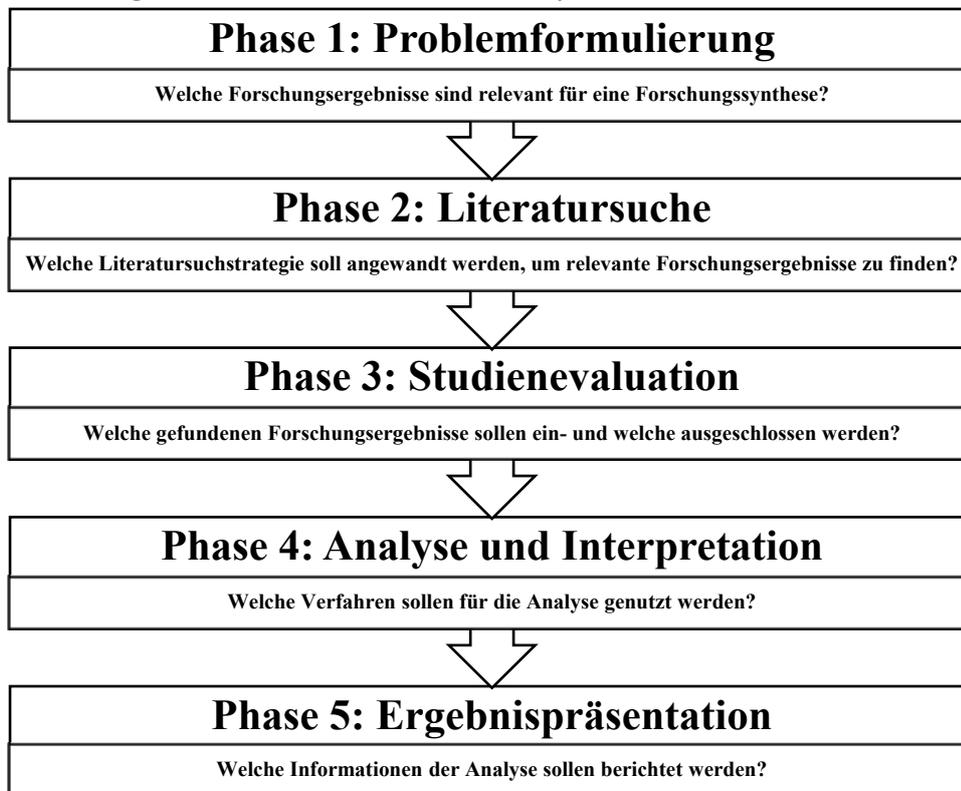
<sup>8</sup> Dies gilt insbesondere für die Sozialwissenschaften, weil dort der Bedarf an Meta-Analysen aufgrund der zunehmend steigenden Anzahl an empirischen Studien in den kommenden Jahren weiter wachsen wird (Weiß/Wagner 2008: 255).

<sup>9</sup> Bei einer systematischen Literatursuche haben Weiß/Wagner (2008) etwa für den Zeitraum von 1975 bis 2005 durchschnittlich lediglich acht soziologische Meta-Analysen pro Jahr entdeckt. Auch in den darauffolgenden Jahren hat sich dies kaum verändert.

<sup>10</sup> Beziehungen zwischen sozialwissenschaftlichen Variablen, die auf solchen Daten beruhen, sind deshalb in der Regel durch andere Faktoren konfundiert und geben lediglich einen verfälschten Eindruck über den „wahren“ Effekt wieder. Nicht durch die Kontrolle von Randbedingungen, wie in Experimenten üblich, kann der Einfluss von Drittvariablen ausgeschlossen werden, sondern etwa durch die Durchführung von Mediations- und Suppressionsanalysen in Form von multivariaten Analysen (Klein/Kopp/Rapp 2013: 235).

patient data), die auf Grundlage der Ergebnisse von Primärstudien vorgeht. Die Ausführungen in dieser Arbeit beschränken sich auf die APD-Form.<sup>11</sup>

**Abbildung 1: Fünf Phasen einer Meta-Analyse**

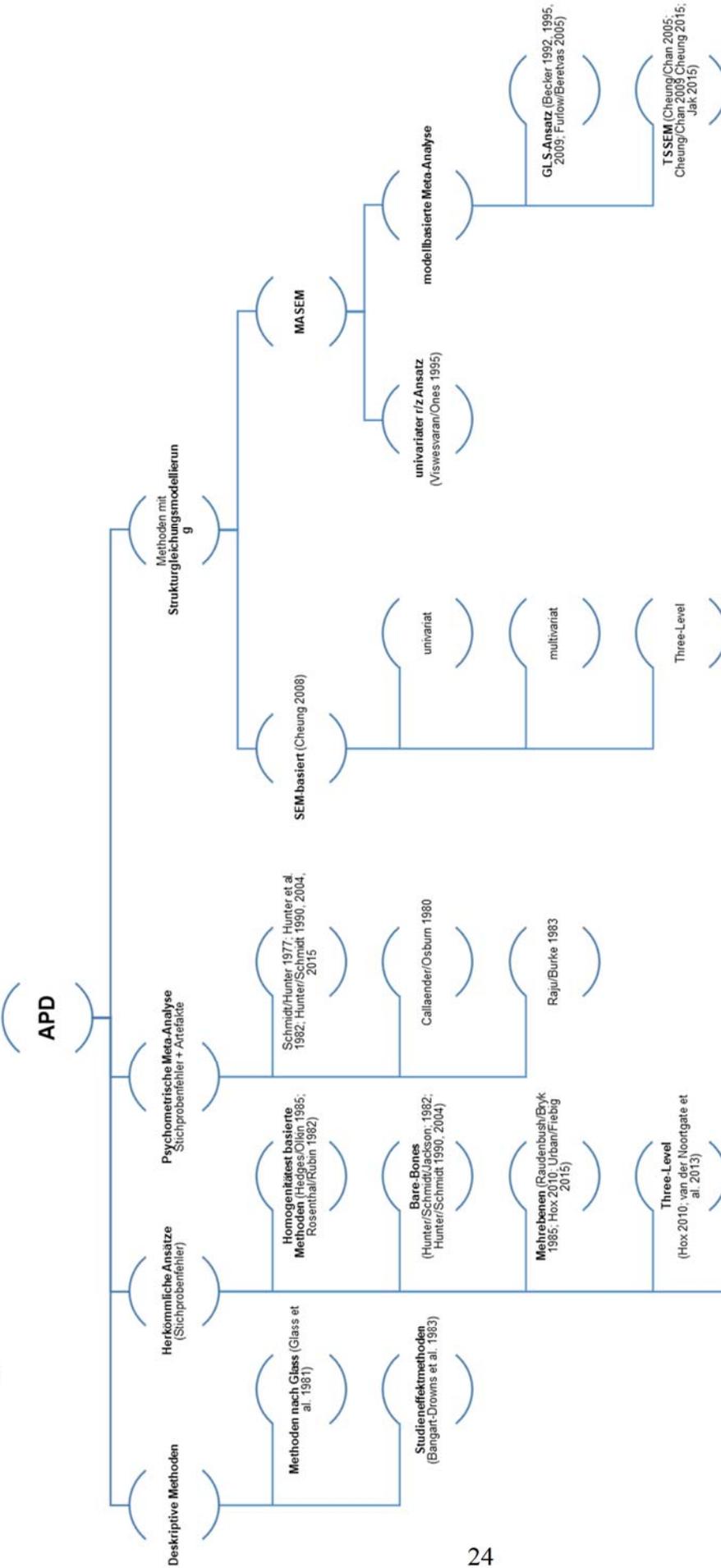


**Anmerkungen:** Eigene Abbildung in Anlehnung an Cooper/Hedges (2009b).

Eine umfassende, aber dennoch unvollständige Übersicht über die Vielzahl von statistischen Umsetzungsarten aus dem Bereich der Meta-Analysen, die auf Aggregatdaten basieren, ist in Abbildung 2 zu sehen. Auf der linken Seite der Abbildung sind die drei Verfahrensgruppen der herkömmlichen Meta-Analyse abgebildet. Dazu zählen die deskriptiven Methoden, die Ansätze, bei denen der Stichprobenfehler berücksichtigt wird und die Gruppe der psychometrischen Verfahren, die Artefakte statistisch korrigieren (beispielsweise Reliabilität der Messinstrumente). Auf der rechten Seite befinden sich die neueren Verfahren, die auf Strukturgleichungsmodellierung basieren.

<sup>11</sup> Informationen zur immer stärker verbreiteten IPD und den vielen damit verknüpften Vorteilen bei der Verwendung von Rohdaten, wie etwa die Möglichkeit, robuste Schätzer verwenden zu können, können an anderer Stelle nachgelesen werden (Pigott 2012: 109ff.). Bei dieser Form der Meta-Analyse werden nicht die veröffentlichten Ergebnisse von Studien, sondern deren ursprüngliche Datenbasis für die metaanalytische Synthese genutzt (Klein/Kopp/Rapp 2013: 227f.). Allerdings ist diese Form der Meta-Analyse aufgrund der sehr aufwendigen Harmonisierungsarbeit zwischen den Datensätzen und deren Beschaffung nur sehr schwer durchführbar. In vielen Wissenschaftsbereichen werden darüber hinaus die Rohdaten unter anderem aus Datenschutzgründen nicht veröffentlicht, so dass die Verwendung der IPD-Verfahren dort nur selten möglich ist (Cheung 2014a: 38).

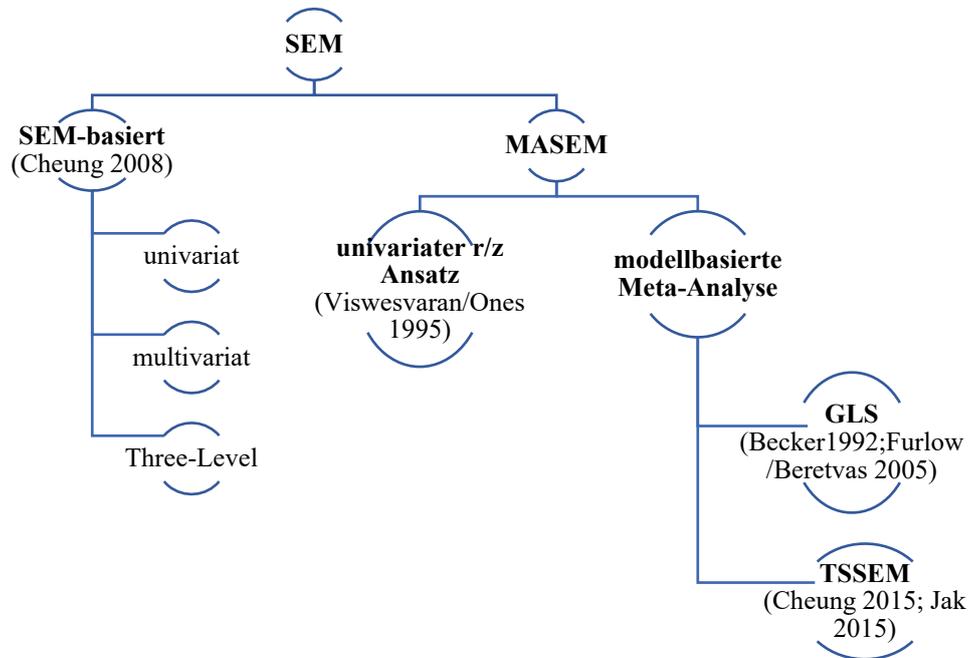
Abbildung 2: Detaillierte Übersicht meta-analytischer APD-Verfahren



**Anmerkungen:** Eigene Abbildung in Anlehnung an Schmidt/Hunter (2015: 465) erweitert. Von jedem Ansatz gibt es noch weitere Abwandlungen, die in dieser Übersicht nicht berücksichtigt werden.

In Abbildung 3 sind nur die auf Strukturgleichungsmodellierung (SEM) beruhenden meta-analytischen Verfahren zu sehen.<sup>12</sup> Zu dieser Verfahrensgruppe zählen die modellbasierten multivariaten Verfahren (Becker 2009), die es erlauben, komplexe theoretisch-analytische Modelle mit mehreren exogenen und endogenen Variablen sowie Mediatorbeziehungen empirisch zu testen.

**Abbildung 3: Fortgeschrittene Verfahren der Meta-Analyse**



**Anmerkung:** Eigene Abbildung.

In der Literatur gibt es einige Verwirrung um die korrekte Bezeichnung dieser Verfahrensgruppe. Beispielsweise sind die folgenden Begriffe (bzw. deren englischen Entsprechungen) sowie Abwandlungen davon im Gebrauch: SEM-basierte Meta-Analyse, meta-analytische Pfadanalyse, verbundene Meta-Analyse, Meta-Analyse von Faktoranalysen, meta-analytische Strukturgleichungsanalyse, Pfadanalyse von meta-analytischen Korrelationsmatrizen, Strukturgleichungsmodellierung von meta-analytischen Korrelationsmatrizen, Pfadanalyse basierend auf meta-analytischen Untersuchungsergebnissen und modellbasierte Meta-Analyse (Becker 2009: 378ff.; Cheung 2014a: 30; Cheung/Hafdahl 2016: 112; Lipsey 1997: 216ff.). Viele dieser Begriffe beschreiben jedoch das gleiche Verfahren: Meta-Analytic-Structural-Equation-Modelling (MASEM).

<sup>12</sup> Unter Strukturgleichungsmodellierung (SEM) wird ein (multivariates) statistisches Verfahren verstanden, bei der simultan alle Beziehungen eines hypothetischen (theoretisch-) analytisch formulierten Modells unter Berücksichtigung von latenten und manifesten Variablen auf Grundlage der beobachteten Varianz-Kovarianzmatrix geschätzt werden können. Komplexe Theorien können somit als testbare Hypothesen mittels Strukturgleichungsmodellierung formuliert und überprüft werden. Es handelt sich also um eine Kombination von Pfad- und Faktoranalyse. Programme wie *Mplus* vereinigen inzwischen neben klassischen SEM-Verfahren auch Mehrebenenmodelle, Latent-Class-Modelle, IRT-Modelle, Bayes-Statistik und viele weitere Ansätze unter einem einheitlichen statistischen Modellierungsframework (Cheung 2015: 5; Muthén/Muthén 2017).

Bei der **SEM-basierten Meta-Analyse** handelt es sich um ein Verfahren, welches SEM-Techniken nutzt, um herkömmliche Fixed- oder Random-Effects-Modelle durchzuführen (Cheung/Hafdahl 2016: 112; Jak 2015: 13). Dort werden meta-analytische Modelle als Spezialfall eines Strukturgleichungsmodells ausgedrückt und getestet. Bei den SEM-basierten Verfahren geht es dementsprechend vorwiegend darum, die Vorteile der Strukturgleichungsmodellierung und der entsprechenden Programme für die praktische Umsetzung von Meta-Analysen nützlich zu machen.

Die **MASEM-Verfahren** werden hingegen darüber definiert, dass die Informationen einer Vielzahl einzelner und gegebenenfalls sehr unterschiedlicher Studien mithilfe von Strukturgleichungsmodellierung zusammengefasst werden, mit dem Ziel, theoretisch-analytisch hergeleitete Modelle zu testen (Becker 1992; Cheung/Chan 2005b; Viswesvaran/Ones 1995). Die zusammengefassten Ergebnisse, in Form einer kumulierten Korrelationsmatrix, dienen als Grundlage für den Test des theoretisch-analytisch postulierten Modells (Landis 2013: 251). Derzeit werden im Wesentlichen drei konkrete MASEM-Ansätze sowie Erweiterungen von diesen diskutiert (Sheng et al. 2016: 189). Dazu gehört der ursprünglich von Viswesvaran/Ones (1995) vorgeschlagene und in der Praxis oftmals angewandte univariate  $r/z$ -Ansatz sowie der von Becker (1992) bzw. Becker/Schram (1994) eingeführte multivariate Generalized-Least-Square-Ansatz. In Abgrenzung zu den Schwächen dieser beiden Verfahren haben Cheung/Chan (2005b) den TSSEM-Ansatz entwickelt.

Mit den multivariaten Ansätzen können die Mechanismen (in Form von Mediatoren) eines Erklärungsmodells sichtbar gemacht werden (Tang/Cheung 2016: 473). Somit können Modelle und Forschungsfragen untersucht werden, die den Komplexitätsgrad von bivariaten Meta-Analysen deutlich überschreiten. In dieser Arbeit gelten nur die multivariaten Ansätze GLS und TSSEM als modellbasierte Verfahren, da nur dort die Abhängigkeit der Effektstärken untereinander berücksichtigt wird (Becker 2009: 378). Im Vergleich zu den herkömmlichen Methoden wird die modellbasierte Meta-Analyse bislang kaum eingesetzt (Cheung/Cheung 2016: 141; Cheung/Hafdahl 2016: 112). Die Fallzahl einer MASEM ist in der Regel höher als in einer SEM-Analyse und hat damit einhergehend eine höhere Teststärke (Bergh et al. 2014: 478; Landis 2013: 252).<sup>13</sup> Eine MASEM liefert Effektstärken, die nach anderen Modellvariablen kontrolliert wurden, und stellt Fit-Indizes über die Gesamtpassung des Modells bereit. MASEM liefert somit einen besseren und vor allem detailreicheren Ansatz zur Zusammenfassung von Forschungsergebnissen als die traditionelle Meta-Analyse.

---

<sup>13</sup> Die MASEM-Parameterschätzungen und die Fit-Indizes sind stabiler als in herkömmlichen SEM-Analysen.

Der größte Vorteil liegt darin, dass komplexe Modelle, mit einer Vielzahl von unterschiedlichen Wirkpfaden, getestet werden können. Dabei muss keine der Primärstudien alle Untersuchungsvariablen berücksichtigt haben, sondern es genügt, wenn jeweils nur Teilausschnitte des Modells getestet wurden (Landis 2013: 251; Sheng et al. 2016: 187; Viswesvaran/Ones 1995: 866). Somit können Forschungsfragen beantwortet werden, die in keiner Primärstudie direkt gestellt wurden und die unter Umständen sehr komplex sind und eine hohe Anzahl von Variablen aufweisen.

Wissenschaftler sind bei empirischen Untersuchungen häufig auf der Suche nach validen kausalen Aussagen. Dies gilt natürlich ebenso für meta-analytische Fragestellungen. Ergebnisse von Meta-Analysen, deren Grundlage experimentelle Studien sind, können kausal interpretiert werden. Meta-Analysen hingegen, die auf bivariaten korrelativen Variablenbeziehungen beruhen, sind lediglich eine schwache Basis für valide Inferenzschlüsse, da sie den Einfluss von Drittvariablen nicht kontrollieren (Steinmetz/Isidor 2013: 1). Darüber hinaus kann die „wahre“ kausale Struktur, die sich hinter einer bivariaten Korrelation verbirgt, komplex sein und bivariate Analysen zu falschen Schlussfolgerungen führen. Deshalb sollten Drittvariablen und Mediatoren als Erklärungsmechanismen bei der statistischen Modellierung berücksichtigt werden.<sup>14</sup> Dies ist mit gebräuchlichen meta-analytischen Verfahren kaum möglich. Die Kombination von Strukturgleichungsmodellierung und Meta-Analyse (MASEM) ist dazu in der Lage, da die vielen Vorteile des SEM-Verfahrens, wie etwa die simultane Schätzung aller Beziehungen eines Pfadmodells, auf meta-analytische Fragestellungen angewendet werden können.<sup>15</sup>

Da durch das MASEM-Verfahren alle vorhandenen empirischen Daten für die Überprüfung eines spezifischen Beziehungsgeflechts verwendet werden, erhöht sich die externe Validität im Vergleich zu univariaten meta-analytischen Verfahren deutlich. Im Gegensatz zu herkömmlichen bivariaten Meta-Analysen kann eine Vielzahl von Prädiktoren mit einer oder mehreren Outcome-Variable(n) in Beziehung gesetzt und somit direkte und indirekte Effekte getestet werden (Becker 2009: 380). Darüber hinaus können Beziehungen aufgedeckt werden, die in

---

<sup>14</sup> In der herkömmlichen empirischen Forschung sollten deshalb in einem Regressionsmodell stets Drittvariablen aufgenommen werden, um einen „reineren“ Effekt zu erhalten oder Pfadanalysen mit mehreren abhängigen Variablen zum Auffinden von Mediatoren durchgeführt werden (Urban/Mayerl 2018).

<sup>15</sup> Mit dieser Kombination kann der sogenannte „conformation bias“, also die Voreingenommenheit mit dem eigenen Modell, vermieden und verschiedene Modelle im Hinblick auf ihre meta-analytische Erklärungskraft miteinander verglichen werden (Cheung/Chan 2005b: 40). Ein solcher Modellvergleich ist mit den verbreiteten meta-analytischen Methoden nicht möglich.

der bisherigen Forschung kaum untersucht wurden. Somit können ganze „Theorien“ meta-analytisch evaluiert und etwa konkurrierende Erklärungsansätze miteinander verglichen und somit ein Beitrag zur Theorieentwicklung geleistet werden (Card 2012: 305).

Das Untersuchungsmodell kann in unterschiedlichen Stichproben und unter Verwendung von unterschiedlichen Messinstrumenten getestet werden (Cheung 2015: 215; Sheng et al. 2016: 187). Eine solche Modellüberprüfung ermöglicht die statistische Überprüfung der externen Validität. Es können alternative Erklärungsmodelle spezifiziert und überprüft werden, welches Modell am besten zu den empirischen Daten passt. Die hinter den Effekten vermuteten indirekten Mechanismen (Mediatorbeziehungen) können direkt ins Modell aufgenommen und modellspezifisch getestet werden. Dadurch können Black-Boxes geöffnet und somit differenziertere Erklärungen für soziale Phänomene gefunden und darüber hinaus die Generalisierbarkeit der Ergebnisse erhöht werden (Bergh et al. 2014: 478; Cheung/Cheung 2016: 140). Ein Erklärungsmodell, das sich in einem solch strengen Test bewährt, kann anschließend Grundlage für weitere Forschungsvorhaben bilden.

## **2.2 Dokumentation von Meta-Analysen**

Wie bereits angemerkt, ist der Ablauf einer Meta-Analyse in mehrere aufeinanderfolgende Phasen gegliedert. Die Dokumentation der Meta-Analyse orientiert sich an diesen Phasen. Bezüglich der Grundstruktur gibt es keinen Unterschied zwischen einer herkömmlichen und einer modellbasierten Meta-Analyse. Die sorgfältige Dokumentation aller methodologischen Entscheidungen ist bei der Durchführung einer Meta-Analyse von zentraler Bedeutung.

Das Ziel der meta-analytischen Darstellung ist es, die Ergebnisse verständlich und systematisch zu präsentieren, damit Rezipienten den größtmöglichen Nutzen aus einer Meta-Analyse ziehen können (Card 2012: 315f.; Clarke 2009: 522; Möser/Schmidt 2016: 20ff.). Meta-Analysen wirken, aufgrund der komplexen, teilweise hierarchischen Datenstruktur, der ungewohnten Analysestrategien und der Vielzahl von Modellierungsschritten, oftmals technisch und sind nur schwer nachvollziehbar (Borman/Grigg 2009: 498ff.). Deshalb ist es wichtig, dass die Ergebnisse nachvollziehbar beschrieben und visuell anschaulich präsentiert werden.

Die Präsentation und Interpretation von meta-analytischen Ergebnissen folgt dabei den allgemeinen Empfehlungen der empirischen Forschung. Das heißt, die Ergebnisse werden nachvollziehbar beschrieben und zusammengefasst, Implikationen für die Theorie und Praxis herausgearbeitet, Forschungslücken aufgezeigt und methodische Einschränkungen diskutiert (Eisend 2017: 12). Allerdings gibt es bei Meta-Analysen spezifische Aspekte, die berücksichtigt werden müssen. Es wird empfohlen, die anerkannten Berichtstandards „Meta-Analysis-Reporting-

Standards“ (MARS) und/oder „Preferred Reporting Items for Systematic Reviews and Meta-Analyses“ (PRISMA) zu beachten (American Psychological Association 2008; Liberati et al. 2009). Die Standards decken die relevanten methodischen Aspekte einer Meta-Analyse ab. In Tabelle 1 ist eine verkürzte Übersicht über die MARS-Reporting-Standards zu sehen. Ein Großteil der Inhalte von MARS bezieht sich dabei auf die ersten Phasen einer Meta-Analyse.

**Tabelle 1: Meta-Analysis Reporting Standards (MARS)**

Abschnitt	Unterpunkte	Inhalt
Titel		Es sollte bereits dort deutlich werden, dass es sich um eine Forschungssynthese mit Meta-Analyse handelt
Abstract		Problemdefinition Selektionskriterien Probanden Meta-analytische Methoden/Modellierungsart Zentrale Ergebnisse Schlussfolgerungen Implikationen für Forschung und Praxis
Einleitung		Theoretisch-analytische Herleitung der Forschungsfrage und des Untersuchungsmodells Forschungsstand Operationale Definition der Kernkonstrukte
Methodik	Selektionskriterien	Grundgesamtheit Forschungsdesign Zeitraumen Geographische/kulturelle Eingrenzung
	Moderatoranalyse	Definition der Kodinkategorien Dokumentation der Datenerhebung (Datenbanken, Suchbegriffe, Zeitraumen)
	Suchstrategien	Dokumentation des Screening
	Kodierprozedur	Art der Effektstärke Umgang mit fehlenden Werten Studienqualität
	Statistische Analysemethoden	Modellierungsart (Fixed vs. Random Effects) Form der Moderatoranalyse Überprüfung Publikationsbias Gewichtung Umgang mit statistischer Abhängigkeit Umgang mit fehlenden Werten Programme
Ergebnisse		Anzahl der Primärstudien Deskriptive Informationen (Effektstärke und Fallzahl) Kumulierte Effektstärke Tabellen und Graphiken Ergebnisse der Moderator- und Mediatoranalysen
Diskussion		Zentrale Erkenntnisse Alternative Erkenntnisse Generalisierbarkeit der Schlussfolgerungen Schwächen der Studie Implikationen für Forschung und Praxis

**Anmerkungen:** In Anlehnung an Clarke (2009) stark gekürzt und verändert in Anlehnung an Card (2012).

Zu den genannten Reporting-Standards liegen ausführliche Checklisten vor, die beim Verfassen einer Meta-Analyse zu Rate gezogen werden können (Clarke 2009: 523f.; Liberati et al. 2009:

3ff.).<sup>16</sup> Der große Schwachpunkt der vorgefertigten Standards, aus Sicht der vorliegenden Arbeit, besteht vor allem darin, dass modellbasierte Meta-Analysen bislang nicht berücksichtigt werden. Für den meta-analytischen Test von komplexen Variablenbeziehungen sind diese Empfehlungen bislang nicht erweitert worden. Zwar gibt es für diese Analysemodelle bislang keine ausgearbeiteten Berichtstandards, aber man kann sich an vorliegenden MASEM-Studien orientieren (Gu et al. 2015; Lesener/Gusy/Wolter 2019; Scherer/Siddiq/Tondeur 2019). Zum PRISMA-Standard gehört ein Flussdiagramm, in dem alle relevanten Informationen zur Erhebung der Primärstudien dargestellt sind (Liberati et al. 2009: 4). Aus dieser Darstellung wird ersichtlich, wie viele Studien mit der Suchstrategie gefunden und nach Screening und Kodierung in der Meta-Analyse berücksichtigt werden.

Neben der schriftlichen Dokumentation aller wichtigen Entscheidungspunkte ist die graphische Darstellung einer Meta-Analyse von zentraler Bedeutung. Üblicherweise werden die Ergebnisse von Meta-Analysen in Tabellen und Abbildungen präsentiert (Anzures-Cabrera/Higgins 2010: 67ff.; Card 2012: 329ff.; Schewe/Hülshager/Maier 2014: 199). Dabei werden Ergebnistabellen präsentiert, in denen auf den ersten Blick Effekte, Signifikanzen, Moderatoreffekte, Heterogenität etc. zu erkennen sind (Borman/Grigg 2009: 498).

Die typischen Darstellungsformen für Meta-Analysen sind sogenannte Forest- oder Konfidenzintervallplots (Eisend 2017: 12).<sup>17</sup> Dort wird das Ergebnis der Effektschätzung im Kontext zu den anderen eingeschlossenen Studien veranschaulicht (Anzures-Cabrera/Higgins 2010: 67ff.). Jede Studie wird in einer solchen Abbildung mit einem Punktschätzer, einem Konfidenzintervall und teilweise mit einem relativen Gewicht dargestellt (Borenstein et al. 2011: 366f.). So können der mittlere Effekt und die einzelnen Studieneffekte sofort miteinander verglichen werden. Werden die Effektstärken der Größe nach sortiert, spricht man von einem Caterpillar-Plot. In einer modellbasierten Meta-Analyse sind die Darstellungsoptionen hingegen begrenzt, da normalerweise mehr als eine bivariate Variablenbeziehung betrachtet wird. Die graphischen Darstellungsformen kommen dabei schnell an ihre Grenzen. Jedoch können die Darstellungsmöglichkeiten für die bivariaten Zusammenhänge eingesetzt werden. Eine wichtige Rolle spielt die tabellarische Darstellung etwa der kumulierten Korrelationsmatrix und der zugehörigen Heterogenität (Jak 2015: 62; Quinn 2016: 36ff.).

---

<sup>16</sup> Die Reporting-Standards sind vorwiegend aus medizinischer und psychologischer Sicht entwickelt und fokussieren sich deshalb oftmals auf experimentelle Arbeiten. Allerdings decken sich diese im Wesentlichen mit den entsprechenden Empfehlungen aus sozialwissenschaftlicher Sicht (Card 2012: 317ff.).

<sup>17</sup> Neben den genannten Darstellungsformen gibt es noch eine Vielzahl weiterer Möglichkeiten der visuellen Ergebnisdarstellung für spezifische Formen der Meta-Analyse (Anzures-Cabrera/Higgins 2010; Borman/Grigg 2009).

Das Gesamtergebnis einer modellbasierten Meta-Studie lässt sich am besten in Form eines Pfaddiagramms darstellen, bei denen die Parameterschätzungen mit den zugehörigen Signifikanzen (mit Konfidenzintervall) und den Residualvarianzen zu sehen sind (Becker 2009: 387ff.; Cheung 2017b: 37ff.; Jak 2015: 68). In dieser Arbeit wird eine Ergebnisdarstellung angewandt, die sich an den genannten meta-analytischen Empfehlungen orientiert, in Verbindung mit Vorschlägen aus vorliegenden MASEM-Studien.

### **2.3 Exkurs: Kritische Würdigung des Verfahrens**

Bevor das Forschungsdesign der Arbeit dargelegt wird, folgt die Würdigung der gegenüber von Meta-Analysen geäußerten Kritik (Borenstein et al. 2011: 377ff.; Card 2012: 23ff.; Klein/Kopp/Rapp 2013: 254f.; Kopp/Lois 2014: 209). Diese Kritikpunkte gelten für herkömmliche und modellbasierte Meta-Analyse gleichermaßen. Die diskutierten Aspekte stellen eine ernstzunehmende Gefahr für die Validität der meta-analytischen Schlüsse dar und bedrohen damit deren Generalisierbarkeit (Matt/Cook 2009: 540). Diese Faktoren sollten deshalb bereits bei der Planung einer Meta-Analyse besonders bedacht werden. Anknüpfend an die vorgebrachten Schwächen von Meta-Analysen werden die entsprechenden Gegenargumente vorgebracht, um das Für und Wider der Methodik angemessen abwiegen zu können.

Oftmals wird behauptet, dass einer empirischen Meta-Analyse die kreative detailreiche qualitative Finesse eines narrativen Reviews fehlt (Card 2012: 24f.). Die Autoren von Meta-Analysen, so der Vorwurf, vernachlässigen die qualitative theoretische Einbettung der Ergebnisse in den Forschungskontext. Der Vorwurf kann leicht abgewiesen werden, da jede quantitative Meta-Analyse nach der Erhebungsphase einen qualitativen Review beinhaltet, der die gefundenen Effektstärken klassifiziert und bereits erste Bewertungen vornimmt (Brüderl 2004: 86).

Im Vergleich zum narrativen Review sind Meta-Analysen in vielen Belangen deutlich überlegen. So sind die aktuellen meta-analytischen Techniken weniger auf die Signifikanz der Primärstudienresultate fixiert, sondern betreiben eine statistisch abgesicherte Ergebniszusammenfassung und nehmen oftmals eine Erklärung der Ergebnisvariabilität vor. Darüber hinaus ist eine Meta-Analyse intersubjektiv besser nachvollziehbar, da viele Entscheidungen bei der Durchführung, wie die Offenlegung von Literatursuchstrategie und Selektionskriterien, äußerst transparent gehandhabt werden und die statistischen Methoden leicht replizierbar sind (Cheung/Vijayakumar 2016: 122). Ohne einen bestimmten Umfang an herkömmlicher Literatursynthese ist eine Meta-Analyse kaum denkbar. Vor allem der Rückbezug der statistischen Ergebnisse einer Meta-Analyse an den Forschungsstand und das Aufzeigen neuer Forschungspfade gehört zu deren elementaren Bestandteilen (Urban/Fiebig 2015: 181ff.).

Der modellbasierte Ansatz bietet hierzu erhebliche Potenziale, da nicht nur die Erkenntnisse aus der Synthese einer bivariaten Beziehung eine Rolle spielen, sondern ein Geflecht von Beziehungen meta-analytisch überprüft werden. So kann etwa die empirisch zunächst ernüchternde Erkenntnis, dass zur Überprüfung einer bestimmten Teilbeziehung des Erklärungsmodells die entsprechenden empirischen Informationen fehlen, ein fruchtbarer Ansatzpunkt für neue Primärstudien darstellen.

Ein weiterer Vorwurf ist das sogenannte „Apples-and-Oranges“-Problem, also der Vorwurf, dass Äpfel mit Birnen verglichen werden, sprich dass sehr heterogene Studien zur Analyse verwendet und deshalb keine bedeutungsvollen Schlüsse aus einem solchen „Meta-Analyse-Fruchtsalat“ gezogen werden können (Card 2012: 25). Dies stellt allerdings nur ein Problem dar, wenn die Schlüsse, die man ziehen möchte, lediglich auf eine eng abgegrenzte Population von Studien (beispielsweise die Wirkkraft eines bestimmten Medikaments) zutreffen soll. Liegt allerdings ein breites Erkenntnisinteresse vor, wie in dieser Studie, dann ist es sinnvoll, eine breite heterogene Studienlandschaft zu analysieren. Denn Primärstudien sind niemals vollständige Replikationen voneinander und mit dem Einschluss von Studien, die unterschiedlich in Bezug auf Methodologie, verwendete Messinstrumente sowie die Untersuchungspopulation sind, kann die Generalisierbarkeit der meta-analytischen Ergebnisse erhöht werden (Eisend 2014: 70). Die Stärke der Meta-Analyse liegt darin, potenzielle Heterogenität mittels Moderatoranalysen zur statistischen Erklärung der Variabilität von Effektstärken heranzuziehen (Borenstein et al. 2011: 358f). Durch die Mixed-Effects-Modellierung wird das „Problem“ leicht beherrschbar. Lediglich bei der Interpretation der Ergebnisse muss darauf geachtet werden, wie heterogen die zugrunde gelegten Primärstudien sind und welche Schlüsse auf der jeweiligen Datenbasis zulässig und inhaltlich gehaltvoll sind.

Das sogenannte „File-Drawer“-Problem (auch Publikationsbias genannt), also die Möglichkeit, dass die in der Meta-Analyse eingeschlossenen Studien keine repräsentative Stichprobe der durchgeführten Studien ist, muss in jeder Form der Forschungssynthese thematisiert werden (Borenstein et al. 2011: 378f.). Aber nicht nur Literaturreviews haben mit diesem Problem zu kämpfen, sondern jede wissenschaftliche Arbeit kann nur diejenigen Publikationen für die eigene Arbeit berücksichtigen, die auch gefunden werden (Eisend 2014: 72). Dies ist jedoch unbestreitbar eine ernstzunehmende Gefahr für Meta-Analysen, da es dort direkt zu Ergebnisverzerrungen kommen kann. Jedoch gibt es zum einen die Möglichkeit, durch eine entsprechende Literatursuchstrategie die Gefahr zu minimieren, und zum anderen durch statistische Diagnose- und Korrekturmöglichkeiten das Vorliegen eines Bias aufzudecken.

Ein weiterer häufig vorgebrachter Einwand ist das sogenannte „garbage in, garbage out“-Argument, also die Annahme, dass wenn methodologisch schwache bzw. qualitative schlechte Studien<sup>18</sup> in der Meta-Analyse berücksichtigt werden, auch nur qualitativ schwache meta-analytische Ergebnisse erzeugt werden können (Card 2012: 26f.). Allerdings gibt es auch die Möglichkeit, durch die Verwendung von bestimmten statistischen Artefaktkorrekturmethode die Schwächen der Primärstudien auszugleichen (Schmidt/Hunter 2015). Somit ist es möglich, Studienergebnisse, die auf Grundlage von wenig reliablen Messinstrumenten entstanden sind, unter Zuhilfenahme von Reliabilitätskoeffizienten zu korrigieren. Manche Autoren empfehlen einen Ausschluss von Studien mit minderwertiger Qualität als Selektionskriterium, die meisten warnen allerdings davor, da dadurch ebenfalls Ergebnisverzerrungen entstehen können (Borenstein et al. 2011: 380f.).

Neben diesen eher konzeptionellen Kritikpunkten gibt es auch das vorwiegend statistische Problem, dass fehlende, weil in den Primärstudien nicht berichtete, Werte (unter anderem auch Reporting-Bias genannt) die meta-analytischen Ergebnisse beeinflussen (Eisend 2014: 78ff.). Dabei sind unterschiedliche Formen von fehlenden Werten möglich. Es können ganze Studien, Effektstärken oder Studiencharakteristiken fehlen. Die bei fehlenden Daten hierbei am häufigsten empfohlene Strategie ist die Kontaktierung der Primärstudienautoren (Pigott 2009: 400). Nur falls dies keinen Erfolg bringt, sollte man die methodischen Möglichkeiten zur Ersetzung fehlender Werte in Betracht ziehen. Dies sind die Methoden, die aus der herkömmlichen statistischen Analyse bekannt sind, nämlich paarweiser Fallausschluss, listenweiser Fallausschluss, einfache Imputation (Mittelwertersetzung, Regressionsersetzung), multiple Imputation und Verfahren, die Maximum Likelihood-Schätzungen nutzen (Cheung 2015: 290ff.; Enders 2010; Pigott 2009: 403ff.).<sup>19</sup>

Allerdings gibt es konzeptionelle Einwände gegen Datenersetzungsverfahren bei Meta-Analysen (Rosopa/Kim 2017: 218ff.). In der Regel sollen mithilfe einer Meta-Analyse die in der Realität vorliegenden Studien eines Forschungszweigs zusammengefasst und Aussagen über die

---

<sup>18</sup> Um den Begriff der Studienqualität gibt es in der meta-analytischen Forschungspraxis eine lebhafte Diskussion. In der Regel wird dieses Phänomen unter Validitätsaspekten betrachtet (Valentine 2009: 134f.). Unstrittig ist lediglich, dass es keine allgemeingültige Form von Studienqualität gibt. Die Mehrzahl der Autoren empfiehlt deshalb eine Operationalisierung von einzelnen Indikatoren der methodischen Güte wie beispielsweise die Reliabilität von Messinstrumenten etc. Ebenso gibt es bereits umfangreiche Versuche, Qualitätsskalen zu entwickeln – dies gestaltet sich jedoch aufgrund des sehr vielschichtigen Phänomens als sehr schwierig. Andere Autoren (Urban/Fiebig 2015) verwenden eine subjektive Einschätzung der methodischen Güte der Primärstudien als Proxyvariable.

<sup>19</sup> Eine im MASEM-Kontext diskutierte Ersetzungsmöglichkeit besteht darin, die Werte durch „plausible“ Werte, etwa durch Experteneinschätzung, zu ersetzen (Viswesvaran/Ones 1995: 876). Auch können fehlende Werte durch den sehr konservativen Wert „0“ oder durch den Mittelwert ersetzt werden (Eisend 2014: 80). Diese „Verfahren“ sind jedoch ungeeignet, da sie mitunter zu starken Verzerrungen der geschätzten Werte führen (Enders 2010: 42ff.).

tatsächlich durchgeführten Studien getroffen und keine synthetischen Effektstärken von Studien geschätzt werden, die niemals angefertigt wurden. Keines dieser Verfahren wurde speziell für den meta-analytischen Anwendungsfall entwickelt. Die beiden derzeit am besten evaluierten Verfahren, multiple Imputation und Ansätze, die auf dem Maximum Likelihood-Schätzungen basierenden (Enders 2010), liefern nur bei großen Fallzahlen und geringer Missingquote effiziente Ergebnisse. In Meta-Studien liegen meist nur wenige Studien vor, so dass der Anteil fehlender Werte häufig hoch ist und diese Verfahren somit untauglich sind.

Listenweiser Fallausschluss ist aufgrund des großen Datenverlustes in Meta-Analysen ebenfalls ungeeignet, da dadurch die statistische Teststärke stark verringert wird. Viel häufiger eingesetzt wird in Meta-Analysen der paarweise Fallausschluss (Cheung/Chan 2005b: 40ff.; Rosopa/Kim 2017: 218). Dieser führt jedoch bei den meisten modellbasierten Verfahren zu dem Problem, dass die einzelnen Zellen der Korrelationsmatrix auf unterschiedlichen Fallzahlen beruhen, wodurch sich das Folgeproblem ergibt, eine adäquate Fallzahl für die SEM-Analyse zu finden.<sup>20</sup>

Generelle Strategie sollte es für Meta-Analysen deshalb sein, fehlende Werte durch sparsame statistische Modelle, in Bezug auf Studiencharakteristiken, inhaltliche Moderatoren und zu spezifizierende Mediatorbeziehungen, zu vermeiden. Darüber hinaus sollten die angesprochenen Verfahren zur Ersetzung fehlender Werte zurückhaltend verwendet werden, da deren Performanz in Bezug auf Meta-Analysen oftmals ungeklärt ist (Rosopa/Kim 2017: 232). Dies gilt insbesondere für die modellbasierten meta-analytischen Verfahren, die auf Korrelationsmatrizen beruhen.

Eine der größten Herausforderungen der MASEM-Verfahren besteht darin, eine angemessene Strategie zum Umgang mit fehlenden Werten zu finden (Eisend 2017: 11), denn Beziehungen, die in keiner Primärstudie getestet wurden, können auch meta-analytisch nicht überprüft werden (Becker 2009: 384). Dieses Problem ergibt sich auch, wenn einige Stellen des analytischen Modells bzw. Zellen der kumulierten Korrelationsmatrix unbesetzt bleiben. Dieses spezifische MASEM-Problem ist bei der herkömmlichen Meta-Analyse kaum relevant und es gibt deshalb

---

<sup>20</sup> Ein weiteres Problemfeld bei fehlenden Werten in Meta-Analysen liegt unter anderem darin begründet, dass diese auf zwei unterschiedlichen Ebenen auftreten können. Zum einen sind fehlende Werte innerhalb der Primärstudien auf Individualebene (bspw. Religiositätsmessung fehlt) und zum anderen auf Studienebene (bspw. Mediatorbeziehung wurde nicht spezifiziert) möglich (Rosopa/Kim 2017: 232). Somit sind prinzipiell auch die Ersetzungsverfahren auf zwei Ebenen möglich. Eine solche „Doppelimputation“ kann natürlich auch unwissentlich vorgenommen werden, wenn die entsprechenden Ersetzungsverfahren in den Primärstudien nicht dokumentiert werden – was immer noch weitverbreitete Praxis ist. Ebenfalls muss bedacht werden, dass den multiplen Imputationsverfahren bestimmte Annahmen zugrunde liegen (Enders 2010). Werden diese Annahmen verletzt, müssen die Ergebnisse dieser Verfahren in Frage gestellt werden (Matt/Cook 2009: 544).

auch keine Lösungsstrategien, etwa für die Literatursuchstrategie, um alle Beziehungspaare eines komplexen meta-analytischen Modells mit ausreichend empirischem Material zu füllen.

Im Gegensatz zur Möglichkeit, indirekte Beziehungen (Mediatoren) im Modell berücksichtigen zu können, haben die meisten MASEM-Verfahren generell Probleme mit dem Test von Moderationshypothesen (Bergh et al. 2014: 479). Die Verfahren bleiben, mit Ausnahme des neu entwickelten OSMASEM-Ansatzes, zumeist beschränkt auf den Test von kategorialen Moderatoren (Männer vs. Frauen) mithilfe einer Subgruppenanalyse (Sheng et al. 2016: 190). Der Test einer Vielzahl von kontinuierlichen und kategorialen Moderatoren, wie sie in einem meta-analytischen Mixed-Effects-Modell üblich sind<sup>21</sup>, können in einer MASEM-Analyse kaum durchgeführt werden, da dort die notwendigen empirischen Informationen (Korrelation mit den Produkttermen) in der Regel fehlen.

Auch eine modellbasierte MASEM-Studie kann kausale Aussagen<sup>22</sup> nur dann treffen, wenn experimentelle Studien vorliegen (Bergh et al. 2014: 479). Allerdings sind die Aussagen einer MASEM-Modellierung insofern gegenüber der univariaten Meta-Analyse überlegen, da nach weiteren Variablen kontrolliert werden kann. Aufgrund der Möglichkeit, die Modellpassung zu prüfen, kann herausgefunden werden, welches Modell am besten zur Beschreibung der Realität geeignet ist. Somit können Hinweise für zukünftige Forschungsvorhaben mit experimentellem Charakter generiert werden.

Eine weitere Einschränkung des MASEM-Ansatzes besteht darin, dass alle statistischen Verfahren, die Rohdaten benötigen, nicht durchgeführt werden können, da MASEM lediglich auf den (Kovarianz- bzw.) Korrelationsmatrizen beruht. Dazu gehören etwa ML-Schätzungen bei fehlenden Werten auf Ebene der Subjekte, die Analyse von binären oder kategorialen Daten, robuste Teststatistiken (und Standardfehler), Mixture-Modelling auf Subjektebene und die Modellierung von nicht linearen Beziehungen (Cheung 2015: 217).

Auch eine MASEM-Studie ist vor Konstruktvaliditätsproblemen in den Primärstudien nicht sicher. Die Qualität der MASEM-Ergebnisse kann deshalb nur so gut sein wie die methodologische Qualität der Primärstudien (Bergh et al. 2014: 479). Auch wenn ein Modell eine gute Passung auf die kumulierte (durchschnittliche) Korrelationsmatrix aufweist (einen guten Modellfit hat), bedeutet dies nicht, dass das Modell für jede (Teil-) Population passend ist (Becker 2009: 384). Deshalb sollte fast immer eine Random-Effects-Modellierung genutzt werden.

---

<sup>21</sup> Diese oftmals atheoretische Praxis bei der Durchführung von Moderatoranalysen kann jedoch im Hinblick auf die Teststärke auch kritisch betrachtet werden.

<sup>22</sup> Zur Erläuterung des Problems der kausalen Aussagen auf Basis von nicht-experimentellen Arbeiten in den Sozialwissenschaften siehe etwa Antonakis et al. (2010).

Der größte Nachteil der (modellbasierten) MASEM-Verfahren liegt aber für viele praxisorientierte Anwender in der vergleichsweise komplizierteren Umsetzung im Vergleich zu den klassischen meta-analytischen Techniken (Becker 2009: 379; Cheung/Hafdahl 2016: 116). Der MASEM-Anwender muss ein Grundverständnis für die Techniken der Strukturgleichungsmodellierung aufweisen (Card 2012: 287).<sup>23</sup>

Es gibt somit einige Schwierigkeiten bei der Durchführung (modellbasierter) Meta-Analysen. Diese sind unabhängig von der konkreten statistischen Modellierungsart und vorwiegend von konzeptioneller Art.<sup>24</sup> Nach einhelliger Meinung der Forschungsgemeinschaft stellen die hier präsentierten Einwände jedoch keinen prinzipiellen Grund zur Ablehnung der Methode an sich dar (Kopp/Lois 2014: 210). Neben der entsprechenden Planung muss die Interpretation der Ergebnisse im Hinblick auf die Datengrundlage zurückhaltend vorgenommen und es dürfen keine Schlüsse gezogen werden, die durch die Grundgesamtheit der Studien und der zugrundeliegenden Modellierungsform nicht gedeckt werden.

---

<sup>23</sup> Allerdings gibt es inzwischen Einführungsliteratur, die die Konzepte und Prinzipien der SEM-Analyse anwendungsorientiert vermitteln (Little/Kline 2016; Urban/Mayerl 2014). Ebenfalls gibt es einführende Arbeiten von Becker (2009); Cheung (2015); Jak (2015), die die statistischen Aspekte von MASEM diskutieren und die Umsetzung etwa mit dem R-Paket metaSEM (Cheung 2014b) erheblich erleichtern.

<sup>24</sup> Die spezifischen (statistischen) Probleme der modellbasierten Ansätze, wie etwa das Problem der nicht positiv definiten Korrelationsmatrix, wird bei der Erläuterung dieser Ansätze in den nachfolgenden Kapiteln noch ausführlich behandelt.

### 3 Forschungsstand und konzeptionelle Überlegungen

Jede MASEM-Studie muss eine klare Vorstellung von den interessierenden Konstrukten und deren angenommener Beziehungsstruktur haben (Landis 2013: 255). Eine der ersten Entscheidungen bei der Durchführung einer Meta-Analyse stellt die Problemdefinition dar, also die Definition der Kernkonzepte und deren Beziehungen zueinander (Cooper 2010: 20ff.). Dabei ergeben sich wesentliche Auswirkungen auf die Frage, welche statistischen Methoden für die Untersuchung angemessen sind und worin die Grundgesamtheit der Meta-Analyse besteht (Hedges 2009: 38).

Deshalb folgt zunächst die inhaltliche Klärung des Gegenstands, da sich das nachfolgende Forschungsdesign auf die konzeptionellen Festlegungen bezieht. Dazu werden beide sozialen Konstrukte nacheinander konzeptionell definiert und klar von angrenzenden Phänomenen abgegrenzt.<sup>25</sup> Im Anschluss an die Konzeptdefinitionen wird der Zusammenhang zwischen den beiden latenten Konstrukten empirisch-analytisch hergeleitet sowie die diskutierten Erklärungsmechanismen systematisiert, um daran anknüpfend ein psychosoziales Erklärungsmodell zu konstruieren. Im vorliegenden Forschungskontext ist es von besonderer Bedeutung, dass die beiden Konstrukte überschneidungsfrei definiert werden, da ansonsten tautologische Beziehungen, die möglicherweise zu verzerrten Ergebnissen führen, entstehen können. Konzeptionelle Ungenauigkeiten haben in der Vergangenheit bereits zu einer Vielzahl von tautologischen Befunden in diesem Forschungsfeld geführt (Koenig 2008: 352; Wong/Rew/Slaikau 2006: 177).<sup>26</sup> Auch eine zu breite Konzeptualisierung muss vermieden werden, um überhaupt Erklärungsmechanismen spezifizieren zu können.

Die theoretisch-analytischen Überlegungen sind bei allen Formen der Meta-Analyse von zentraler Bedeutung. Dies gilt jedoch besonders für modellbasierte meta-analytische Verfahren, da diese ganze Modelle testen. Die Besonderheit der vorliegenden Arbeit liegt jedoch darin, dass zunächst offenbleibt, welche Erklärungsmechanismen (Mediatoren) berücksichtigt werden. Zunächst werden die Definitionen der bisherigen Meta-Analysen, aber auch anderer konzeptionelle Arbeiten, betrachtet und eine eigene Definition für die beiden Untersuchungskonzepte abgeleitet. Durch diese konzeptionelle Festlegung wird gleichzeitig auch die Grundgesamtheit der Meta-Analyse bestimmt, da in die Analyse nur Studien eingeschlossen werden können, die dieser Konzeptionierung entsprechen.

---

<sup>25</sup> Eine operationale Definition (Cooper 2010: 33ff.) folgt in Kapitel 4 und 6 durch die Einschränkung der zu berücksichtigenden Messinstrumente.

<sup>26</sup> Für eine Übersicht von Fehlschlüssen aufgrund sich überschneidender Konzeptualisierungen in dem Forschungsbereich siehe Koenig/King/Carson (2012: 42).

### 3.1 Religiosität und Spiritualität

Das erste Konstrukt stellt die unabhängige Variable Religiosität/Spiritualität (im Folgenden auch: R/S) dar. Um die Definition der beiden genannten Konzepte und deren Beziehung zueinander gibt es einen hohen Grad an Verwirrung (Dehaan/Yonker/Affholter 2011: 190ff.; Hill 2013: 49; Hofmann 2009: 31ff.; Ivtzan et al. 2013: 916; Pickel 2011: 16ff.). Bereits die Etymologie der Begriffe ist vielschichtig und nicht eindeutig zu klären und bezüglich der Definition gibt es keinen Konsens (Utsch/Klein 2011: 27ff.). Es liegt eine große Heterogenität bei der Definition und Messung der beiden Konzepte vor (Albani et al. 2002: 307; Mytko/Knight 1999: 440; Tiliouine 2009: 92). Dies führte bereits in der Vergangenheit zu einer Variation der Studienergebnisse, wie auch verschiedene Meta-Studien zeigen konnten (Hackney/Sanders 2003: 49ff.; Witter et al. 1985: 177; Yonker/Schnabelrauch/Dehaan 2012: 305).

Bei der Konzeptionierung ist die Balance zwischen einer zu breiten und einer zu engen Konzeptualisierung entscheidend. Es muss vermieden werden, jegliche Glaubensüberzeugungen und -aktivitäten, die mit einer bestimmten Bedeutung versehen sind, als religiös zu definieren, da ansonsten fast alle menschlichen Phänomene religiös bzw. spirituell deutbar sind (Hill et al. 2000: 71; Koenig 2008: 353). Religion ist ein komplexes Phänomen, das sowohl aus individuellen Prozessen als auch aus dem sozialen Kontext heraus analysiert werden kann und oftmals als universelles menschliches Phänomen bezeichnet wird (Hayward/Krause 2014; Hill et al. 2000: 53ff.; Huber 2003: 99f.).<sup>27</sup> Mit dem Begriff Religion werden mehr oder weniger stark organisierte institutionelle soziale Gebilde beschrieben. Religiosität hingegen bezieht sich explizit auf die individuelle Seite von Religion und ist für diese Untersuchung das maßgebliche Konzept (Utsch/Klein 2011: 27f.).

Eine wichtige Entscheidung betrifft die Abgrenzung von Religiosität zu Spiritualität.<sup>28</sup> Einige Autoren gehen davon aus, dass es unterscheidbare Konzepte sind (Worthington/Sandage 2001: 473; Zinnbauer et al. 1997: 561ff.). Andere Autoren hingegen sind davon überzeugt, dass die Konzepte inhaltlich sehr stark miteinander verknüpft und kaum getrennt erfassbar sind (Hofmann 2009: 48; Koenig/King/Carson 2012: 50). Hinzu kommt die empirische Gegebenheit, dass die meisten Menschen spirituelle Erfahrungen in einem institutionellen religiösen Kontext

---

<sup>27</sup> Manche Autoren weisen darauf hin, dass die Konzepte Religiosität und Spiritualität jeweils im westlichen (christlichen) Kulturkontext entstanden sind und deshalb nicht kulturübergreifend verwendet werden sollen (Oman 2013: 42). Allerdings erscheint diese Deutungsweise radikal, da es sich augenscheinlich um ein universales Phänomen handelt und in Ermangelung besserer Begrifflichkeiten sollte an den Konzepten festgehalten werden.

<sup>28</sup> Die großen Diskussionsstränge zur Frage der Überschneidung oder Abgrenzung können bei Utsch/Klein (2011) nachgelesen werden.

machen sowie die beiden Konzepte ohnehin nicht voneinander unterscheiden und eine analytisch konstruierte Trennung deshalb überflüssig erscheint (Hill/Pargament 2003: 65; Ostermann/Büssing/Matthiessen 2004: 352).

Für die vorliegende Arbeit folge ich den Argumenten derjenigen Autoren, die Spiritualität als allgemeineres grundlegenderes Konzept und Religiosität als spezifischeres auf konkrete religiöse Traditionen bezogenes Konzept verstehen. Bei Religiosität und Spiritualität handelt es sich dieser Auffassung nach um eine Art von individueller Werteinstellung (Huber 2003: 34ff.). Im Englischen werden die Begriffe inzwischen nahezu synonym verwendet. In vielen Ländern wie etwa Deutschland durchlaufen immer weniger Menschen eine traditionelle religiöse Sozialisation und reagieren gegenüber den Begriffen Religion und Religiosität eher negativ. Diese Menschen („spirituell aber nicht religiös“) suchen aber dennoch nach Antworten auf die grundlegenden menschlichen Erfahrungen wie Leid, Krankheit und Tod (Sergent 2015: 4; Utsch/Klein 2011: 35).

Nun folgt eine Übersicht über die definatorische Praxis in den vorliegenden Meta-Studien. Witter et al. (1985) verzichten vollständig auf eine konkrete Definition von Religiosität. Es wird lediglich deutlich, dass sie darunter ein multidimensionales Konstrukt verstehen, das sowohl religiöse Aktivität, aber auch Frömmigkeit einschließt (Witter et al. 1985: 334). Sawatzky/Ratner/Chiu (2005) hingegen legen eine explizite konzeptionelle Definitionen vor. Sie definieren Spiritualität als die auf das Transzendente bezogene existenzielle Suche nach Sinn und Bedeutung. Diese ist durch den subjektiven Charakter abseits von vorgegebenen Glaubensvorstellungen gekennzeichnet und somit von traditioneller institutionalisierter Religiosität abgrenzbar (Sawatzky/Ratner/Chiu 2005: 156ff.). Allerdings schließen sie die sogenannten existenziellen Skalen, die nicht auf das Transzendente bezogen sind und Wohlbefindensaspekte beinhalten, aus. In der Studie von Yonker/Schnabelrauch/Dehaan (2012) wird Religiosität und Spiritualität nach einem systematischen Review gemeinsam definiert, als aktive persönliche Hingabe und leidenschaftliche Suche, die größtenteils innerhalb eines selbst gesetzten Rahmens einer theologischen Gemeinschaft stattfindet (Yonker/Schnabelrauch/Dehaan 2012: 300). In der Studie von Eger/Maridal (2015) hingegen ist R/S nur eine von vielen Prädiktoren von Wohlbefinden und wird konzeptionell nicht bestimmt.

Bei den vorwiegend gesundheitswissenschaftlichen Meta-Analysen, die die Beziehung von Religion und Gesundheit untersuchen, ist die Konzeptionierung unpräziser. Bergin (1983), Donahue (1985), Baier/Wright (2001), Mahoney et al. (2001), Saroglou (2002), Saroglou/Delpierre/Dernelle (2004), Wu/Wang/Jia (2015) und Smith/McCullough/Poll (2003) etwa

verzichten auf eine explizite Definition von Religiosität/Spiritualität. Zuckerman/Silberman/Hall (2013) konzeptualisieren Religiosität, indem sie die Ausprägung in bestimmten Facetten von Religion (Glaube an übernatürliche Mächte, gemeinschaftliche Rituale, etc.) betrachten. Salsman/Fitchett et al. (2015) konzeptualisieren R/S abstrakt als eine Konstellation von wichtigen affektiven, verhaltensbezogenen und kognitiven Variablen, welche die Gefühle, Handlungen und Gedanken von Menschen in Bezug auf ihre Gesundheit beeinflussen können. Ähnlich ist es bei Hackney/Sanders (2003), die in Bezug auf Religion von kognitiven, emotionalen und verhaltensbezogenen Aspekten ausgehen. Allerdings beschränken sie sich auf die enge Fassung von Religiosität und schließen eine Reihe von Konstrukten wie etwa Spiritualität, Mystizismus, religiöses Coping und transzendente Erfahrungen aus, weil sie diese Aspekte für konzeptionell verschieden halten und für die eher institutionalisierte Form der Suche nach dem Transzendenten nicht relevant sind (Hackney/Sanders 2003: 46).

Es zeigt sich, dass innerhalb der vorhandenen Studien eine große Heterogenität hinsichtlich Qualität und Umfang der Definitionen vorliegt. Ausgehend von diesen konzeptionellen Definitionen wird nun eine eigene Konzeption entwickelt, die für die vorliegende modellbasierte Meta-Analyse anwendbar ist. Diese Bestimmung ist weder rein substanziell noch funktional und somit anschlussfähig an die sozial- und gesundheitswissenschaftlichen Definitionen (Heil 2010: 7; Koenig/Wolf 2013: 6ff.; Pickel 2011: 19ff.; Pickel/Sammet 2014: 16ff.; Pollack 2014: 22f.; Stolz/Buchard 2014: 57ff.).<sup>29</sup>

Spiritualität ist als Konzept individualistisch angelegt. Das Konzept wird als Oberbegriff für eine Vielzahl von religiös-spirituellen Einstellungen verwendet, von der herkömmlichen Kirchenmitgliedschaft bis zum Sammelbegriff „gläubig aber nicht religiös“, wobei sich inzwischen zentrale Gemeinsamkeiten bei der Definition des Begriffs zeigen (Chiu et al. 2004: 409ff.; Frick et al. 2016: 449; Knoblauch 2005: 129; Koenig/King/Carson 2012: 36).

Das ursprünglich christliche Konzept Spiritualität (Bucher 2014: 11ff.; Koenig 2008: 349) wird heute definiert als „the feeling thoughts experiences and behaviors that arise from a search for the sacred. The term „search“ refers to attempts to identify, articulate, maintain, or transform. The term „sacred“ refers to a divine being, divine object, Ultimate Reality, or Ultimate Truth as perceived by the individual.“ (Hill et al. 2000: 66). Spiritualität meint die Suche nach der letzten Wirklichkeit bzw. der Bedeutung des menschlichen Lebens durch Gefühle der Verbun-

---

<sup>29</sup> Einen großen systematischen Überblick über die Vielzahl von unterschiedlichen bereits vorgebrachten Definitionen aus dem Bereich der Gesundheitswissenschaften kann an anderer Stelle ausführlich nachgelesen werden (Heil 2010: 7ff.; Hofmann 2009: 31ff.; Körtner 2009: 7ff.; Oman 2013: 23ff.).

denheit mit sich selbst, anderen und einem höheren Wesen sowie der Selbsttranszendenz (Bucher 2014: 40; Chiu et al. 2004: 423f.; McClain/Rosenfeld/Breitbart 2003: 1603). Dabei stehen der Suchcharakter sowie der Bezug auf das Transzendente im Fokus (Koenig/King/Carson 2012: 46; Oman 2013: 39). Anders ausgedrückt ist damit die Erkenntnis bzw. der Glaube verbunden, dass es jenseits der sinnlich erfahrbaren Welt noch etwas gibt.<sup>30</sup>

Einigkeit besteht vor allem darin, dass Religiosität im Gegensatz zu Spiritualität einen institutionellen Aspekt beinhaltet, der sich vorwiegend durch vorgegebene religiöse Lehren auszeichnet, aber nicht auf die formelle Mitgliedschaft reduziert werden kann (Dezutter/Soenens/Hutsebaut 2006: 808; Tay et al. 2014). Eine Religion ist es nach Hill et al. (2000) in Anlehnung an ihre Spiritualitätsdefinition wenn „The means and methods [...] of the search [...] receive validation and support from within an identifiable group of people.“ (Hill et al. 2000: 66).

Der Wesenskern von Religiosität liegt also vor allem darin, dass soziale Faktoren bei der Suche nach dem Heiligen relevant sind. Eine eher populärwissenschaftliche Definition, die beide Konzepte einschließt, besagt: „Religiosität ist eine Fähigkeit oder Eigenschaft: das mehr oder weniger ausgeprägte Persönlichkeitsmerkmal, eine Religion im weitesten Sinn zu haben und sich so auf Transzendentes zu beziehen - im Erleben, Denken, Fühlen und Handeln.“ (Vaas/Blume 2009: 20). Die Leistungen oder Funktionen liegen etwa darin, dass sie Menschen eine Vision davon geben, wonach sie im Leben streben (Emmons/Cheung/Tehrani 1998: 403f.), in der Transzendierung der biologischen Natur des Menschen (Pickel/Sammet 2014: 17), dass sie das menschliche Bedürfnis nach „Kosmisierung“ befriedigen (Meulemann 2007: 275), dass sie eine Lösung für das Bewährungsproblem liefern (Pickel 2011: 245ff.) oder in der generellen Bewältigung von Kontingenz (Luhmann 2002: 115ff.). Zusammenfassend kann man Religiosität verstehen als „ein System von Praktiken, das sich speziell mit der Abwehr und Bewältigung krisenhafter Situationen befasst. Religion macht nicht nur das Unerklärliche erklärbar, sondern erhält auch die Handlungsfähigkeit des Menschen in den Situationen, in denen er an seine Grenzen kommt“ (Riesebrodt 2007: 243).

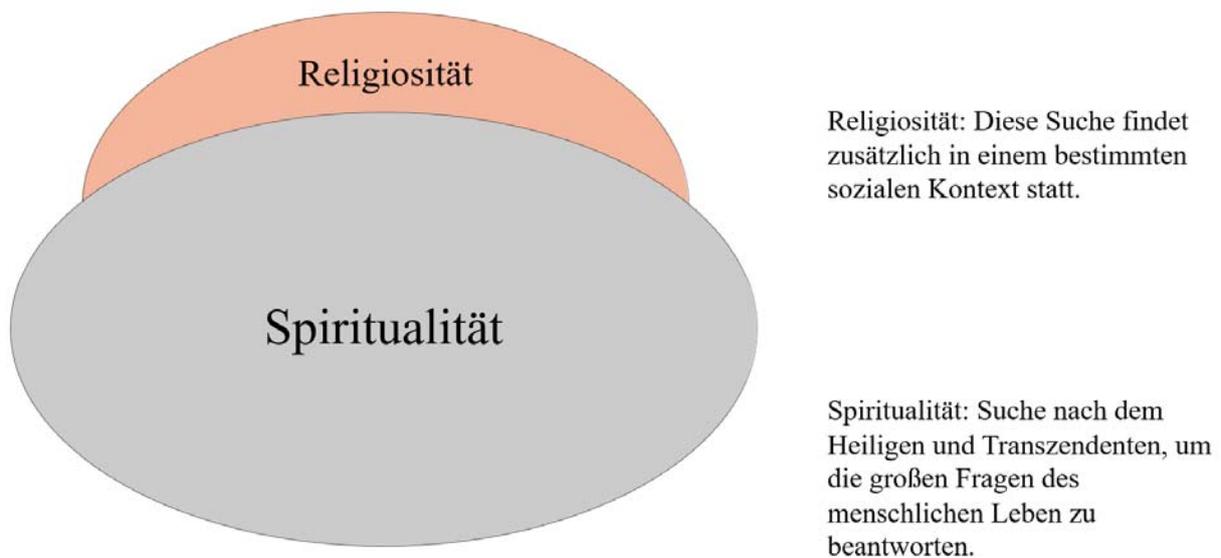
Es handelt sich bei beiden Konzepten um eine Art von Werteinstellung, wobei die Spiritualität den Kern der Religiosität betrifft. Bei Spiritualität geht es vorwiegend um die persönliche Suche nach dem Heiligen und der Transzendenz, also um die großen menschlichen Fragen nach Sinn- und Identität im scheinbar unendlichen Universum. Religiosität hingegen besteht generell nicht nur aus Überzeugungen, Praktiken und Ritualen, die auf das Heilige und Transzendente

---

<sup>30</sup> Viele Wissenschaftler gehen davon aus, dass es sich bei dieser Überzeugung um ein nahezu universelles menschliches Phänomen handelt (Baier 2006: 10).

bezogen sind, sondern hat darüber hinaus auch noch einen bewussten sozialen, gesellschaftlichen und manchmal institutionellen (kirchlichen) traditionellen Bezug (Hill/Pargament 2003: 65; Koenig/King/Carson 2012: 42; Pargament 2002: 243). Diese Konzeptualisierung ist zum besseren Verständnis nochmals in Abbildung 4 visualisiert. Dort sind der große Überschneidungsbereich sowie die verbleibende Spezifität von Religiosität erkennbar.

**Abbildung 4: Definition von Religiosität und Spiritualität**



**Anmerkung:** Eigene Abbildung.

Diese bewusst breite Definition eignet sich für das vorliegende Forschungsvorhaben besonders gut, da es eine Vielzahl von individuellen religiösen und spirituellen Überzeugen, Traditionen und Kulturen einschließt, aber dennoch nicht zur Verzerrung der Forschungsergebnisse beiträgt. Der Grundgedanke bei dieser Konzeptionierung liegt darin, dass alle religiöse Vorstellungen einen gemeinsamen Kern haben. Der gemeinsame Kern bezieht sich auf die Suche nach dem Transzendenten, um die oben genannten Anforderungen (Funktionen) zu erfüllen.

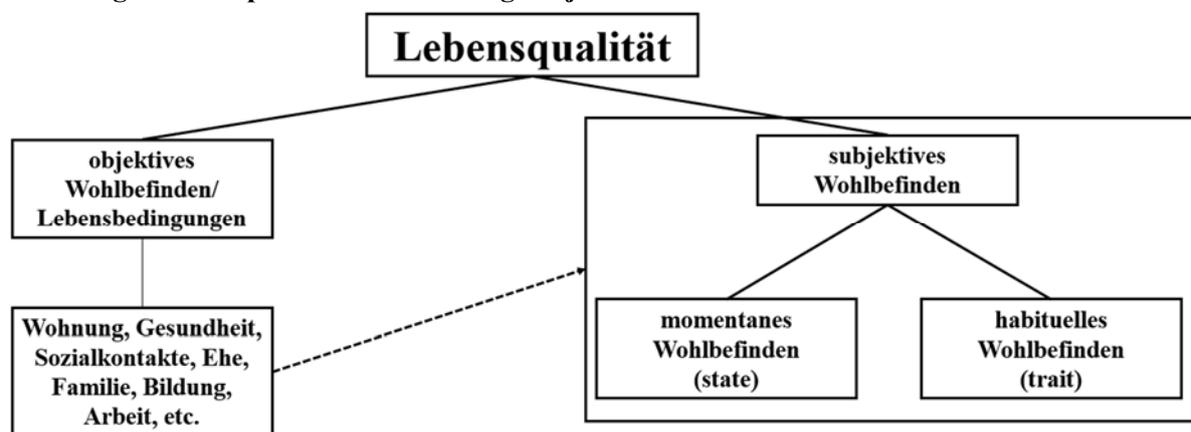
Eine zu starke Ausdehnung des Spiritualitätsbegriffs, wie beispielsweise beim Einschluss von spirituellem Wohlbefinden (Frey/Daaleman/Peyton 2005), würde zu tautologischen Erklärungen führen, da in einer solchen Konzeption bereits positive psychische Befindlichkeiten wie etwa Ausgeglichenheit oder Lebenssinn eingeschlossen sind (Garssen/Visser/Jager Meezenbroek 2016: 145f.; Klein/Albani 2011b: 218). Konzeptualisierungen, die solche Aspekte berücksichtigen, führen zu tautologischen Schlüssen, da Teile der abhängigen Variable bereits in der unabhängigen enthalten sind und dort keine sinnvollen Ursache-Wirkungsketten mehr festgelegt werden können (Bai/Lazenby 2015; Jager Meezenbroek et al. 2012: 350f.; Koenig/King/Carson 2012: 42ff.). Ausgeschlossen werden müssen deshalb diejenigen Konzepte, die zwar thematisch ähnlich sind, aber nicht auf die Transzendenz oder das Heilige bezogen

sind. Darunter fallen positive psychologische Zustände, wie etwa „meaning and purpose“, also das Empfinden von Lebenssinn oder andere unspezifische persönliche Einstellungen und Überzeugungen, die eher die Erklärungsmechanismen des Zusammenhangs darstellen, aber dennoch bereits in vielen Messinstrumenten zur Erfassung von Spiritualität enthalten sind (Monod et al. 2011). Empirische Studien, die ausschließlich solche Konstrukte zur Operationalisierung von Religiosität/Spiritualität verwenden, müssen konsequent aus der vorliegenden Meta-Analyse ausgeschlossen werden, um verunreinigte Ergebnisse zu vermeiden.

### 3.2 Lebenszufriedenheit

Im nächsten Schritt wird die abhängige Variable Lebenszufriedenheit als Teil des subjektiven Wohlbefindens (SWB) konzeptualisiert (Maggino 2014). Das übergeordnete Konzept Lebensqualität wird meist unterteilt in eine subjektive und objektive Komponente (Berth/Klein/Balck 2011: 47ff.; Diener et al. 2009: 9ff.).<sup>31</sup> Diese beiden Konzepte werden häufig synonym verwendet, wobei Wohlbefinden das subjektive Erleben des Individuums in den Vordergrund stellt. SWB ist dabei wie R/S jedem Individuum nur introspektiv über die eigene Erfahrung zugänglich, während objektive Konzeptualisierungen von einem externen Beobachter beurteilt werden können (MacLeod 2015: 1075; White/Blackmore 2016: 6f.). In Abbildung 5 ist eine Zusammenfassung und Einordnung dieser Konzeptionierung zu sehen.

**Abbildung 5: Konzeptionelle Einordnung subjektives Wohlbefinden**



**Anmerkung:** Eigene Abbildung aus einer Synthese der Darstellungen von Heidl et al. (2012) und Dette (2005).

Subjektives Wohlbefinden ist Teil der Lebensqualität, welche neben der subjektiven Komponente eine objektive Seite beinhaltet. Die objektive Seite ist auch durch einen externen Beobachter einschätzbar und besteht aus den unterschiedlichen Bereichen des individuellen Lebens (etwa Bewertung der Wohnsituation oder der körperlichen und geistigen Gesundheit). Die eigene Einschätzung über die Zufriedenheiten in diesen Bereichen werden als Bereichszufriedenheiten bezeichnet (Dette 2005: 42ff.).

<sup>31</sup> Teilweise werden auch komplexere Konzeptualisierungen diskutiert (Kelley-Gillespie 2009).

Beide Ansätze sind nicht frei von Problemen. Die subjektive Bewertung kann im Sinne der sozialen Erwünschtheit vom befragten Subjekt angepasst werden. Beim objektiven Ansatz ist vorwiegend die Auswahl der konkreten Dimensionen und die anschließende Indikatorenwahl problematisch (Argyle 2002: 19; Diener et al. 2009: 23ff.). Da neben der Auswahl der relevanten objektiven Lebensbedingungen auch deren Gewichtung und Bewertung individuell sehr unterschiedlich ausfallen kann und darüber hinaus hedonistische Adaption an die objektiven Lebensbedingungen eine starke Rolle spielt (Frederick/Loewenstein 2003; Gilboa/Schmeidler 2001), ist die Konzeptualisierung von Lebensqualität auf Basis des subjektiven Wohlbefindens für die vorliegende Arbeit am besten geeignet (Moons/Budts/Geest 2006: 899). In den Humanwissenschaften hat sich die (subjektive) Selbstbewertung zur Messung der Lebensqualität inzwischen als Goldstandard durchgesetzt (Cohen/Vengassery/Garcia Aracena 2017). Wie bereits in Abbildung 5 zu sehen ist, stellen die objektiven Lebensbedingungen wichtige Prädiktoren für das subjektive Wohlbefinden dar (Heidl et al. 2012: 5f.; Schimack 2008: 117f.), auch wenn die Korrelationen eher schwach (0.1 – 0.3) sind (Okun/Stock 1987: 109).<sup>32</sup>

Ebenfalls hat sich empirisch gezeigt, dass objektive Veränderungen der eigenen Lebenssituation (Todesfall, Arbeitslosigkeit, Trennung etc.) nicht zu einer (langfristigen) Veränderung des subjektiven Wohlbefindens führen (Berth/Klein/Balck 2011: 58). Körperliche und geistige Gesundheit eines Individuums werden als wichtigste Prädiktoren bzw. kovariierende Faktoren von Wohlbefinden betrachtet (Pavot/Diener 2008). Subjektives Wohlbefinden wird als Prädiktor von Gesundheit und einem langen Leben aufgefasst und deshalb kausal vorgelagert konzeptualisiert (Diener et al. 2017: 94).

Gegenwärtig liegt keine einheitliche Definition von subjektivem Wohlbefinden vor, was zu einer Vielzahl von unterschiedlichen Operationalisierungen führt (Charlemagne-Badal et al. 2015: 321). Viele Autoren verwenden Wohlbefinden, je nach Forschungsbereich, synonym zu den Begriffen Glück, Happiness, Lebensqualität, Lebenszufriedenheit, positiver/negativer Affekt, Wohlergehen („welfare“), Wellness, psychologischem Wohlbefinden, Utility („Nutzen“) und auch Gesundheit (Charlemagne-Badal et al. 2015: 306ff.; Dette 2005: 31ff.; Diener/Ryan 2009: 391; Glaesmer et al. 2011: 127; Veenhoven 2010: 1). Vor allem die synonyme Verwendung von Gesundheit und Wohlbefinden scheint problematisch, weil es sich auf den ersten

---

<sup>32</sup> Die wichtigsten Prädiktoren für das subjektive Wohlbefinden sind Alter, Bildung, finanzieller Status, Arbeit, Geschlecht, Ethnie, Familienstand, soziale Unterstützung, psychologische Faktoren (Optimismus, Hoffnung, etc.), Kontrollüberzeugung, Aktivität, Gesundheit, Genetik und Persönlichkeitsfaktoren (Koenig/King/Carson 2012: 124ff.; Neville/Alpass 1999: 361ff.). Keine der sozio-demographischen Variablen hat dabei jedoch mehr als 10 % Varianzaufklärungsleistung (Andrews/Withey 1976: 247ff.).

Blick um deutlich unterscheidbare Konzepte handelt.<sup>33</sup> Die Messinstrumente für Teile des subjektiven Wohlbefindens (bspw. „Satisfaction with Life Scale“) stellen jedoch messtheoretisch eine sehr gute und vergleichsweise einfach zu erhebende Proxymessung für Gesundheit dar (Charlemagne-Badal et al. 2015; Pavot/Diener 2008). Dies wird sich bei der nachfolgenden Entwicklung des Untersuchungsmodells noch als bedeutsam erweisen, da bislang eher Erklärungsmodelle aus den Humanwissenschaften vorliegen, die Gesundheit als abhängige Variable nutzen.

Zunächst werden die Definitionen der gefundenen Meta-Studien betrachtet und mit weiteren konzeptionellen Arbeiten ergänzt. Wie zuvor bei Religiosität liegen in den meta-analytischen Arbeiten oftmals, wie etwa bei Witter et al. (1985) und DeNeve/Cooper (1998), keine expliziten konzeptionellen Definitionen für subjektives Wohlbefinden vor. Lediglich die operationalen Definitionen geben Auskunft darüber, welche Konzeptualisierungen den Arbeiten zugrunde liegen. Witter et al. (1985) halten fest, dass sie spezifische Skalen zur Messung von Glück, Lebenszufriedenheit und zur geistig-seelischen Verfassung berücksichtigen und domainspezifisches Wohlbefinden und mentale Gesundheit ausschließen (Witter et al. 1985: 334f.). Ähnlich ist die Konzeptionierung in der Meta-Studie von Okun/Olding/Cohn (1990), die SWB als übergeordnetes Konstrukt konzeptualisieren, welches die affektiven Reaktionen eines Individuums in Bezug auf seine Lebenserfahrungen auf einem positiv-negativ-Kontinuum abbildet. Capaldi/Dopko/Zelenski (2014) unterscheiden die beiden großen philosophischen Positionen zum Konzept Wohlbefinden: auf der einen Seite das eudaimonische und auf der anderen Seite die hedonistische Sichtweise.<sup>34</sup> Sie nutzen SWB als zweidimensionales Konstrukt mit der traditionellen Unterscheidung in eine affektive und eine kognitive Komponente.

Sawatzky/Ratner/Chiu (2005) verwenden für ihre Studie das Konzept Lebensqualität. Dieses definieren sie als sehr breites subjektives Phänomen, das nahezu alle Aspekte des menschlichen Lebens einschließt und die persönlichen Perspektiven, Vorstellungen und Erfahrungen eines Menschen in Bezug auf verschiedene Aspekte der Lebensqualität (soziales Wohlbefinden, psychologisches Wohlbefinden etc.) beinhaltet. Diese stellen jedoch keine eigenständigen Dimensionen dar, sondern entsprechen einem unidimensionalen Konstrukt, das synonym mit weiteren

---

<sup>33</sup> Wenn man allerdings die moderne Konzeptionierung von Gesundheit durch die WHO betrachtet, die im Gegensatz zu früheren Defizitmodellen und deren Fixierung auf die binäre Unterscheidung Krankheit/Gesundheit heutzutage das allumfassende körperliche, seelische sowie soziale Wohlbefinden in den Fokus stellt, wird deutlich, warum beide Konzepte synonym verwendet werden (Heidl et al. 2012: 7; WHO o. J.).

<sup>34</sup> Bei der eudaimonischen Sichtweise steht die Ausschöpfung des individuellen menschlichen Potenzials im Vordergrund, während die hedonistische Sichtweise, in der Tradition der positiven Psychologie, die Maximierung des Vergnügens und der Freude sowie die Erfüllung der Bedürfnisse und Wünsche, die zur individuellen Zufriedenheit führen, in den Mittelpunkt stellt (Capaldi/Dopko/Zelenski 2014: 3; Haybron 2008: 22ff.; MacLeod 2015: 1073ff.; Ryff/Boylan 2014: 54ff.).

subjektiven Konzepten wie Lebenszufriedenheit und Wohlbefinden zu sehen ist (Sawatzky/Ratner/Chiu 2005: 158f.). Eine auf die Messinstrumente bezogene Konzeption verwenden Yonker/Schnabelrauch/Dehaan (2012). Sie definieren Wohlbefinden als ein Konstrukt, das über die Abwesenheit von Krankheit hinaus geht und verschiedene Outcomes wie Lebenszufriedenheit, Glück, Selbstbewusstsein und positive Stimmung umfasst (Yonker/Schnabelrauch/Dehaan 2012: 300). Dabei handelt es sich jedoch um eine tautologische Definition, da bereits in die Konzeption von Wohlbefinden Konstrukte eingeschlossen sind, die gleichzeitig Prädiktoren (bspw. Selbstbewusstsein) darstellen.

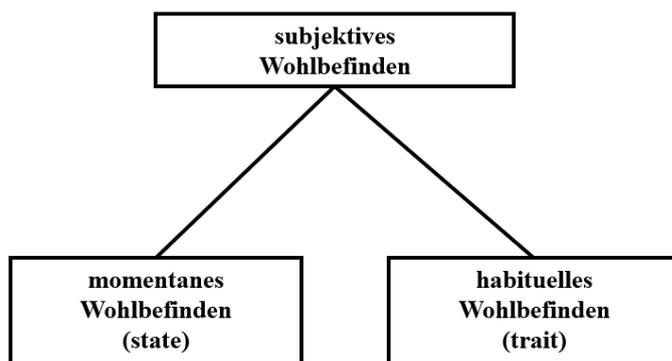
Eger/Maridal (2015) legen eine häufig verwendete Konzeption von Wohlbefinden vor. Sie definieren Wohlbefinden als eine Zusammensetzung der beiden Dimensionen Lebenszufriedenheit und Glück (Eger/Maridal 2015: 46). Dies entspricht der Unterscheidung in Affekt als emotionale Komponente und Lebenszufriedenheit als kognitives Urteil (Diener 1984, 2006). Anschließend unterscheiden sie evaluatives Wohlbefinden sowie hedonistisches bzw. affektives Wohlbefinden und nehmen eine Einteilung in subjektive und objektive Variablen vor (Eger/Maridal 2015: 47f.). Sie fokussieren sich in ihrer Analyse auf die objektiven Lebensbedingungen (Freunde, Familie, Gesundheit, Kriminalität etc.). Fischer/Boer (2011) definieren SWB als subjektive Evaluation des eigenen Lebens, welches emotionale Reaktionen in Hinblick auf persönliche Erfahrungen, Stimmungen und eine Vielzahl von Bewertungen verschiedener Lebensbereiche (bspw. Arbeit, Familie, Einkommen etc.) beinhaltet.

Einigkeit herrscht vor allem darüber, dass es sich um ein mehrdimensionales Konstrukt handelt, das sowohl positive wie auch negative Aspekte beinhalten kann (Charlemagne-Badal et al. 2015; Heidl et al. 2012: 7). Demgemäß wird subjektives Wohlbefinden definiert als „a person’s cognitive and affective evaluations of his or her life“ (Diener/Oishi/Lucas 2011: 187). Subjektives Wohlbefinden ist also die Gesamtbeurteilung des eigenen Lebens und der eigenen emotionalen Erfahrungen und dabei mehrdimensional zu erfassen sowie präziser in der Verwendung als das unpräzise Glück („happiness“) (Diener et al. 2017). SWB besteht somit aus zwei Dimensionen, einer globalen Evaluation der eigenen Lebensqualität sowie einer eher momentanen affektiven Gefühlslage, bestehend aus positiven und negativen Gefühlen (Myers/Diener 1995: 11; Pavot/Diener 1993, 2008).

Einige Autoren bezweifeln jedoch die faktische Korrektheit der Trennung in eine affektive und eine kognitive Komponente (Becker 1991; Mayerl 2001: 21). Vielmehr sollte eine Aufteilung in eine habituelle („trait“) sowie in eine momentane („state“) Komponente vorgenommen wer-

den. Die habituelle Seite beschreibt die eher langfristige, über die Zeit hinweg aggregierte, relativ stabile Eigenschaft auf Basis von kognitiven Urteilen und die „state“-Komponente eine eher auf den Augenblick bezogene Befindlichkeit des eigenen Erlebens (Dette 2005: 43; Heidl et al. 2012: 8). Es geht vorwiegend darum, dass eine Person ihr Leben als zufriedenstellend bewertet sowie, dass sie häufig positive und eher selten negativen Affekt erlebt. Dieser Konzeption nach gehören negative Gefühle und Erlebnisse zweifelsfrei zum menschlichen Leben (Grinde 2016).<sup>35</sup> Diese Konzeptionierung ist in Abbildung 6 zu sehen.<sup>36</sup> Das Konstrukt besteht aus dem momentanen Wohlbefinden, das in der ursprünglichen Konzeptionierung oftmals affektives Wohlbefinden genannt, und dem habituellen Wohlbefinden (Lebenszufriedenheit), welches ursprünglich als kognitives Wohlbefinden bezeichnet wurde. Jede der Dimensionen nimmt also einen unterschiedlichen Aspekt (kurzfristige Perspektive vs. langfristige Perspektive) des subjektiven Wohlbefindens in den Blick.<sup>37</sup>

**Abbildung 6: Definition subjektives Wohlbefinden**



**Anmerkung:** Eigene Abbildung.

Ebenfalls müssen die Konzeptualisierungen ausgeschlossen werden, die objektive Kriterien wie etwa die Gesundheitseinschätzung durch medizinische Erhebungsinstrumente oder physiologische Phänomene wie Herzschlag oder Stresslevel als SWB-Messung verwenden. Konzepte wie

<sup>35</sup> Die typischerweise angeführten Prädiktoren zur Vorhersage des subjektiven Wohlbefindens sind Alter, Bildung, finanzieller Status, Beschäftigungsstatus, Geschlecht, Ethnie, Familienstand, soziale Integration, Optimismus, Hoffnung, Selbstwert, Lebenssinn, Kontrollüberzeugung, Gesundheit, aktive Lebensführung und Genetik (Koenig/King/Carson 2012: 124ff.). Viele dieser Prädiktoren beziehen sich auf die oben angeführten objektiven Lebensbedingungen und einige werden im weiteren Verlauf dieser Arbeit noch eine Rolle spielen, andere hingegen stellen Erklärungsmechanismen über den zu Grunde liegende Zusammenhang zwischen R/S und SWB dar.

<sup>36</sup> Gelegentlich wird diese Konzeption noch um die sogenannte Bereichszufriedenheiten erweitert, also die Zufriedenheit in speziellen Lebensbereichen, also den mehr oder weniger objektiven Kriterien des guten Lebens (siehe oben), wie etwa der Arbeit oder dem Eheleben (Casas et al. 2015; Cummins 1996; Diener et al. 1999). Eine solche Definition wird jedoch in der vorliegenden Konzeptualisierung ausgeschlossen, weil sie zum einen äußerst komplexe nicht-lineare Beziehungen mit Lebenszufriedenheit aufweist (González et al. 2010) und zum anderen ein anderes Erklärungsmodell in Bezug auf die unabhängige Variable und den nachfolgenden Erklärungsmechanismen ergeben würde.

<sup>37</sup> Es gibt zwar kulturelle Unterschiede in Bezug auf das subjektive Wohlbefinden wie etwa Unterschiede im Durchschnittsniveau, diese lassen sich jedoch zum Großteil mit Unterschieden bezüglich der unterschiedlichen Verhältnisse der Nationen wie etwa Wohlstand erklären (Diener et al. 2017: 93).

psychologisches Wohlbefinden (Compton 2001; Ryff 1989; White/Blackmore 2016: 20ff.) werden ebenfalls aus der Definition ausgeschlossen, da dort häufig Bestandteile (bspw. der Sinn des Lebens) enthalten sind, die wichtige Prädiktoren von subjektiven Wohlbefinden bzw. eigentlich die Mediatorvariablen der Erklärungsmechanismen zwischen R/S und SWB darstellen (Argyle 2002; Dette 2005: 45f.; Mayring 1991). Zur Vereinfachung des meta-analytischen Untersuchungsmodells konzentriert sich die vorliegende Studie auf die habituelle Seite des subjektiven Wohlbefindens, also das Konzept der Lebenszufriedenheit. Dies hat neben der Modellvereinfachung den großen Vorteil, dass zur Erfassung dieses Konstrukts nur wenige Messinstrumente in der empirischen Praxis verwendet werden.

Nach dem die beiden zentralen Konzepte dieser Untersuchung definiert sind, wird im nächsten Schritt deren Beziehung zueinander empirisch-analytisch begründet (Kapitel 3.3) und ein Erklärungsmodell für die nachfolgende modellbasierte Meta-Analyse entwickelt (Kapitel 3.4).

### **3.3 Zusammenhangsbegründung**

Religiosität/Spiritualität wird in Abgrenzung zu den frühen (psychologischen) Arbeiten heutzutage zumeist als protektiver bzw. salutogener Faktor, das heißt als der Gesundheit und dem menschlichen Wohlbefinden zuträgliche Ressource, aufgefasst (Klein/Albani 2011b: 215). Dabei wird analytisch inzwischen immer seltener ein direkter Zusammenhang zwischen Religiosität/Spiritualität und Lebenszufriedenheit bzw. Gesundheit vermutet. Inzwischen ist man vielmehr dazu übergegangen, die dahinterliegenden Erklärungsmechanismen (Mediatoren) stärker in den Fokus zu nehmen (Pargament 2002: 243). Dabei ist eine große Diskrepanz zwischen den inzwischen sehr elaborierten Theorien und den durchgeführten empirischen Arbeiten zu erkennen. Trotz weit fortgeschrittener Theorieentwicklung über den Zusammenhang wird oftmals nur die direkte Beziehung untersucht (Park 2007).

Eine Vielzahl von empirischen Quer- und Längsschnitтарbeiten sowie Literaturreviews (Donahue/Benson 1995; Rizvi/Hossain 2016; Schreiber/Brockopp 2012; Visser/Garssen/Vingerhoets 2010) zeigt einen zwar schwachen, aber positiven Zusammenhang zwischen Religiosität und Lebenszufriedenheit.<sup>38</sup> Der Großteil dieser Studien wurde bezüglich des Zusammenhangs von Religiosität und Wohlbefinden angefertigt (Barton/Miller 2015: 830ff.; Handal/Fenzel 2000:

---

<sup>38</sup> Eine große Auswahl von Studien zeigt dies (Abdala et al. 2015; Abu-Raiya/Agbaria 2016; Bergan/McConatha 2001; Cohen 2002; Cohen/Johnson 2011: 2ff.; Diener/Tay/Myers 2011; Doane 2013: 49; Donahue/Benson 1995; Eichhorn 2012; Ellison 1991; Ellison/Gay/Glass 1989; Ferriss 2002: 199ff.; Francis et al. 2004; Gartner/Larson/Allen 1991; Headey et al. 2010: 73ff.; Heinemann/Wörmann 2010; Ivztan et al. 2013; Judd 1986: 71ff.; Kashdan/Nezlek 2012: 1523ff.; Kim-Prieto/Diener 2009: 447ff.; Klein/Albani 2007; Koenig 2012: 4; Koenig/Larson 2009: 71; Leondari/Gialamas 2009; Myers 2000: 63, 2008: 234ff.; Payne et al. 1991: 14; Stavrova/Fetchenhauer/Schlosser 2013; Steffen/Masters/Baldwin 2016; Tiliouine 2009: 94; Wong/Rew/Slaikau 2006: 174).

225ff.). Eine geringe Anzahl von Studien findet hingegen auch negative oder keine Zusammenhänge (Koenig/King/Carson 2012: 134ff.).

Viele Studien haben soziodemographische Kontrollvariablen (Alter, Geschlecht, Rasse, Bildung, Arbeit, Familie, Finanzstatus etc.) berücksichtigt und festgestellt, dass auch nach deren Kontrolle ein eigenständiger Effekt bestehen bleibt (Eichhorn 2012; Green/Elliott 2010: 155ff.; Visser/Garssen/Vingerhoets 2010: 566). Gleich ob R/S mit Glück, Lebenszufriedenheit oder subjektivem Wohlbefinden empirisch in Beziehung gesetzt wird, der Zusammenhang ist in der Regel schwach positiv (Cohen/Johnson 2011: 3).

Die drei Meta-Studien, die die Verbindung zwischen den beiden Variablen direkt untersuchen, finden einen schwachen positiven Zusammenhang. Witter et al. (1985) berichten eine durchschnittliche Effektstärke von .16 über die 28 berücksichtigten Studien hinweg. Fast die gleiche Effektstärke zeigt später die Studie von Yonker/Schnabelrauch/Dehaan (2012) in Bezug auf den Zusammenhang von Religiosität/Spiritualität und Wohlbefinden bei den von ihnen untersuchten Studien an Jugendlichen (.15). Einen deutlich höheren Zusammenhang finden Sawatzky/Ratner/Chiu (2005). Bei ihren Analysen können sie im Random-Effects-Modell einen Zusammenhang zwischen Spiritualität und Lebensqualität von durchschnittlich .34 zeigen.

Auch die meta-analytischen Arbeiten, die die Beziehung zwischen Religiosität/Spiritualität und verschiedene Gesundheitsaspekte untersuchen, zeigen schwach positive Korrelationen, wobei stets eine positive Wirkrichtung von R/S auf Gesundheit kolportiert wird. Für den Zusammenhang von R/S und mentaler Gesundheit zeigen die Studien von Salsman/Pustejovsky et al. (2015) und Hackney/Sanders (2003) durchschnittliche Effektstärken von .19 und .1. Auch bei Bergin (1983) zeigt sich ein leicht positiver Effekt. Ähnlich bei der Studie von Smith/McCullough/Poll (2003), die einen sehr schwachen negativen Zusammenhang zwischen Religiosität und Depression (-.096) gefunden haben. McCullough et al. (2000) zeigen eine 29 % höhere generelle Überlebenschance von religiösen Menschen gegenüber nicht-religiösen Menschen. In einer ähnlichen Untersuchung konnten Chida/Stephoe/Powell (2009) ebenfalls einen protektiven Effekt sowohl bei Gesunden als auch bei Kranken nachweisen. Auch in Bezug auf Suizid hat Religiosität einen schützenden Effekt (Wu/Wang/Jia 2015). Religiöse Menschen haben ihren Ergebnissen zufolge ein deutlich geringeres Risiko (Odds Ratio = .38), Suizid zu begehen. Shor/Roelfs (2013) zeigen, dass religiöse Partizipation ein Prädiktor für ein langes Leben darstellt. Jim et al. (2015) weisen einen schwachen positiven Zusammenhang (.153) zwischen R/S und physischer Gesundheit nach.

Alle Forschungsbefunde zusammengefasst zeigen, dass die der Lebenszufriedenheit (und der Gesundheit) zuträglichen Wirkungen der Religiosität gegenüber den negativen Wirkungen deutlich überwiegen (Klein/Albani 2011b: 223). Diese Studienlandschaft ist jedoch in Bezug auf Definitionen der Kernkonzepte, verwendeten Operationalisierungen, Messungen und Analysemethoden äußerst heterogen. Aus diesem Grund kommt es immer wieder zu unterschiedlichen und teilweise auch widersprüchlichen Studienergebnissen (Lewis/Cruise 2006; Sergent 2015). Die Vielfalt der Ergebnisse ist ein Indiz für die Notwendigkeit von Meta-Analysen, denn diese können die Variabilität der Ergebnisse als eigenständige Forschungsfrage in den Blick nehmen.

### **3.4 Ableitung des Erklärungsmodells**

Nachdem der Zusammenhang der beiden Konstrukte empirisch begründet ist, stellt sich nun im nächsten Schritt die Frage, welche Art von Erklärungsmodellen die bisherigen Untersuchungen heranziehen. Ein empirisch bewährtes Erklärungsmodell liegt für die Beziehung bislang nicht vor. Die verschiedenen vorliegenden Ansätze werden zu einem Erklärungsmodell vereinigt, welches die Grundlage für die anschließende Meta-Analyse darstellt.

Die gefundenen Meta-Studien über die Beziehung zwischen Religiosität/Spiritualität und (subjektiven) Wohlbefinden/Lebenszufriedenheit haben kein einheitliches theoretisch-analytisches Framework (Eger/Maridal 2015; Sawatzky/Ratner/Chiu 2005; Witter et al. 1985; Yonker/Schnabelrauch/Dehaan 2012). Die konzeptionellen Abhandlungen sind in der Regel sehr kurz und beschränken sich auf eine vorwiegend empirische Herleitung des Forschungsgegenstands. Witter et al. (1985) haben etwa lediglich vage Vermutungen über die direkten und indirekten Wirkweisen anhand von verschiedenen empirischen Ergebnissen gesammelt und diese nicht systematisiert (Witter et al. 1985: 332). Die Studie von Sawatzky/Ratner/Chiu (2005) legt ebenso eine empirisch-analytische Konzeption zugrunde, welche von einem positiven Zusammenhang von R/S und Lebensqualität ausgeht und dazu eine Vielzahl von empirischen Ergebnissen anführt. Sie stützen ihre Analyse ausschließlich auf Messungen, die den subjektiven Charakter von Lebensqualität und Spiritualität berücksichtigen. Yonker/Schnabelrauch/Dehaan (2012) untersuchen den Zusammenhang von R/S und verschiedenen psychologischen Outcomes, die für eine gesunde Entwicklung von Jugendlichen relevant sind. Sie haben ebenfalls ein empirisch-analytisches Konzept zugrunde gelegt, das im Mittelpunkt das Risikoverhalten und entsprechende protektive Faktoren im Fokus hat sowie auf empirische Ergebnisse verweist

(Yonker/Schnabelrauch/Dehaan 2012: 299). Die Studie von Eger/Maridal (2015), die Religiosität nur als einen unter vielen Prädiktoren von Wohlbefinden erforscht, zieht im Gegensatz dazu unterschiedliche Konzeptionen über das Entstehen von Wohlbefinden heran.<sup>39</sup>

Ein ähnliches Bild ergibt sich in den Meta-Studien, die den Zusammenhang von R/S und Gesundheit untersuchen. Dort wird auf empirische Vorergebnisse hingewiesen und ebenfalls kein Erklärungsmodell zugrunde gelegt (Hackney/Sanders 2003). Manche Studien führen psychosoziale Mechanismen, wie etwa gesundheitsbezogenes Verhalten oder soziale Unterstützung, an (Chida/Stephoe/Powell 2009: 82; Jim et al. 2015: 3761; Wu/Wang/Jia 2015: 14, 2015: 2, 2015) oder verweisen auf bestimmte Moderatorvariablen wie Alter, Geschlecht oder ethnische Herkunft (McCullough et al. 2000: 211f.). Keine dieser Meta-Studien zieht ein geschlossenes Erklärungsmodell heran.<sup>40</sup> Somit zeigt sich eine Diskrepanz zwischen den geforderten realistischen Erklärungsmodellen, die der Komplexität gerecht werden, aber dennoch in der empirischen Praxis angewandt werden können (Klein/Albani 2011b: 223).

Da die empirischen Studien fast ausschließlich positive Effekte berichten, wird ein Erklärungsmodell für die positiven Wirkungen von Religiosität formuliert. Es sind jedoch auch negative Auswirkungen analytisch plausibel und haben sich in einigen Studien bereits gezeigt (Cohen/Johnson 2011: 13f.; Hill et al. 2000: 56; Wong/Rew/Slaikue 2006: 178).<sup>41</sup> Auch eine reziproke Beziehung zwischen den Konstrukten ist möglich (Klein/Albani 2011b: 216; Koenig/King/Carson 2012: 127ff.; Park/Slattery 2013: 551f.) und wurde bereits empirisch überprüft (Doane/Elliott 2016). Darüber hinaus gibt es bereits Erklärungsmodelle, die den Zu-

---

<sup>39</sup> Übersichten über die verschiedenen human- und sozialwissenschaftlichen Theorien über die Entstehung von (subjektivem) Wohlbefinden, wie etwa die broaden-and-build-Theorie, die self-determination-Theorie sowie bottom-up und top-down u. v. w. finden sich an anderer Stelle (Argyle 2002, 2003; Bruni/Porta 2014; Dette 2005: 50ff.; Diener/Oishi/Lucas 2011: 189ff.; Dolan/Peasgood/White 2008; Fredrickson 2004; George 2010; Guerin/Nuvolati 2014; Haybron 2008; Kahneman 2003; Mayerl 2001; Myers/Diener 1995; Schimack 2008; Veenhoven 2008: 44, 2010: 74ff.; White/Blackmore 2016: 1ff.). In der Regel werden drei sozialwissenschaftliche Theorien bzw. Abwandlungen davon prominent diskutiert: die Vergleichstheorie, die Folkloretheorie sowie die Lebensqualitätstheorie (Veenhoven 2007: 1; Veenhoven/Ehrhardt 1995). Weitere Ansätze wie etwa der von Ormel et al. (1999), die eine Verbindung von sozialen Produktionsfunktionen und herkömmlichen Ansätzen zum subjektiven Wohlbefinden vornehmen, sind bislang jedoch kaum rezipiert worden, sind aber durchaus vielversprechend.

<sup>40</sup> Eine Ausnahme bildet hierbei die soziologisch-kriminologische Meta-Analyse von Baier/Wright (2001) zum Zusammenhang von Religion und Kriminalität, die explizite theoretische Modelle heranzieht.

<sup>41</sup> Diese Befunde werden etwa durch den empirischen Zusammenhang von Religiosität mit einigen geistigen Erkrankungen, wie Angst- und Zwangsstörungen, erklärt (Bergin 1991; Klein/Albani 2007: e4ff.). Weitere Erklärungsansätze für negative Wirkungen von Religiosität sind negatives religiöses Coping, ein negatives strenges, weil bei persönlichen Verfehlungen strafendes Gottesbild, das Vorliegen bestimmter Werte und Einstellungen wie Autoritarismus und Abhängigkeit sowie negative Gefühle durch sozialen Druck und soziale Ausgrenzung, mehr Stress durch Rigorismus in Bezug auf religiöse Vorschriften, Verstärkung von Schuldgefühlen und Ängsten sowie psychologischen Belastungen und dem Gefühl der eigenen Sündhaftigkeit oder dem Vorliegen bestimmter Persönlichkeitseigenschaften (Klein/Albani 2011b: 226ff.; Koenig/King/Carson 2012: 311f.; Mehnert/Höcker 2011: 247ff.; Murken 1998: 72; Park/Slattery 2013: 550f.).

sammenhang von Religiosität und körperlicher Gesundheit und Wohlbefinden mittels physiologischer Funktionen erklären (Kusche 2013: 17ff.; Mehnert/Höcker 2011: 247ff.). Dort werden Wirkmechanismen wie Allostase, Immunfunktionen, hormonelle Wirkungen und andere körperliche Prozesse zur Erklärung des Zusammenhangs diskutiert. Die Wirkungen werden dabei jedoch den psychosozialen Erklärungsmechanismen nachkommend konzeptualisiert und deshalb aus dem Erklärungsmodell ausgeschlossen (Koenig/King/Carson 2012: 39ff.). Diese Überlegungen werden zur Vereinfachung der Modellkonstruktion ausgeschlossen.

Ein umfassendes Erklärungsmodell, welches positive und negative Wirkweisen von Religiosität und psychischer Gesundheit sowie Wechselwirkungen mithilfe von psychosozialen Mechanismen berücksichtigt, liegt mit dem aus der Psychologie bzw. Medizin stammenden Diathese-Stress-Modell vor (Klein/Albani 2011a: 33ff.; Zwingmann/Klein 2012: 13ff.). Dieser Ansatz erklärt, wie Religiosität bzw. Spiritualität zu einer besseren geistigen Gesundheit führt. Aufgrund der hohen konzeptionellen Nähe von psychischer Gesundheit zu dem Konzept subjektives Wohlbefinden mit den entsprechenden sehr hohen Korrelationen (Koenig/King/Carson 2012: 126) wird dieses humanwissenschaftliche Erklärungsmodell dem hier zu konstruierenden sozialwissenschaftlichen Erklärungsmodell analytisch zugrunde gelegt und durch ein umfassendes multidisziplinäres Literaturreview erweitert. Dies lässt sich damit begründen, dass in der Literatur die gleichen Erklärungsmechanismen für die unterschiedlichen abhängigen Variablen (psychische) Gesundheit und subjektives Wohlbefinden diskutiert werden (Sergent 2015: 1ff.). Das Diathese-Stress-Modell, welches sukzessive von Murken (1997), Dörr (2001) und Klein/Albani (2007) entwickelt wurde, bildet die Grundlage für das Erklärungsmodell. Dabei findet eine Fokussierung auf die Ausschnitte des Modells statt, die für eine sozialwissenschaftliche Untersuchung relevant sind. Das Modell besagt, dass die Befindlichkeit eines Individuums aus den bestehenden Anforderungen wie Alltagsstress, chronische Stressoren (etwa Krankheit) und kritische Lebensereignisse (bspw. Todesfall) resultiert. Das Ergebnis dieser Reaktion wird von den Anforderungen und der genetischen Prädisposition des Individuums (bspw. neurotische Persönlichkeit) sowie der Umwelt (bspw. stark religiöse Kultur) mitbestimmt.

Hierbei wird eine Vielzahl unterschiedlicher Mechanismen vermutet. Dabei sind direkte und indirekte Pfade möglich (Koenig/King/Carson 2012: 127; Ventis 1995: 33ff.). Überzeugender sind jedoch indirekte Variablenbeziehungen in Form von Mediatorbeziehungen. Viele empirische Studien ignorieren dies bislang jedoch. Diese Mechanismen sollen nun, so die weit verbreitete Meinung in der Literatur, stärker in den Fokus rücken und mit den entsprechenden

Methoden untersucht werden.<sup>42</sup> Die modellbasierte Meta-Analyse kann hierbei ihre Vorzüge zeigen, da der Ansatz auch Studien berücksichtigen kann, die lediglich Ausschnitte des Modells untersucht haben.

Bei dem hier entwickelten Modell handelt es sich demnach um ein psychosoziales Erklärungsmodell, das Wechselwirkungen zwischen den psychologischen Prozessen auf Individualebene vermutet und darüber hinaus auch den jeweiligen sozialen Kontext berücksichtigt (Tay et al. 2014). Dies entspricht der skizzierten Konzeptualisierung, die von einer Wechselwirkung zwischen sozialer Umwelt und inneren Bewusstseinsprozessen ausgeht. Die verschiedenen Mediatoren schließen sich dabei nicht aus, sondern stellen sich ergänzende Erklärungspfade dar. Zunächst folgt nun eine Sammlung der Erklärungsmechanismen.<sup>43</sup> Die Mechanismen werden in den empirischen Arbeiten zumeist unsystematisch eingeführt (Garssen/Uwland-Sikkema/Visser 2015: 1249ff.; Sergent 2015: 12ff.; Ward 2010: 21ff.). Der humanwissenschaftlichen Literatur zufolge sind sechs unterschiedliche theoretische Zugänge relevant (siehe Tabelle 2). Wie in Tabelle 2 deutlich wird, sind diese sechs Erklärungspfade jedoch nicht überschneidungsfrei konzeptualisiert.

Die humanwissenschaftlichen Arbeiten konzeptualisieren die Erklärungspfade direkt auf die physische und psychische Gesundheit.<sup>44</sup> Ein solches analytisches Modell ist empirisch und vor allem meta-analytisch (> 30 Mediatorpfade) nicht umsetzbar und muss vereinfacht werden. Um einen Überblick zu erhalten, werden die Erklärungsmechanismen zunächst zu Clustern zusammengefasst.

---

<sup>42</sup> Diese Forderung wird in der Literatur gestellt (Diener/Clifton 2002: 208f.; Doane 2013; Eichhorn 2012; George et al. 2000: 110; Hayward/Krause 2014; Koenig/King/Carson 2012: 143; Park et al. 2015: 3792; Pressman/Cohen 2005: 963; Unterrainer/Lewis/Fink 2014: 391; Ward 2010: 1; Wong/Rew/Slaikeu 2006: 178).

<sup>43</sup> Für eine detailreichere Erläuterung der verschiedenen Mechanismen in diesem Bereich siehe Murken (1997), Murken (1998) und Klein/Albani (2007) sowie Klein/Albani (2011b). Dort sind ebenfalls Verweise auf die Erläuterungen der verschiedenen Konzepte zu finden.

<sup>44</sup> Diese besagen zumeist, dass das Stressniveau abgesenkt wird und somit positive Emotionen in Form von Wohlbefinden und Gesundheit hervorgerufen und darüber hinaus noch direkte neuroimmunologische und physiologische Effekte verursacht werden (Klein/Albani 2007: e6).

**Tabelle 2: Übersicht Erklärungsmechanismen**

Mechanismen	Erläuterung	Variablen	Quellen
1) Theorie der sozialen Unterstützung (Kohäsion)	Soziale Netzwerke führen zu emotionaler (Trost, Anteilnahme, Mitgefühl etc.) und materieller Unterstützung	Emotionale Unterstützung (Hilfe, Trost, Anteilnahme, Mitgefühl), Integration und soziale Eingebundenheit	(Chida/Steptoe/Powell 2009: 82; Diener/Ryan 2009: 397; Doane 2013: 53ff.; Ellison/Levin 1998: 705; George et al. 2000: 110f.; George/Ellison/Larson 2002; Hayward/Krause 2014; Kortt/Dollery/Grant 2015; Krok 2014: 67; Lau et al. 2015; Lim/Putnam 2010; McCullough/Willoughby 2009: 70; Mehnert 2006: 783; Moreira-Almeida/Lotufoto Neto/Koenig 2006: 247; Myers 2008: 326; Mytko/Knight 1999: 445; Park/Roh/Yeo 2012: 641ff.; Park/Slattery 2013: 547f.; Sergent 2015: 12f.; Seybold/Hill 2001: 23; Shor/Roelfs 2013: 121; Smith/McCullough/Poll 2003; Steffen/Masters/Baldwin 2016; Thoresen 1999: 296; van der Jagt-Jelsma et al. 2015: 845; Wu/Wang/Jia 2015: 2; Yoon et al. 2014: 134)
2) Theorie der kognitiven Orientierung/der Kohärenz	Religiöse/spirituelle Strömungen liefern Erklärungen dafür, warum die Welt so ist wie sie ist (bspw. Kontingenzproblem, Todesfälle, Schicksalsschläge) und liefern somit Halt sowie Orientierung und lassen die Gläubigen Sinn an den Geschehnissen finden und Gefühle wie Hoffnung und Kohärenz vermitteln	Kohärenz, Hoffnung, Optimismus, Lebenssinn	(George et al. 2000: 111f.; George/Ellison/Larson 2002; Klein/Albani 2007: e5; Mehnert 2006: 783; Myers 2008: 328; Park/Slattery 2013: 547ff.; Seybold/Hill 2001: 23; Zarzycka/Rydz 2014)
3) Theorie der Verhaltensregulierung bzw. gesunder Lebensstil	Religiöse/spirituelle Traditionen formulieren oft konkrete Handlungsvorgaben zu Körperpflege, körperlicher Aktivität, Ruhezeiten, Sexualverhalten, Ernährung und Substanzkonsum. R/S führt in der Regel zu gesünderem Verhalten und niedrigeren Sensation-Seeking, also Verhaltensweisen, die direkt dem Erhalt der Gesundheit dienen oder Stress vermeiden	Drogenkonsum, Alkoholkonsum, Tabakkonsum, sportliche Aktivität, sexuelles Risikoverhalten, gesunde Ernährung	(Chida/Steptoe/Powell 2009: 82; Ellison/Levin 1998: 704; George et al. 2000: 110; George/Ellison/Larson 2002; Headey/Hoehne/Wagner 2014; Klein/Albani 2007: e5; McCullough/Willoughby 2009: 70; Mytko/Knight 1999: 444; Park/Slattery 2013: 548; Seybold/Hill 2001: 23; Shor/Roelfs 2013: 121; Thoresen 1999: 296; van der Jagt-Jelsma et al. 2015: 845; Wu/Wang/Jia 2015: 2; Yoon et al. 2014: 134),
4) Theorie der positiven Gefühle/Emotionen/des Selbstwerts	R/S kann zu einem positiven Selbstwertgefühl führen, zu einem Gefühl wie unbedingtes Angenommensein, Vertrauen oder fürsorgliche Liebe sowie emotionale Instabilität reduzieren und Einsamkeitsgefühle reduzieren.	Vertrauen, Angenommen-Sein, fürsorgliche Liebe, positive Gottesbeziehung, Selbstwertgefühl, weniger Einsamkeit, weniger emotionale Instabilität	(Chida/Steptoe/Powell 2009: 82; Ellison/Levin 1998: 708; Kim-Prieto/Diener 2009; Klein/Albani 2007: e6; Moreira-Almeida/Lotufoto Neto/Koenig 2006: 247; Park/Slattery 2013: 549f.; Seybold/Hill 2001: 23; van Cappellen et al. 2016)
5) Theorie der alternativen Werte	R/S vermitteln alternative Werte wie Demut, Verzeihen und Dankbarkeit	Demut, Verzicht, Selbstverwirklichung, Genuss, Harmonie, Verzeihen, Dankbarkeit	(Klein/Albani 2007: e5f.; Myers 2008: 328ff.; Park/Slattery 2013: 548; Sergent 2015: 13f.).
6) Theorie des religiösen Copings	Die religiöse/spirituelle Deutung der eigenen Belastungssituationen hilft dabei, mit den schwierigen Situationen umzugehen, indem das Gefühl, hoffnungslos ausgeliefert zu sein, reduziert und Geborgenheit sowie Sicherheit vermittelt wird.	Bewältigungsstrategien	(Cohen/Johnson 2017; Ellison/Levin 1998: 707; Klein/Albani 2007: e6; Krok 2014: 67; Lau et al. 2015; McCullough/Willoughby 2009: 70; Moreira-Almeida/Lotufoto Neto/Koenig 2006: 247; Mytko/Knight 1999: 445; Park/Slattery 2013: 549; Seybold/Hill 2001: 23; Shor/Roelfs 2013: 121f.; van der Jagt-Jelsma et al. 2015: 845; Yoon et al. 2014: 134),

Die sechs Erklärungspfade lassen sich für die analytische Darstellung in drei Cluster zusammenfassen (siehe Abbildung 7). Eine ähnliche Einteilung der Mediatoren nehmen etwa auch

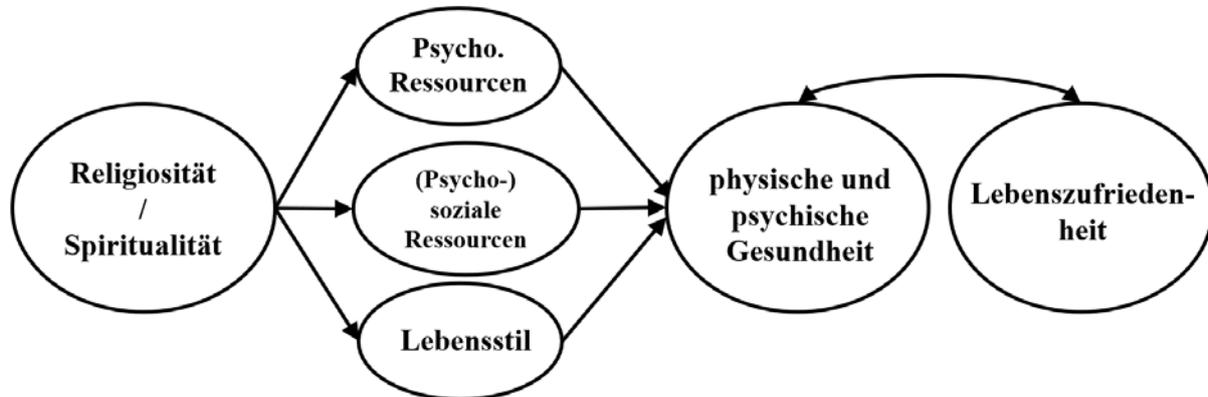
Morton/Lee/Martin (2017) vor. Der erste Bereich stellt die **psychologische Ressourcen** dar. Diese Ressourcen entstehen unter anderem durch Religiosität oder werden durch sie genährt und haben einen Einfluss auf Lebenszufriedenheit (Ellison/Levin 1998: 706f.; Hayward/Krause 2014; Kortt/Dollery/Grant 2015; Lau et al. 2015; Myers 2008: 327f.; Shor/Roelfs 2013: 121; Thoresen 1999: 296; Wu/Wang/Jia 2015: 2). Die spirituellen/religiösen Überzeugungen führen zu einer Vielzahl unterschiedlicher individueller Überzeugungen, Werte und Bewältigungsstrategien, die ihren Beitrag dazu leisten, die Lebenszufriedenheit zu erhöhen. Religiöse Menschen können etwa durch religiöses Coping besser mit kritischen Lebensereignissen (bspw. Krankheit, Alter) umgehen, was sich wiederum in ihrer langfristigen Lebenszufriedenheit niederschlägt (Klein/Albani 2011a: 30; Murken 1997: 162). Ebenso fällt in diesen Bereich die sogenannte Kohärenzhypothese, die besagt, dass Gläubige durch religiöse Interpretationen kognitive Kohärenz erlangen und die Welt als beeinflussbar und kontrollierbar wahrnehmen (Murken 1997: 162). Ansonsten Unerklärliches, wie etwa Todesfälle, wird bewältigbar. Positive Emotionen (Hoffnung, Optimismus, Vertrauen, Selbstvertrauen, Lebenszweck etc.) und Lebenssinn („meaning and purpose“) entstehen (Koenig/King/Carson 2012: 579ff.).

Der nächste Bereich beinhaltet **(psycho-)soziale Ressourcen**, die einen Einfluss auf die Lebenszufriedenheit haben (George/Ellison/Larson 2002; Hayward/Krause 2014; Park/Slattery 2013: 548; van der Jagt-Jelsma et al. 2015: 845). Es ist sowohl emotionale und materielle Unterstützung in Religionsgemeinschaften gemeint (soziale Netzwerke), als auch das Gefühl, sich einer bestimmten sozialen Gruppe zugehörig zu fühlen (soziale Identitätsressource) (Klein/Albani 2011a: 25f.). Dies ist der einzige Mediatorbereich, der auch von einer der betrachteten Meta-Analysen angesprochen wird (Witter et al. 1985). Die Autoren gehen davon aus, dass der Effekt zwischen Religiosität und Wohlbefinden vor allem aufgrund der Kohäsionshypothese, also auf Grund sozialer Einbettung, wirksam ist.

Der dritte Bereich umfasst diejenigen Erklärungsmechanismen, die aufgrund von bestimmten (gesunden) Verhaltensweisen (**Lebensstil**) zu einem erhöhten Wohlbefinden durch positives gesundheitsbezogenes Verhalten beitragen können und nicht ausschließlich auf physiologische Auswirkungen beschränkt sind. Dies wird auch als „religiös motiviertes Gesundheitsverhalten“ (Klein/Albani 2011a: 33) bezeichnet. Religiöse bzw. spirituelle Menschen rauchen in der Regel

weniger, trinken weniger Alkohol, nehmen weniger Drogen und verhalten sich insgesamt gesünder, indem sie Sport treiben oder bestimmte Diät- bzw. Hygienevorschriften einhalten oder generell, aus religiösen Gründen, auf ihren Körper achten.<sup>45</sup>

**Abbildung 7: Erklärungsmodell mit Mechanismusclustern**



**Anmerkung:** Eigene Abbildung.

Allerdings entspricht dieser Teil des Erklärungsmodells nicht dem festgelegten Kriterium, dass lediglich diejenigen Variablenbeziehungen berücksichtigt werden sollen, die eine direkte Auswirkung auf die Lebenszufriedenheit haben. Deshalb wird das Erklärungsmodell um diejenigen Mechanismen verringert, die eher eine Wirkung auf die Gesundheit haben. Das „Cluster“ Lebensstil wird deshalb aus dem Untersuchungsmodell entfernt, da hier eher direkte Wirkungen auf Gesundheit zu erwarten sind.<sup>46</sup>

Ebenfalls ausgeschlossen wird die Theorie des religiösen Copings als Bestandteil der psychologischen Ressourcen, da diese vorwiegend auf die psychische Gesundheit eines Individuums Einfluss nehmen und in Bezug auf die Erhebungsinstrumente komplex zu erfassen sind. Durch den Verzicht dieser beiden Bereiche wird das Erklärungsmodell deutlich vereinfacht und für den sozialwissenschaftlichen Anwendungsfall passender gestaltet (siehe Abbildung 8).

Auch nach der Reduzierung des Erklärungsmodells bleiben noch über 20 potenzielle Mediatorvariablen sowie deren unterschiedlichen Operationalisierungen bestehen. Die Mediatoren werden deshalb so detailliert wie möglich, mit einer möglichst geringen Anzahl an unterschiedlichen Messinstrumenten, erfasst.

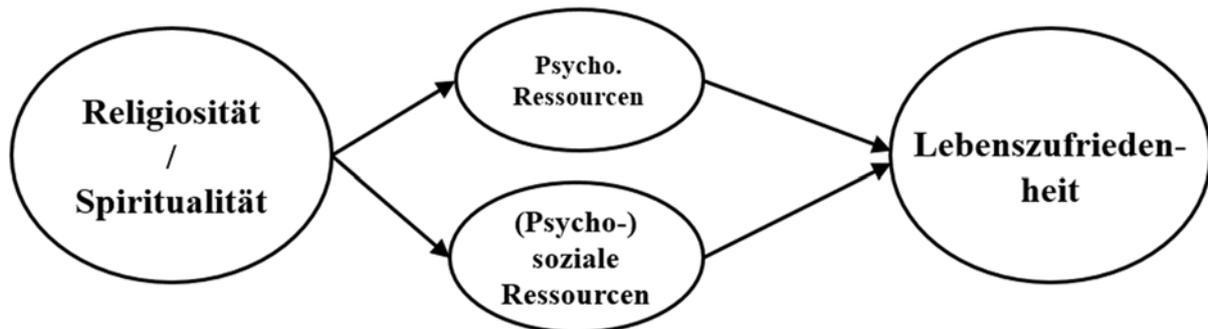
Es muss auch bedacht werden, dass die von Klein/Albani (2007) vorgeschlagenen Erklärungspfade für den Zusammenhang von organisierter Religiosität und Gesundheit formuliert wurden.

<sup>45</sup> Dies wird unter anderem auch dadurch erklärt, dass religiöse bzw. spirituelle Menschen ein höheres Maß an Selbstbeherrschung und Kontrollüberzeugung aufweisen und wird deshalb auch unter dem Begriff der „Verhaltensregulierung“ diskutiert (Cohen/Johnson 2017; Lau et al. 2015; McCullough/Willoughby 2009: 86; Mehnert 2006: 783; Murken 1997: 160ff.; Osborne/Milojev/Sibley 2016).

<sup>46</sup> Etwa durch Variablen wie niedriger Tabak- und Alkoholkonsum, die sich direkt auf die körperliche und geistige Gesundheit und dadurch nur indirekt auf das habituelle Wohlbefinden auswirken.

Für den Zusammenhang der spirituellen Komponente werden deshalb noch weitere in der Literatur identifizierte Mediatoren berücksichtigt, die ebenfalls in den Bereich der psychologischen Ressourcen eingeordnet werden. Dies ist zum einen die Selbstwirksamkeitserwartung (self-efficacy) und zum anderen die Selbstbeherrschung bzw. Kontrollüberzeugung, also die Überzeugung, inwiefern ein Individuum selbst verantwortlich ist für das eigene Leben (Abdel-Khalek/Lester 2017; Jackson/Bergeman 2011; Osborne/Milojev/Sibley 2016).

**Abbildung 8: Sozialwissenschaftliches Erklärungsmodell**



**Anmerkung:** Eigene Abbildung.

Darüber hinaus werden nun verschiedene **Moderatorvariablen** erörtert, die ebenfalls in der Literatur diskutiert werden. Ein sehr bedeutsamer Moderator der Beziehung ist sicherlich das Ausmaß von stressigen Lebensereignissen (bspw. Todesfall eines Angehörigen, Trennung vom Partner etc.), mit der ein Individuum zu kämpfen hat (Smith/McCullough/Poll 2003: 614). Weiterhin hat auch die genetische Veranlagung in Form von bestimmten Persönlichkeitseigenschaften („traits“), wie etwa Neurotizismus einen moderierenden Einfluss auf die Beziehung (Aghababaei 2013; Crivelli/Della Nella/Lucchini 2014: 176ff.; Diener/Lucas 2003; Ferriss 2002: 212; Joshanloo/Afshari 2011; Lucas 2008; Lucchini/Della Nella/Crivelli 2003; McCullough/Willoughby 2009: 87).

Die Relevanz dieser Variablen konnte eine Meta-Analyse aus dem Jahr 2015 zeigen, da dort die genetischen Faktoren für etwa 30 % der Varianz von Lebenszufriedenheit verantwortlich sind (Bartels 2015). Die Persönlichkeit ist somit einer der wichtigsten Prädiktoren des Wohlbefindens überhaupt und kann vergleichsweise mehr Varianz aufklären als die objektiven Indikatoren Gesundheit und Einkommen (Lucas 2008: 188f.).<sup>47</sup> Aber auch soziodemographische Moderatoren wie Rasse/Ethnie, Bildung, Konfession und Einkommen sind für die Variablenbeziehung relevant (Lim 2015: 699; Peterson/Webb 2006: 114; Salsman/Fitchett et al. 2015:

<sup>47</sup> Manche Autoren interpretieren solche Ergebnisse insofern, dass das Ausmaß von individuellem Glück und Wohlbefinden zum Großteil ein genetischer „Zufall“ ist und die objektiven Einflussvariablen kaum einen Einfluss haben (Lykken/Tellegen 1996).

3758; Sergent 2015: 15f.; Tiliouine 2009: 94, 2009: 94). Auch nach Kontrolle der soziodemographischen Faktoren bleibt ein eigenständiger Effekt bestehen (Hackett/Marshall/Schiller 2019).

Allerdings können die genannten Moderatoren in einer modellbasierten Meta-Analyse kaum umgesetzt werden, da die entsprechenden Variablen meist nicht berichtet werden. Lediglich ausgewählte Moderatoranalysen können durchgeführt werden. Dies können zum einen Studiencharakteristiken, wie etwa der Publikationsstatus, und zum anderen kulturelle Makroindikatoren sein. Die bekannten Meta-Studien in diesem Bereich haben ebenfalls Moderatoranalysen durchgeführt. Die Geschlechtsvariable hat keinen signifikanten Einfluss (Sawatzky/Ratner/Chiu 2005; Witter et al. 1985; Yonker/Schnabelrauch/Dehaan 2012). Im Gegensatz dazu hat die Art der Religiositätsmessung eine Erklärungsleistung (Sawatzky/Ratner/Chiu 2005; Witter et al. 1985). Aber auch die Moderatoren für Alter, Rasse sowie Publikationsjahr sind teilweise signifikant (Witter et al. 1985; Yonker/Schnabelrauch/Dehaan 2012). Die Studienqualität hingegen scheint keinen Einfluss zu haben (Witter et al. 1985).

Der Einfluss des **kulturellen Kontexts** bei der Beziehung zwischen Religiosität und Lebenszufriedenheit wird zunehmend in den Fokus gerückt (Cohen/Johnson 2011: 10ff.; Diener 2008: 504ff.; Diener/Clifton 2002: 208; Eichhorn 2012; Hill et al. 2000: 69f.; Jagodzinski 2010; Masters/Hooker 2013: 209ff.; Okulicz-Kozaryn 2009: 1ff.; Okulicz-Kozaryn 2010; Peterson/Webb 2006; Sabatier et al. 2011; Stavrova/Fetchenhauer/Schlosser 2013; Tay et al. 2014). Viele Erkenntnisse über die individuellen Auswirkungen von Religiosität sind anhand westlicher Stichproben untersucht worden und deren Übertragbarkeit auf andere Kulturbereiche bleibt diskussionswürdig (Cohen/Wu/Miller 2016). Die Beziehung zwischen Religiosität und Lebenszufriedenheit kann etwa mit dem Ausmaß staatlicher Regulation des „religiösen Marktes“ oder der persönlichen Freiheit variieren (Elliott/Hayward 2009). Ebenso haben das Ausmaß der Religiosität in einem Land und die gesellschaftlichen Umstände einen Einfluss auf die Beziehung zwischen den beiden Variablen (Diener/Tay/Myers 2011: 1290; Eichhorn 2012).

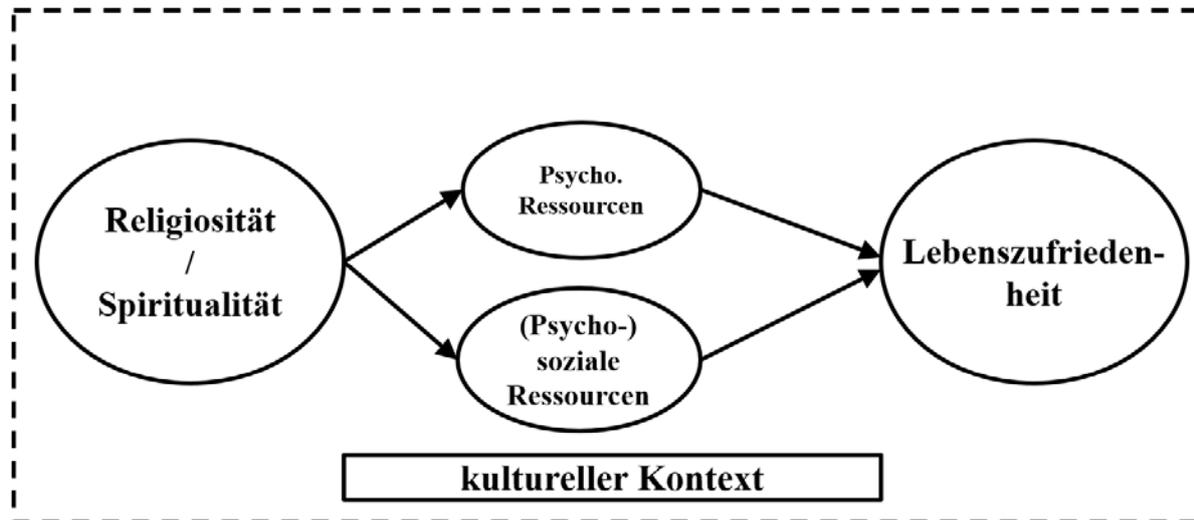
Die bislang präsentierten Überlegungen führen zu einem sozialwissenschaftlich geprägten psychosozialen Erklärungsmodell (siehe Abbildung 9). Der Effekt von Religiosität auf Lebenszufriedenheit ist dabei indirekt, über Mediatoren vermittelt, konzeptualisiert.<sup>48</sup> Dabei wird die

---

<sup>48</sup> Sowohl die Genetik eines Menschen, vor allem bemerkbar etwa in seinen Persönlichkeitseigenschaften oder Prädispositionen für bestimmte Krankheiten (Depressionen etc.), als auch die Umwelt in Form von Lebensereignissen (Trennungen, familiäre Tragödien, Arbeitsplatzverlust etc.) haben einen moderierenden Einfluss auf die Beziehung.

Wichtigkeit der überschneidungsfreien Konzeptualisierung deutlich, da viele breite Konzeptualisierungen Aspekte einschließen, die eher Mediatorvariablen entsprechen.

**Abbildung 9: Vollständiges Erklärungsmodell**



**Anmerkung:** Eigene Abbildung.

Der kulturelle Kontext wird in dieser Konzeptionierung mit zwei Proxyvariablen erfasst: zum einem mit dem Human Development Index (HDI) und zum anderen mit dem aggregierten Wert der Variable „Wichtigkeit von Religion“ aus dem World Value Survey. Diese Variablen dienen zur Exemplifizierung der Konzeptionierung, da sie leicht zu erheben sind.

Beim HDI handelt es sich um einen leicht zu konstruierenden Wohlstandindikator, der neben Einkommen auch Bildung und Lebenserwartung eines Landes berücksichtigt (McGillivray 1991; Sagar/Najam 1998). Nichtsdestotrotz handelt es sich um eine vielfach verwendete Proxy-Messung von Wohlstand, die sich aufgrund ihrer Einfachheit für die nachfolgende meta-analytische Untersuchung eignet. Ein hoher HDI-Wert geht definitionsgemäß mit einem hohen Grad an menschlicher Entwicklung einher. Auf eine ähnliche Art wird bei der Variable „Wichtigkeit von Religion“ verfahren. Die Werte dieser Variable werden aus den unterschiedlichen Wellen des World Value Survey entnommen.<sup>49</sup> Dabei werden jeweils diejenigen Werte den Stichproben zugeordnet, die entweder dem Erhebungszeitpunkt der Daten oder wenn dieser nicht vorliegt, dem Publikationsdatum der Studie am nächsten liegt. Im World Value Survey wird die Variable als Teil einer Itembatterie über verschiedene Lebensbereiche erhoben: „For each of the following, indicate how important it is in your life. Would you say it is: Religion“. Die Befragten konnten ihre Antwort auf einer vierstufigen Skala von „very important“, „rather important“, „not very important“ und „not at all important“ abstimmen. Die Prozentanzahl der Zustimmung

<sup>49</sup> Siehe: <http://www.worldvaluessurvey.org/WVSONline.jsp>, aufgerufen am 25.06.2018.

für „very important“ wird hierbei als aggregierter Wichtigkeitswert für die Meta-Analyse übernommen.

Das Untersuchungsmodell im Gegensatz zum Erklärungsmodell, anhand dessen die modellbasierte Meta-Analyse durchgeführt werden kann, folgt jedoch erst nach der vollständigen Erhebung der Studien sowie deren Kodierung, da erst dann abzusehen ist, für welche der Mediatorpfade ausreichend empirisches Material vorliegt. Dies ist auch im Hinblick auf die statistische Analyse notwendig, da die einzusetzenden Verfahren bei einer großen Anzahl an leeren Zellen Konvergenzprobleme aufweisen. Es ist also zunächst einmal eine empirische Frage, welche Pfade untersucht werden können. Dies ist Ausdruck des explorativen Charakters dieser Untersuchung. Die Definitionen der eingeschlossenen Mediatoren ergeben sich aus den operationalen Definitionen während des Erhebungsprozesses. Das Erklärungsmodell auf der Individualebene ist eingebettet in den kulturellen Kontext und es werden Wechselwirkungen zwischen Mikro- und Makroebene vermutet.

## **4 Forschungsdesign**

In den folgenden Abschnitten wird das Forschungsdesign der modellbasierten Meta-Analyse festgelegt. Das Forschungsdesign beruht auf den Erfahrungen und Desiderata der methodologischen Praxis und den empfohlenen meta-analytischen Vorgehensweisen aus den klassischen meta-analytischen Methodenhandbüchern (Card 2012; Cooper 2010; Cooper/Hedges/Valentine 2009; Hammerstrøm et al. 2016; Urban/Fiebig 2015).

Zunächst wird ein methodologischer Review der meta-analytischen Praxis in diesem Forschungsfeld angefertigt (siehe Kapitel 4.1). Daran anknüpfend wird das Forschungsdesign entwickelt (Kapitel 4.2 bis 4.5). Dazu zählt zunächst die Festlegung von Selektionskriterien, anhand derer entschieden wird, welche Primärstudien für die Meta-Analyse relevant sind. Diese Selektionskriterien definieren gemeinsam mit der Definition der zentralen Konstrukte das Studienuniversum potenziell relevanter Studien. Anschließend wird eine Literatursuchstrategie entwickelt, um möglichst alle für die Forschungsfrage relevanten empirischen Studien, unter Berücksichtigung der Schwierigkeiten beim Auffinden von grauer Literatur, zu finden. Daran anknüpfend wird ein Screeningverfahren entwickelt, welches sicherstellt, dass nur die im Sinne der formulierten Selektionskriterien relevanten empirischen Studien in die Untersuchung aufgenommen werden. Dann folgen die Inhalte des Kodierschemas mit den zugehörigen Kodierregeln. Mit deren Hilfe werden die gefundenen Studien im Hinblick auf die interessierenden Merkmale (bspw. methodische Studiencharakteristiken) kodiert. Diese Studiencharakteristika können, wie bei der herkömmlichen Meta-Analyse, anschließend für die Durchführung von Moderatoranalysen zur Erklärung der Variabilität der Effektstärkekoeffizienten genutzt werden (Becker 2009: 386). Die spezifische statistische Methodik für modellbasierte Meta-Analysen wird gesondert in Kapitel 5 behandelt, da diese im Mittelpunkt dieser Arbeit steht.

### **4.1 Methodischer Review der meta-analytischen Praxis**

Im Folgenden wird die methodologische Praxis der bislang durchgeführten Meta-Studien über den Zusammenhang von Religiosität/Spiritualität und subjektivem Wohlbefinden/Lebenszufriedenheit evaluiert. Um einen Eindruck über die methodisch-statistische Praxis in diesem Forschungsbereich sowie der Dokumentationspraxis zu erlangen, wird diese anhand des idealtypischen Ablaufs einer Meta-Analyse betrachtet (Cooper/Hedges 2009b: 9). Das Ende dieses Kapitels stellt eine Zusammenfassung der meta-analytischen empirischen Praxis in diesem Forschungsfeld sowie eine Diskussion der vorliegenden Desiderata dar. Die mittels systematischer

Datenbanksuche gefundenen 34 Meta-Studien werden nun systematisch gesichtet.<sup>50</sup> Zunächst folgt eine Gesamtübersicht über alle berücksichtigten Meta-Studien und deren zugrundeliegenden statistischen Analysemodelle sowie den jeweiligen Literatursuchzeitraum (siehe Tabelle 3).

Den ersten Bereich bilden vier Meta-Studien, die die beiden Untersuchungskonzepte direkt miteinander in Beziehung setzen (Eger/Maridal 2015; Sawatzky/Ratner/Chiu 2005; Witter et al. 1985; Yonker/Schnabelrauch/Dehaan 2012). Fasst man das der Arbeit zugrunde liegende Erkenntnisinteresse eng, so sind lediglich drei Studien für die genauere Durchsicht relevant, da die Studie von Eger/Maridal (2015) vorwiegend das Wohlbefinden fokussiert. Yonker/Schnabelrauch/Dehaan (2012) untersuchen die Auswirkungen von Religiosität/Spiritualität auf verschiedene psychologische Outcomevariablen bei Jugendlichen und jungen Heranwachsenden. Sawatzky/Ratner/Chiu (2005) untersuchen den Zusammenhang von Spiritualität mit dem breiteren Konzept Lebensqualität. Die Studie von Witter et al. (1985) ist hingegen die einzige Studie, die eine identische Forschungsfrage untersucht.

Der zweite Bereich besteht aus Studien, die den Zusammenhang von R/S und „Gesundheit“ im weiteren Sinne meta-analytisch untersuchen: körperliche Gesundheit (Jim et al. 2015), mentale Gesundheit (Bergin 1983; Hackney/Sanders 2003; Salsman/Pustejovsky et al. 2015; Smith/McCullough/Poll 2003), soziale Gesundheit (Sherman et al. 2015), Sterblichkeit/Lebensdauer (Chida/Stephoe/Powell 2009; McCullough et al. 2000; Shor/Roelfs 2013) und Suizid (Wu/Wang/Jia 2015).

Im dritten Bereich lassen sich Studien verorten, die den Zusammenhang zwischen Religiosität und anderen Konzepten untersuchen. Dazu gehören die Zusammenhänge von intrinsischer und extrinsischer Religiosität (Donahue 1985), Religiosität und Persönlichkeit (Saroglou 2002), Religiosität und Werte (Saroglou/Delpierre/Dernelle 2004), Religiosität und Intelligenz (Zuckerman/Silberman/Hall 2013), Religiosität und Ehe/Elternschaft (Mahoney et al. 2001), Religiosität und Kriminalität (Baier/Wright 2001), religiöses Coping und psychologischer Anpassung (Ano/Vasconcelles 2005) sowie eine Studie zur Effektivität von spirituellen Therapieformen (Smith/Bartz/Scott Richards 2007).

---

<sup>50</sup> Neben den 28 hier dargestellten Meta-Studien werden sechs weitere zu illustrativen Zwecken ebenfalls für diesen Überblick verwendet. Diese sind in thematisch verwandten Kontexten angesiedelt und können über die methodische Praxis in dem Forschungsfeld Auskunft geben (Goyal et al. 2014; Lamers et al. 2012; Masters/Spielmanns 2007; Pinquart 2002; Prati/Pietrantonio 2009; Sedikides/Gebauer 2009). Diese werden allerdings aus Gründen der Übersichtlichkeit in den Tabellen nicht ausgewiesen.

**Tabelle 3: Basisinformationen der gefundenen Meta-Studien**

	Studie	Konzept 1	Konzept 2	Zeitraum	Modell
1) R/S & SWB	Witter et al. 1985	R	SWB	bis 1979	ME
	Sawatzky et al. 2005	S	Lebensqualität	ab 1991	ME
	Yonker et al. 2012	R/S	psych. Outcomes	1990 – 2010	ME
	Eger/Maridal 2015	R	SWB	bis 2013	ME
2) R/S & Gesundheit	Bergin 1983	R	psych. Gesundheit	bis 1979	FE
	McCullough et al. 2000	R	Sterblichkeit	bis 1999	MEMA
	Hackney/Sanders 2003	R	psych. Gesundheit	1990 – 2001	ME
	Smith et al. 2003	R	Depression	bis 2000	ME
	Chida et al. 2009	R/S	Sterblichkeit	bis 2008	ME
	Shor/Roelfs 2013	R	Lebenserwartung	bis 2009	ME
	Jim et al. 2015	R/S	körp. Gesundheit	bis 2013	GEE
	Salsman et al. 2015	R/S	psych. Gesundheit	bis 2013	GEE
	Sherman et al. 2015	R/S	soziale Gesundheit	bis 2013	GEE
Wu et al. 2015	R/S	Suizid	2000 – 2015	ME	
3) R/S & X	Donahue 1985	R	Extrinsische R.	bis 1982	FE
	Baier/Wright 2001	R	Kriminalität	bis 1998	ME
	Mahoney et al. 2001	R	Ehe/Elternschaft	1980 – 1999	ME
	Saroglou 2002	R	Persönlichkeit	bis 2000	FE
	Saroglou et al. 2004	R	Werte	bis 2003	ME
	Ano/Vasconellas 2005	relig. Coping	Stressanpassung	-	FE
	Smith et al. 2007	relig. Therapie	Besserung	1981 – 2005	ME
	Zuckerman et al. 2013	R	Intelligenz	-	ME
4) X & SWB	Okun et al. 1990	Interventionen	SWB	bis 1987	FE
	DeNeve/Cooper 1998	Persönlichkeit	SWB	bis 1995	FE
	Pinquart/Sörensen 2000	SÖS/soziales Netzwerk/ Kompetenz	SWB	bis 1999	ME
	Fischer/Boer 2011	Geld/ Autonomie	Nationales WB	1997 – 2005	MEMA
	Weis/Speridakos 2011	Interventionen	SWB/Hoffnung/ Verzweiflung	1994 – 2011	ME
	Capaldi et al. 2014	Naturverbundenheit	SWB/Vitalität	bis 2013	ME

**Anmerkungen:** FE=Fixed-Effects-Modell; RE=Random-Effects-Modell; ME=Mixed-Effects-Modell (mit Moderatoren); MEMA=Mehrebenen-Meta-Analyse. Aus Darstellungsgründen wird der fünfte Bereich hier nicht ausgewiesen.

Der vierte Bereich umfasst Studien, die sich mit dem Konzept des subjektiven Wohlbefindens auseinandersetzen. Hierzu zählen zwei Studien, die den Einfluss verschiedener (psychotherapeutischer) Interventionen auf das subjektive Wohlbefinden untersuchen (Okun/Olding/Cohn 1990; Weis/Speridakos 2011), eine Studie über den Zusammenhang von langfristigen Persönlichkeitsfaktoren und SWB (DeNeve/Cooper 1998), eine nationenübergreifende Studie über die Frage, ob Autonomie oder Geld wichtiger für das Wohlbefinden ist (Fischer/Boer 2011), eine ähnliche Studie über den Einfluss des sozioökonomischen Status, des sozialen Netzwerks und der Kompetenz auf das subjektive Wohlbefinden (Pinquart/Sörensen 2000) sowie eine Studie über den Zusammenhang von Naturverbundenheit und Glück (Capaldi/Dopko/Zelenski 2014).

Im fünften Bereich hingegen sind Studien zu finden, die lediglich in ähnlichen thematischen Kontexten angefertigt wurden. Diese Studien wurden in den Überblick aufgenommen, um einen besseren Eindruck über die generelle meta-analytische methodische Praxis in diesem Forschungsbereich zu erlangen und werden aus Platzgründen in den Tabellen nicht ausgewiesen.<sup>51</sup>

Die gefundenen Meta-Analysen sind vorwiegend in einem psychologischen (Hackney/Sanders 2003; Yonker/Schnabelrauch/Dehaan 2012) und medizinischen (McCullough et al. 2000; Wu/Wang/Jia 2015), aber auch sozialwissenschaftlichen (Eger/Maridal 2015; Shor/Roelfs 2013) Forschungskontext entstanden (siehe Tabelle 3). Dies unterstreicht deutlich die Multidisziplinarität des Forschungsfeldes. Allerdings sind die sozialwissenschaftlichen gegenüber den humanwissenschaftlichen Arbeiten deutlich in der Unterzahl.

Der Großteil der Studien verfolgt dabei (implizit) den Ansatz, alle relevanten Studien, also alle Primärstudien, die den Inklusionskriterien entsprechen, mit in die Untersuchung aufzunehmen. Bei vier Studien war es nicht möglich herauszufinden, welcher Such- bzw. Erhebungszeitraum den Studien zugrunde gelegt wurde. Lediglich acht Studien haben den Such- bzw. Inklusionszeitraum auf eine bestimmte Zeitspanne eingeschränkt. Diese Zeiträume sind nicht theoretisch-inhaltlich begründet (Ausnahme: (Mahoney et al. 2001)), sondern vermutlich aus pragmatischen Gründen gewählt (Wilson 2009: 163).

Betrachtet man das jeweilige zugrunde gelegte statistische Modell, zeigt sich eine große Ähnlichkeit über alle Studien hinweg. Nur vergleichsweise wenige Studien modellieren ein Fixed-Effects-Modell. Die meisten Studien verwenden somit ein Random-Effects- oder ein Mixed-Effects-Modell und modellieren die Variation zwischen den Studien (Heterogenität) explizit (Random-Effects) bzw. versuchen mithilfe von Moderatoren diese Variabilität statistisch zu erklären (Mixed-Effects-Modell). Drei der Meta-Studien, die in einem gemeinsamen Projekt entstanden sind (Salsman/Fitchett et al. 2015), nutzen den sogenannten Generalized Estimating Equation (GEE) Ansatz. Lediglich zwei Studien führen eine Mehrebenen-Meta-Analyse (MEMA) durch (Fischer/Boer 2011; McCullough et al. 2000), welche die hierarchische Daten-

---

<sup>51</sup> Die Inhalte dieser Studien sind deshalb in einem weiter gesteckten Themenfeld angesiedelt. Eine Studie etwa untersucht den Zusammenhang von emotionalem Wohlbefinden auf die Langzeitrehabilitation (Lamers et al. 2012), eine weitere beschäftigt sich mit den Auswirkungen unterschiedlicher Meditationsprogramme auf psychischen Stress und Wohlbefinden (Goyal et al. 2014). Pinquart befasst sich mit dem Zusammenhang von Lebenssinn und einer Vielzahl von Outcomevariablen (Pinquart 2002), Masters und Spielmans untersuchen in ihrem Beitrag die Wirksamkeit von (Fürbitt-)Gebeten auf die Gesundheit (Masters/Spielmans 2007). Weitere Studien in diesem Bereich untersuchen den Zusammenhang von Optimismus, sozialer Unterstützung und unterschiedlichen Coping-Strategien mit dem psychologischen Konzept posttraumatisches Wachstum (Prati/Pietrantonio 2009) sowie die Frage, ob Religiosität ein Mittel zur Selbststeigerung ist (Sedikides/Gebauer 2009).

strukturen von Individuen und Studien statistisch berücksichtigt (Fiebig/Urban 2014). Insgesamt haben somit fünf der betrachteten Studien das Problem abhängiger Effektstärken mit der Verwendung entsprechender Verfahren berücksichtigt.

Die Dokumentation des Vorgehens hat sich in Einklang mit den immer systematischer werdenden Reporting-Standards für Meta-Analysen seit den 1980er Jahren verbessert. So sind die methodologischen Angaben bei den frühen Meta-Analysen eher dürftig (Bergin 1983; Donahue 1985; Witter et al. 1985). Bei den neueren Studien (Salsman/Pustejovsky et al. 2015) hingegen werden explizit Suchstrategien, verwendete Zitationsdatenbanken sowie Suchbegriffe nachvollziehbar und begründet dargelegt. Damit einhergehend sind die Ergebnisse dieser Studien leichter zu replizieren und entsprechen damit eher dem wissenschaftlichen Ideal (Cooper/Hedges 2009b: 4). Die statistischen Modelle haben sich im untersuchten Forschungsfeld über die Jahre hinweg jedoch kaum gewandelt.

#### **4.1.1 Literatursuche**

Im nachfolgenden Abschnitt werden die gefundenen Meta-Studien im Hinblick auf ihre Literatursuchstrategien betrachtet. Dabei geht es vorwiegend darum, nachzuvollziehen, welche Strategien zur Suche nach relevanter Primärliteratur angewandt wurden. Vor allem bei den älteren Studien sind die Aussagen über den genauen Ablauf der Literatursuche vage oder überhaupt nicht dokumentiert (Bergin 1983; Donahue 1985). Nahezu alle Studien haben nicht nur eine einzelne Strategie verwendet, sondern mehrere Suchprozedere miteinander verknüpft.

Einige Meta-Studien haben bereits vor der eigentlichen Literatursuche bekannte empirische Studien in die Untersuchung aufgenommen (Saroglou 2002; Saroglou/Delpierre/Dernelle 2004; Weis/Speridakos 2011). Dies ist allerdings für eine systematische Literatursuche eher ungeeignet, da davon auszugehen ist, dass mit diesem Vorgehen bestimmte Studien überrepräsentiert sind und so die Gefahr für verzerrte Ergebnisse steigt. Darüber hinaus haben nur wenige Autoren unsystematische, händische Suchen in Bibliotheken durchgeführt (Pinquart 2002; Pinquart/Sörensen 2000; Witter et al. 1985).

Eine zentrale Strategie bei der Literatursuche ist die Verwendung der Primärliteratur von älteren, bereits bekannten Meta-Studien bzw. (narrativen) Literaturreviews (DeNeve/Cooper 1998; McCullough et al. 2000; Prati/Pietrantonio 2009; Smith/Bartz/Scott Richards 2007; Weis/Speridakos 2011). Hierbei ist zu berücksichtigen, dass auch potenzielle Fehler bei der Erhebung der Primärstudien weitergetragen werden. Über diese Vorgehen hinaus haben manche Studien (Ano/Vasconcelles 2005; Sawatzky/Ratner/Chiu 2005) durch die Verwendung besonders zentraler Arbeiten, unter Zuhilfenahme von Zitationsdatenbanken, Vorwärtssuchen in die Zukunft

durchgeführt („ancestry-approach“). Alle Studien, mit der Ausnahme von Okun/Olding/Cohn (1990), die ihre Literatursuche ausschließlich in bekannten Journals durchführten, bedienen sich der elektronischen Zitationsdatenbanksuche. Manche Studien führen darüber hinaus eine händische Suche in den zuvor identifizierten zentralen Journals durch (DeNeve/Cooper 1998; Hackney/Sanders 2003; Mahoney et al. 2001; Zuckerman/Silberman/Hall 2013). Viele Autoren durchsuchen die Literaturverzeichnisse der gefundenen Studien, um weitere relevante Studien mittels „footnote-chasing“ zu finden (Baier/Wright 2001; Capaldi/Dopko/Zelenski 2014; Fischer/Boer 2011; Hackney/Sanders 2003; Mahoney et al. 2001; McCullough et al. 2000; Sawatzky/Ratner/Chiu 2005; Shor/Roelfs 2013; Smith/McCullough/Poll 2003; Zuckerman/Silberman/Hall 2013).

Einige Autoren haben darüber hinaus zur Vervollständigung der Literatursuche potenziell „wichtige“ Kollegen im Forschungsfeld kontaktiert und diese um Vervollständigung ihrer Literaturlisten gebeten (DeNeve/Cooper 1998; McCullough et al. 2000; Prati/Pietrantonio 2009). Auch Anfragen in speziellen E-Maillisten, etwa von Fachgesellschaften, wurden von einigen Studienautoren durchgeführt, um vor allem nicht publizierte Studien bzw. Konferenzvorträge zu erlangen (Capaldi/Dopko/Zelenski 2014; Jim et al. 2015; Salsman/Pustejovsky et al. 2015; Sherman et al. 2015).

Bezüglich der Vielfältigkeit der Literatursuche ist die empirische Praxis der berücksichtigten Meta-Analysen angemessen, da jeweils verschiedene Verfahren zur Erhebung von Primärstudien verwendet werden. Die Verknüpfung unterschiedlicher Suchverfahren kann die Nachteile einzelner Methoden ausgleichen.

Nun folgt eine Übersicht über die wichtigste Literatursuchform (Card 2012: 46f.), die Suche in Fach- bzw. Zitationsdatenbanken (siehe Tabelle 4). Trotz der Vielzahl der unterschiedlichen Forschungsfelder, die in dieser Übersicht betrachtet werden, gibt es in Bezug auf die verwendeten Datenbanken große Überschneidungen. Am häufigsten verwendet werden folgende Fach- bzw. Zitationsdatenbanken: PsycINFO (23 von 34 Meta-Studien), gefolgt von MEDLINE (10 von 34), CINAHL (Cumulative Index to Nursing and Allied Health Literature) und SSCI (Social Science Citation Index) mit jeweils acht von 34 sowie ERIC (Education Resources Information Center) und PubMed mit jeweils sechs von 34. Die restlichen Datenbanken werden seltener verwendet und verteilen sich entsprechend der unterschiedlichen Fachdisziplinen auf die jeweiligen Studien.

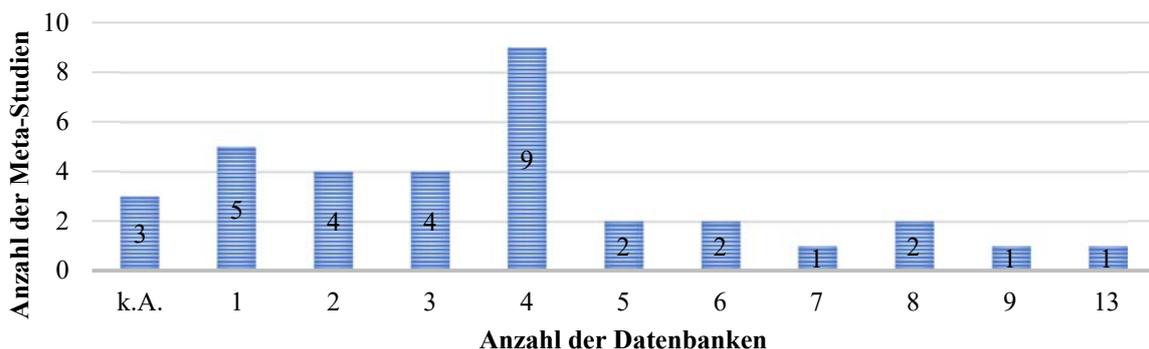
**Tabelle 4: Verwendete Zitationsdatenbanken**

Bereich	Studie	N	Zitationsdatenbanken
1) R/S & SWB	Witter et al. 1985		-
	Sawatzky et al. 2005	8	CINAHL, PubMed EMBASE, HEALTHSTAR, PsycINFO, Social Science Citation Index, Science Citation Index Expanded, Arts and Humanities Index
	Yonker et al. 2012	9	PsycINFO, PsychARTICLES, ALTA, Christian Periodical Index Online, Catholic Periodical & Literature Index Online, Social Science Abstracts, Dissertation Abstracts, Web of Knowledge, Google Scholar
	Eger/Maridal 2015	6	JSTOR, ProQuest, Web of knowledge, Dissertation Abstracts International, Google Scholar, Database of Happiness
2) R/S & Gesundheit	Bergin 1983		-
	McCullough et al. 2000	6	MEDLINE, PsycINFO, Sociofile, CINAHL, ERIC, Dissertation Abstracts
	Hackney/Sanders 2003	4	PsycINFO, Social Sciences Abstracts, MEDLINE, Humanities Abstracts
	Smith et al. 2003	13	CINAHL, ERIC, HealthSTAR, MEDLINE, Mental Health Abstracts, PsycINFO, Religion Index, Programme Appliqué à la Sélection et la Compilation Automatique de Littérature, Social Science Abstracts, Social SciSearch, Sociological Abstracts via SocioFile, Social Work Abstracts, TGG Health & wellness database
	Chida et al. 2009	4	MEDLINE, PsycINFO, Web of Science, PubMed
	Shor/Roelfs 2013	5	Web of Knowledge, PubMed, Google Scholar, EMBASE, CINAHL
	Salsman et al. 2015	4	PubMed, PsycINFO, CINAHL, Cochrane Library
	Jim et al. 2015	4	PubMed, PsycINFO, CINAHL, Cochrane Library
	Sherman et al. 2015	4	PubMed, PsycINFO, CINAHL, Cochrane Library
	Wu et al. 2015	2	PubMed, Web of Science
3) R/S & X	Donahue 1985	1	Social Science Citation Index
	Baier/Wright 2001	4	Sociological Abstracts, PsycINFO, Social Science Citation Index, Academic Universe
	Mahoney et al. 2001	3	PsycINFO, SocioInfo, Social Science Citation Index
	Saroglou 2002	1	PSYCLIT
	Saroglou et al. 2004	1	PsycINFO
	Ano/Vasconellas 2005	1	PsycINFO
	Smith et al. 2007	4	PsycINFO, Social Science Abstracts, ERIC, ALTA
	Zuckerman et al. 2013	3	PsycINFO, Google Scholar, Archive for Psychology of Religion
4) X & SWB	Okun et al. 1990	8	Psychological Abstracts, Sociological Abstracts, ERIC, National Technical Information System, Dissertation Abstracts, Dissertation Abstracts international, Social Science Citation Index, National Institute of Mental Health
	DeNeve/Cooper 1998	4	PsychLIT, Social Science Citation Index, Dissertation Abstracts, ERIC
	Pinquart/Sörensen 2000	3	PsycINFO, MEDLINE, PSYINDEX
	Fischer/Boer 2011	1	PsycINFO
	Weis/Speridakos 2011	5	Academic Search Complete, ERIC, MEDLINE, PsycINFO, Dissertation Abstract Database
	Capaldi et al. 2014	2	PsycINFO, Dissertation and Theses Full Text

Die Anzahl der pro Studie verwendeten Datenbanken schwankt erheblich (Abbildung 10). Drei Studien haben keine Angaben gemacht, welche Datenbanken sie zur Suche verwendet haben. Manche Studienautoren verlassen sich darüber hinaus bei ihrer Suche, entgegen den Empfehlungen (Reed/Baxter 2009: 90), auf nur eine einzige Datenbank, wohingegen Smith/McCullough/Poll (2003) das Maximum von 13 unterschiedlichen Datenbanken nutzen. Insgesamt verwenden die meisten Autoren mehrere Zitationsdatenbanken (Pinquart/Sörensen 2000). Dies entspricht etwa 3,7 Datenbanken pro Studie.

Ein Großteil der Studien hat sowohl medizinische (MEDLINE oder PubMed) als auch psychologische Datenbanken (PsycINFO oder PSYINDEX) genutzt. Aber auch fachbereichsübergreifende Datenbanken (SCOPUS oder Web of Science) werden von vielen Autoren verwendet (Chida/Steptoe/Powell 2009; Goyal et al. 2014; Sedikides/Gebauer 2009). Besonders relevant, um den sogenannten „publication bias“ vorzubeugen, sind jene Datenbanken, die es erlauben, sogenannte „graue Literatur“ in Form von nicht publizierten Abschlussarbeiten und Dissertationen zu finden.

**Abbildung 10: Anzahl der Zitationsdatenbanken**



**Anmerkung:** Eigene Abbildung.

Diese Datenbanken (etwa Dissertation Abstracts oder Dissertations and Theses Full Text) werden von einigen Autoren berücksichtigt. Darüber hinaus wird das frei zugängliche Google Scholar in einigen der Studien genutzt. Humanwissenschaftliche Arbeiten (Goyal et al. 2014; Jim et al. 2015; Salsman/Pustejovsky et al. 2015; Sherman et al. 2015) nutzten auch die für Meta-Analysen angelegte Cochrane Library.

Im Anschluss an die Übersicht des Literatursuchprozedere sowie der verwendeten Zitationsdatenbanken folgt ein Überblick über die verwendeten Suchbegriffe (Card 2012: 44ff.; White 2009: 62ff.). Die Suchbegriffe unterscheiden sich inhaltlich in den fünf Bereichen. In Tabelle 5 sind die inhaltlich relevanten Suchstrings ausgewiesen.

Bis auf Witter et al. (1985), die leider keinerlei Angaben über die verwendeten Suchbegriffe machen, haben alle Studien im ersten Bereich ihre Suchstrings dokumentiert. Sawatzky/Ratner/Chiu (2005) verwenden, entsprechend ihrer weiten Konzeptualisierung („existential“), sehr differenzierte Suchbegriffe für Religiosität/Spiritualität. Sie verknüpfen diese mit den drei Konzepten „quality of life“, „well-being“ und „life satisfaction“. Im Gegensatz dazu wird bei Yonker/Schnabelrauch/Dehaan (2012) nur mit den drei Begriffen „religion“, „spirit“ und „faith“ nach Religiosität gesucht. Da es in dieser Studie um viele verschiedene Outcomes geht, sind entsprechend viele Suchbegriffe dafür vorgesehen. Eger/Maridal (2015) verwenden lediglich Suchstrings zu SWB. Dabei nutzen sie mehrere Kombinationen der bereits bei Sawatzky/Ratner/Chiu (2005) eingesetzten Begriffe. Zusätzlich suchen sie auch nach „emotional wellbeing“.

Betrachtet man die Suchstrings für R/S in den Bereichen 2 und 3 so zeigen sich kaum Auffälligkeiten. Rechts-Trunkierung wird von vielen Studien zur Vereinfachung der Suche eingesetzt (Chida/Stephoe/Powell 2009; Smith/McCullough/Poll 2003). Der Großteil der Studien verwendet die Begriffe „religion“, „religiousness“, „religiosity“, „spirituality“ und Abwandlungen davon. Auf Religiosität fokussiert, aber dafür mit entsprechender konzeptioneller Klarheit, suchen lediglich drei Studien (Mahoney et al. 2001; Saroglou 2002; Saroglou/Delpierre/Dernelle 2004). Auffällig sind die Begriffe bei Smith/McCullough/Poll (2003), die auch „church“, „mosque“, „synagogue“ und „temple“ sowie „worship“ und „prey“ bei der Suche verwenden. Chida/Stephoe/Powell (2009) haben dies noch um den String „meditation“ erweitert und nutzen somit eine breite Konzeptionalisierung von Religiosität. Humanwissenschaftliche Arbeiten, die sich vorwiegend mit mentaler Gesundheit auseinandersetzen, nutzen auch subjektives Wohlbefinden als Messinstrument für Gesundheit (Hackney/Sanders 2003; Smith/McCullough/Poll 2003).

Der vierte Bereich ist in Bezug auf die Suchbegriffe heterogener. Passt etwa die Konzeption von DeNeve/Cooper (1998) noch zur zugrunde gelegten Zielsetzung dieser Arbeit, ist hingegen die Konzeptionierung von Okun/Olding/Cohn (1990) mit dem Suchbegriff „morale“ unschärfer. Auch die anderen Studien in diesem Bereich legen unterschiedliche Suchbegriffe entsprechend der jeweiligen Fragestellung an. Besonders interessant ist hierbei das Vorgehen von Fischer/Boer (2011), die sich auf wenige formalisierte Messinstrumente zur Erfassung von Gesundheit fokussieren. Ebenso auffällig ist, dass Capaldi/Dopko/Zelenski (2014), die die Beziehung von Naturverbundenheit und Glück untersuchen, lediglich Suchbegriffe zur Naturverbundenheit konkretisieren.

**Tabelle 5: Inhaltliche Suchstrings**

Bereich	Studie	Suchbegriffe R/S	Suchbegriffe SWB
1) R/S & SWB	Witter et al. 1985	-	-
	Sawatzky et al. 2005	Spiritual(ity), religion, religiosity, religiousness, transcendence, existential, transpersonal, sacredness	Quality of life, well-being, life satisfaction
	Yonker et al. 2012	Religion, spirit, faith	Personality, identity, sexual activity, depression, anxiety, anger, well-being, self-esteem, smoking, alcohol, drugs, substance use, deviant behavior, coping, happiness, social support, altruism, empathy
	Eger/Maridal 2015	-	Life satisfaction, satisfaction with life, happiness, quality of life, emotional wellbeing, subjective wellbeing, wellbeing
2) R/S & Gesundheit	Bergin 1983	-	-
	McCullough et al. 2000	Religion, religiousness, religiosity, religious	Mortality, fatality, death, survival
	Hackney/Sanders 2003	Religiosity, religion, religiousness	Mental health, depression, happiness, life satisfaction, self-esteem
	Smith et al. 2003	Relig*, spirit*, church, mosque, synagogue, temple, worship, pray	Depress*, affective disorder, mood, affect
	Chida et al. 2009	Relig*, spirit*, church, mosque, synagogue, temple, worship, pray, meditation	Mortality, survival
	Shor/Roelfs 2013	-	-
	Jim et al. 2015	Religio*, spitual*	Cancer, health outcomes
	Salsman et al. 2015	Religio*, spitual*	
	Sherman et al. 2015	Religio*, spitual*	
Wu et al. 2015	Religion, religious, religiosity, spiritual, spirituality	Suicide, suicidality, suicidal attempt	
3) R/S & X	Donahue 1985	-	<i>Anderer Forschungsgegenstand</i>
	Baier/Wright 2001	-	
	Mahoney et al. 2001	Religion, religiosity	
	Saroglou 2002	Relig*	
	Saroglou et al. 2004	Relig*	
	Ano/Vasconellas 2005	Religion, religiosity, religious coping	
	Smith et al. 2007	Spirit, relig, faith, church, mosque, synagogue, temple, worship, pray	
	Zuckerman et al. 2013	Religion, spirituality, religiosity, religious belief	
4) X & SWB	Okun et al. 1990	<i>Anderer Forschungsgegenstand</i>	Life satisfaction, happiness, morale, well-being, perceived well-being
	DeNeve/Cooper 1998		Subjective well-being, happiness, life satisfaction, quality of life
	Pinquart/Sörensen 2000		-
	Fischer/Boer 2011		General Health Questionnaire/Spielberger State-Trait Anxiety Inventory, STAI / Maslach Burnout Inventory
	Weis/Speridakos 2011		Hope, hope theory, hope enhancement, hopefulness, hopelessness, positive psychological interventions
	Capaldi et al. 2014		Commitment to nature, connectedness to nature, connectivity to nature, emotional affinity toward nature, environmental identity, inclusion of nature in self, nature relatedness

**Anmerkungen:** Daneben haben die Autoren noch weitere, hier nicht berichtete, zumeist auf die Methodik bezogene Suchbegriffe verknüpft oder bestimmte Begriffe ausgeschlossen.

Einige der Studien machen keine Angaben darüber, welche Suchstrings sie zur Suche in Zitations- bzw. Fachdatenbanken verwenden. Dies verhindert die Nachvollzieh- bzw. Replizierbarkeit erheblich. Andere Studien hingegen machen nur grobe Angaben, die die Replizierbarkeit kaum verbessern (Jim et al. 2015; Salsman/Pustejovsky et al. 2015; Sherman et al. 2015). Lediglich zwei Studien haben die Datenbanksuchen von mehreren Personen und zu verschiedenen Zeitpunkten durchgeführt (Smith/McCullough/Poll 2003; Yonker/Schnabelrauch/Dehaan 2012). Dies bedeutet im Umkehrschluss, dass die Literatursuche in den meisten Fällen von einem einzigen Forscher durchgeführt wurde.

#### **4.1.2 Screening**

Der nächste Schritt im Ablauf einer Meta-Analyse ist das Screening der gefundenen Literatur. Aufgrund der größtenteils sehr differenzierten detailreichen Darstellung dieser Kriterien sowie deren starke inhaltliche Bezogenheit auf die jeweilige Forschungsfrage, sind in Tabelle 6 lediglich Auffälligkeiten zu sehen.

Zunächst lässt sich feststellen, dass nahezu alle Studien Kriterien dafür definieren, wann eine Primärstudie in die Meta-Analyse aufgenommen bzw. wann sie ausgeschlossen wird. Nur wenige der betrachteten Meta-Studien haben keine Kriterien zum Einschluss von Studien formuliert. Beinahe alle Studien haben festgelegt, dass eine Studie eingeschlossen wird, wenn sie die interessierende Beziehung untersucht sowie wenn ausreichend statistische Informationen zur Bildung von Effektstärken vorhanden sind. Ebenfalls schließen die meisten Meta-Studien alle theoretischen, qualitativen und nicht zur genauen Konzeptualisierung passenden Primärstudien aus. In einigen Arbeiten werden Studien ausgeschlossen, die als einzige R/S-Messung die Konfession der Befragten berücksichtigen. Ebenfalls werden in zwei Meta-Analysen diejenigen Studien ausgeschlossen, in denen nur die Ergebnisse von multivariaten Regressionen, Strukturgleichungsmodellen oder von Varianzanalysen berichtet wurden.

Auffällig ist die Berücksichtigung von Studien in fünf Sprachen bei Pinquart (2002). Außergewöhnlich ist auch die Exklusion von Studien aufgrund einer speziellen Stichprobengröße (Capaldi/Dopko/Zelenski 2014). Ein solches Kriterium erscheint ungeeignet, da dadurch Ergebnisverzerrungen entstehen können. Inhaltlich besonders interessant ist der Ausschluss von Studien, die lediglich die mentale Gesundheit bzw. domainspezifisches SWB berücksichtigen (Witter et al. 1985). Unglücklich, weil ebenfalls anfällig für statistische Verzerrungen, ist der Ausschluss von Studien, die nicht in begutachteten Zeitschriften erschienen sind (Ano/Vasconcelles 2005; Chida/Stephoe/Powell 2009). Eine solche Strategie des Ausschlusses erhöht nahezu

unweigerlich die Gefahr des Publication-Bias, da erfahrungsgemäß Studien mit nicht signifikanten Ergebnissen weniger häufig publiziert werden.

**Tabelle 6: Auffälligkeiten Inklusions- und Exklusionskriterien**

Bereich	Studie	Inkl.	Exkl.	Auffälligkeiten
1) R/S & SWB	Witter et al. 1985			Inklusion: Literatur aus den USA Exklusion: mentale Gesundheit; domainspezifisches SWB
	Sawatzky et al. 2005	X	X	Exklusion: nicht englischsprachig
	Yonker et al. 2012	X	X	Exklusion: Studien, die multiple Regression oder SEM durchgeführt haben
	Eger/Maridal 2015	X		Inklusion: nationales repräsentatives Sample
2) R/S & Gesundheit	Bergin 1983			
	McCullough et al. 2000		X	Exklusion: Konfession als einzige R.-Messung
	Hackney/Sanders 2003	X		Inklusion: nur enges Konzept von R.
	Smith et al. 2003	X	X	Exklusion: Konfession als einzige R.-Messung
	Chida et al. 2009	X	X	Inklusion: englischsprachig + peer-reviewed Journal Exklusion: Konfession als einzige R.-Messung
	Shor/Roelfs 2013	X	X	Exklusion: private R.-Messung / selbst eingeschätzte R.
	Jim et al. 2015	X	X	
	Salsman et al. 2015	X	X	
	Sherman et al. 2015	X	X	
	Wu et al. 2015	X	X	Inklusion: englischsprachig
3) R/S & X	Donahue 1985			
	Baier/Wright 2001	X		Inklusion: Zusammenhangsmessung vorhanden
	Mahoney et al. 2001	X	X	Exklusion: keine klare operationalen Definition von R.
	Saroglou 2002	X	X	Inklusion: Zusammenhangsmessung
	Saroglou et al. 2004			
	Ano/Vasconellas 2005	X		Inklusion: Artikel muss in einem Fachjournal erschienen sein
	Smith et al. 2007	X	X	Exklusion: nicht englischsprachig
	Zuckerman et al. 2013	X	X	Exklusion: indirekte Messung der Beziehung
4) X & SWB	Okun et al. 1990	X		Inklusion: US-Probanden
	DeNeve/Cooper 1998	X	X	Exklusion: multivariate Regression + Varianzanalyse
	Pinquart/Sörensen 2000	X	X	Inklusion: deutschsprachig / englischsprachig
	Fischer/Boer 2011	X		
	Weis/Speridakos 2011	X	X	Inklusion: englischsprachig
	Capaldi et al. 2014	X	X	Inklusion: experimentelle Studien mit Baseline Messung Exklusion: Studien, die Glück dichotomisiert haben; Samplegröße <10

Die empirische Praxis in Bezug auf das Screening der Primärliteratur ist insgesamt sehr heterogen. Einige Studien dokumentieren die entsprechenden Exklusions- und Inklusionskriterien sehr detailliert, während andere Meta-Studien diese nur vage beschreiben. Viele Studien haben aus pragmatischen Gründen nur englischsprachige Literatur sowie Journalartikel berücksichtigt und laufen damit Gefahr, Verzerrungen in ihren Untersuchungen zu produzieren (Card 2012:

83ff.; Rothstein/Hopewell 2009: 105ff.). Die Begründung der spezifizierten Kriterien fehlt in vielen Studien vollständig.

### 4.1.3 Anzahl der Studien

Mit dem beschriebenen Literatursuchprozedere und dem anschließenden Screening wurde in den betrachteten Meta-Analysen eine überaus heterogene Anzahl von Studien gefunden (Tabelle 7). Die Meta-Analyse mit der geringsten Anzahl an Studien umfasst gerade einmal neun Studien (Wu/Wang/Jia 2015), die mit der höchsten Anzahl an Primärstudien hingegen 286 (Pinquart/Sörensen 2000).

**Tabelle 7: Anzahl Primärstudien sowie Effekt- und Samplegröße**

Bereich	Studie	Studienanzahl	Effektstärken	Befragte
1) R/S & SWB	Witter et al. 1985	28	56	-
	Sawatzky et al. 2005	48	59	22.554
	Yonker et al. 2012	75	-	66.273
	Eger/Maridal 2015	164	-	560 (Länder/Einheiten)
2) R/S & Gesundheit	Bergin 1983	24	30	9.799
	McCullough et al. 2000	29	42	125.826
	Hackney/Sanders 2003	34	264	-
	Smith et al. 2003	147	-	98.975
	Chida et al. 2009	44 (gesund)/ 13(krank)	-	ca. 2.750 / ca. 485
	Shor/Roelfs 2013	74	312	>300.000
	Jim et al. 2015	101	497	>32.000
	Salsman et al. 2015	148	617	<39.000
	Sherman et al. 2015	78	227	14.277
	Wu et al. 2015	9	-	2.339 Suizide / 5.252 Vergleichsfälle
3) R/S & X	Donahue 1985	34 / 28	-	8.271 /4.952
	Baier/Wright 2001	60	79	-
	Mahoney et al. 2001	94	-	-
	Saroglou 2002	16	-	>3.000
	Saroglou et al. 2004	21	-	8.551
	Ano/Vasconellas 2005	49	105	13.512
	Smith et al. 2007	31	-	1.845
	Zuckerman et al. 2013	63	-	-
4) X & SWB	Okun et al. 1990	31	45	-
	DeNeve/Cooper 1998	142	1.538	42.171
	Pinquart/Sörensen 2000	286	-	-
	Fischer/Boer 2011	274 / 164 / 200	396 / - / 245	260.449 / 15.477 / 124.149
	Weis/Speridakos 2011	27	-	2.154
	Capaldi et al. 2014	30	-	8.523

Was die zugrundeliegenden Stichproben betrifft, gibt es ebenfalls eine große Spannweite. Die kleinste Meta-Studie beruht auf lediglich 560 geographischen Einheiten, wobei es sich um eine Makroanalyse handelt (Eger/Maridal 2015). Die größte Studie hingegen beruht auf über

300.000 Befragten (Shor/Roelfs 2013). Betrachtet man nun die Anzahl der Primärstudien in Relation zur Anzahl der verwendeten Effektstärken so fällt auf, dass einige der Meta-Studien mehrere Effektstärken pro Studie, etwa aufgrund mehrerer Outcomemessungen, berücksichtigen. Dies führt zu Verletzungen der Unabhängigkeitsannahme und muss bei der Modellierung beachtet werden (Borenstein et al. 2011: 225ff.).

Insgesamt haben alle Meta-Studien eine hohe Anzahl von Primärstudien aufgenommen und können somit von den Vorteilen dieser Methode in Bezug auf die höhere statistische Teststärke profitieren.<sup>52</sup> Vor allem die Studien mit spezieller Untersuchungsfrage (etwa Wirksamkeit von Fürbittgebeten auf die Gesundheit) haben Schwierigkeiten, eine umfassende Anzahl von Primärstudien zu finden (Masters/Spielmanns 2007; Wu/Wang/Jia 2015). Vor allem die neueren Meta-Studien basieren auf einer hohen Anzahl von Studien und einer entsprechend großen Stichprobe. Was das Verhältnis von Primärstudien zu den in die Untersuchung eingeschlossenen Effektstärken betrifft, gibt es ebenfalls sehr große Unterschiede zwischen den Meta-Analysen. Einige Studien haben durch entsprechende Verfahren sichergestellt, dass pro Primärstudie nur eine Effektgröße in die Untersuchung einfließt, während etwa Hackney/Sanders (2003) fast achtmal so viel Effektstärken berücksichtigen wie Primärstudien vorliegen.

In Bezug auf die Anzahl von Primärstudien, die Anzahl der analysierten Effektstärken sowie die Gesamtstichprobe lässt sich festhalten, dass große Heterogenität vorliegt. Leider ist die Dokumentation vieler Studien in Bezug auf die Stichprobengröße nicht zufriedenstellend. Somit ist es bei vielen Primärstudien im Einzelnen nicht möglich, die Anzahl der in die Untersuchung eingegangenen Effektgrößen sowie eine korrekte Stichprobengröße zu bestimmen (Baier/Wright 2001; Mahoney et al. 2001; Masters/Spielmanns 2007; Pinquart 2002; Prati/Pietrantonio 2009).

#### **4.1.4 Kodierverfahren**

Im Anschluss an die Übersicht der verwendeten Primärstudien folgt ein Überblick über das in den Meta-Studien angewandte Kodierverfahren und damit zusammenhängend ein Überblick über die angewandten Reliabilitätsüberprüfungen (Tabelle 8). Bei Durchsicht aller Meta-Studien wird deutlich, dass das Kodierverfahren ein ungenau dokumentierter Aspekt ist. Einige der berücksichtigten Studien haben keinerlei Angaben zur Kodierung sowie zur Reliabilitätsüberprüfung der Kodingsentscheidungen dokumentiert.

---

<sup>52</sup> Wie umfassend oder repräsentativ das jeweilige Primärstudienensemble im Hinblick auf die Grundgesamtheit an potenziell verfügbaren Primärstudien ist, kann jedoch an dieser Stelle nicht angemessen gewürdigt werden.

**Tabelle 8: Kodierverfahren und Reliabilitätsüberprüfung**

Bereich	Studie	Kodierverfahren	Reliabilitätsüberprüfung
1) R/S & SWB	Witter et al. 1985	Vier Koder; keine nähere Beschreibung	-
	Sawatzky et al. 2005	-	Diskrepanzen per Diskussion gelöst
	Yonker et al. 2012	Methodensektion und Ergebnisse getrennt voneinander kodiert; unabhängige Kodierung von S/R durch zwei Koder	Cohens Kappa; Diskrepanzen per Diskussion gelöst
	Eger/Maridal 2015	-	-
2) R/S & Gesundheit	Bergin 1983	-	-
	McCullough et al. 2000	-	Cohens Kappa; ICC
	Hackney/Sanders 2003	-	Stichprobe gezogen und durch Kollegen kodieren lassen; prozentuale Übereinstimmung
	Smith et al. 2003	Koder wurden intensiv geschult; Methodensektion und Ergebnisse getrennt voneinander kodiert; Kodierung durch zwei Koder	Cohens Kappa; ICC; Diskrepanzen per Diskussion gelöst
	Chida et al. 2009	Studieneinschluss, Qualität sowie Validität durch zwei Koder	Diskrepanzen per Diskussion gelöst
	Shor/Roelfs 2013	-	-
	Jim et al. 2015	Vier Zwei-Koder-Paare haben alle Studien unabhängig voneinander kodiert	Diskrepanzen per Diskussion gelöst
	Salsman et al. 2015	Vier Zwei-Koder-Paare haben alle Studien unabhängig voneinander kodiert	Diskrepanzen per Diskussion gelöst
	Sherman et al. 2015	Vier Zwei-Koder-Paare haben alle Studien unabhängig voneinander kodiert	Diskrepanzen per Diskussion gelöst
	Wu et al. 2015	-	-
3) R/S & X	Donahue 1985	-	-
	Baier/Wright 2001	-	-
	Mahoney et al. 2001	-	-
	Saroglou 2002	-	-
	Saroglou et al. 2004	-	-
	Ano/Vasconellas 2005	Unabhängige Kodierung durch zwei Koder	-
	Smith et al. 2007	Unabhängige Kodierung durch zwei Koder	Cohens Kappa, ICC; Diskrepanzen durch dritten Koder gelöst
Zuckerman et al. 2013	Effektgrößen durch zwei Koder verkodet; nur eine Variable mit subjektiver Einschätzung	Diskrepanzen per Diskussion gelöst	
4) X & SWB	Okun et al. 1990	Drei Koder; nur Studiencharakteristiken und Effektstärken doppelt verkodet	prozentuale Übereinstimmung
	DeNeve/Cooper 1998	Ein Autor hat alles kodiert; 10 % wurden erneut kodiert	prozentuale Übereinstimmung
	Pinquart/Sörensen 2000	-	-
	Fischer/Boer 2011	-	-
	Weis/Speridakos 2011	Unabhängige Kodierung durch zwei Koder	Cohens Kappa; Diskrepanzen per Diskussion gelöst
	Capaldi et al. 2014	Unabhängige Kodierung durch zwei Koder	Prozentuale Übereinstimmung; Cohens Kappa; ICC; Diskrepanzen per Diskussion gelöst

Studien, die ihr Vorgehen berichten, haben überwiegend alle Studien durch zwei Koder kodieren lassen. Teilweise wurden lediglich unterschiedliche Bereiche der Studien, etwa der Methoden- und Analyseteil, von unterschiedlichen Personen kodiert. Eine weitere Variante, die jedoch in den begutachteten Meta-Analysen nur ein einziges Mal angewandt wurde, besteht darin, dass nur eine Sub-Stichprobe vollständig doppelt kodiert wird. In der Studie von DeNeve/Cooper (1998) wird dies angewandt, indem eine Sub-Stichprobe von 10 % aller Studien durch einen zweiten Koder erneut kodiert wurde. Einige Studien haben auf eine Reliabilitätsprüfung verzichtet und auftretende Diskrepanzen per Diskussion zwischen den Kodern gelöst. Andere Meta-Studien haben mittels mathematischer Verfahren die Interraterreliabilität überprüft. Zusammengefasst entspricht das Kodierverfahren, über alle Studien hinweg betrachtet, dem am schlechtesten dokumentierten Bereich. Geben sich einige der Studien große Mühe, in diesem Bereich nachvollziehbar zu dokumentieren (Smith/Bartz/Scott Richards 2007; Yonker/Schnabelrauch/Dehaan 2012), machen andere Autoren keinerlei Angaben.

#### **4.1.5 Statistische Umsetzung**

Im nächsten Schritt folgt ein systematischer Überblick über die statistische Methodologie der Meta-Studien. Das grundlegende statistische Analysemodell wurde bereits zu Beginn besprochen (siehe Tabelle 3). An dieser Stelle werden verschiedene Aspekte der statistischen Konzeption verglichen.<sup>53</sup> Die folgenden sechs Bereiche werden näher betrachtet (siehe Tabelle 9):

1. die verwendeten Effektstärken,
2. die potenziellen Transformationen dieser Effektstärken,
3. die Berücksichtigung der Abhängigkeiten von Effektstärken aufgrund multipler Messungen der interessierenden Beziehungen oder aufgrund von mehreren Messzeitpunkten sowie deren angemessene Modellierung,
4. das potenzielle Gewichtungsverfahren,
5. die Durchführung eines Heterogenitätstests sowie
6. die Verfahren, die angewandt wurden, um den sogenannten Publikationsbias („file drawer“-Problematik) zu überprüfen.

Auch bei dieser Übersicht ist es vor allem bei den älteren Studien schwierig herauszufinden, welche statistischen Verfahren genutzt werden, da deren Dokumentation teilweise unzureichend ist. Die meisten Studien verwenden als Effektstärkemaß den Pearsonschen Korrelationskoeffizient  $r$ . Ein Großteil dieser Studien führt dementsprechend verschiedene Transformationen basierend auf Fishers  $z$ -Transformation durch. Lediglich diejenigen Studien, die sich mit

---

<sup>53</sup> Teilweise wurden noch weitere statistische Anpassungen und Transformationen vorgenommen, die hier nicht weiter dokumentiert sind, da sie spezifische Einzelfälle betreffen.

nicht kontinuierlichen Variablen beschäftigen, verwenden andere Effektstärkemaße. Nur fünf Studien haben keinerlei Transformation vorgenommen oder diese nicht dokumentiert.

Insgesamt beruhen fünf Analysen auf mittelwertbasierten Effektstärkemaßen. Von diesen mittelwertbasierten Meta-Studien haben nur Smith/Bartz/Scott Richards (2007) eine Transformation der Effektstärke vorgenommen. In diesem Fall wurde Cohens  $d$  in eine andere standardisierte Mittelwertdifferenz umgeformt, die jedoch nicht näher bezeichnet wird.

Der nächste Bereich, der hier vergleichend zwischen den Meta-Studien betrachtet wird, ist die Berücksichtigung der Abhängigkeit von Effektstärken innerhalb einer Studie (Hedges 2009: 44). Wird das Verhältnis von Primärstudien zu verwendeten Effektstärken in Tabelle 7 analysiert, wird deutlich, dass die meisten Meta-Studien von dieser Problematik betroffen sind. Die Meta-Analysen nutzen hierbei unterschiedliche Lösungsansätze. Lediglich zwei Studien haben angegeben, dass sie die Abhängigkeit ignoriert haben. Einige Studien, bei denen auch die Anzahl der verwendeten Effektstärken nicht berichtet ist, haben keine Aussage hierzu getroffen. Die drei verwandten Studien zu den Auswirkungen von R/S auf verschiedene Gesundheitsaspekte berücksichtigen die Abhängigkeit der Effektstärken durch Verwendung eines robusten Varianzschätzers. Eine weitere Strategie, die Mittelwertbildung, wenden hingegen eine ganze Reihe von Studien an. Drei der genannten Studien (DeNeve/Cooper 1998; Sedikides/Gebauer 2009; Smith/McCullough/Poll 2003) nutzen zudem den sogenannten „shifting unit of analysis“ Ansatz.

Der vierte Aspekt, der in dieser Übersicht von statistischen Verfahren berücksichtigt wird, ist die Gewichtung der Effektstärken. 15 der 34 Studien verwenden keine Gewichtung. Die am häufigsten verwendete Variante der Gewichtung stellt die Gewichtung mithilfe der jeweiligen Stichprobengröße dar. Immerhin 14 Studien haben somit die Effektstärken gewichtet. Vier Studien setzen unterschiedliche Ausformungen der inversen Varianzgewichtung ein. Ähnlich wird von Sawatzky/Ratner/Chiu (2005) eine Gewichtung mit der geschätzten Varianz durchgeführt. Lamers et al. (2012) nutzen als weitere Variante die inversen Standardfehler der Konfidenzintervalle zur Gewichtung. Nahezu alle Studien, ganz gleich welches statistische Modell (Fixed-Effects, Random-Effects, Mixed-Effects) sie zugrunde legen, haben die Heterogenität der Effektstärken mittels  $Q$ -Statistik überprüft. Lediglich acht Studien haben einen solchen Homogenitätstest nicht durchgeführt bzw. nicht ausgewiesen. Zusätzlich zur  $Q$ -Statistik, die auf einem Chi-Quadrat-Signifikanztest beruht, wird in fünf Studien die sogenannte stichprobenbedingte Varianzstatistik ( $I^2$ ) berechnet.

**Tabelle 9: Ausgewählte methodologisch-statistische Aspekte**

Bereich	Studie	ES	Trans	Abhängig.	Gewicht.	Q / I <sup>2</sup>	Publikations-Bias
1) R/S & SWB	Witter et al. 1985	r	-	Jackknifes Konfidenzintervall	S	Q	-
	Sawatzky et al. 2005	r	z	Auswahl	Varianz	Q	Fail-safe-N
	Yonker et al. 2012	r	z	-	S	Q	Fail-safe-N
	Eger/Maridal 2015	r	z	ignoriert	-	-	Egger-Test
2) R/S & Gesundheit	Bergin 1983	r	z	a	-	-	-
	McCullough et al. 2000	OR	ln	a	-	Q	Funnel-Plot; Rangkorrelationstest; fail-safe-N
	Hackney/Sanders 2003	r	z	Stat. getestet	Inverse Varianz	Q	-
	Smith et al. 2003	r	z	a & Shifting unit of analysis	-	Q	Mixed ANOVA zwischen publiziert & unpubliziert; fail-safe-N; Trim-and-Fill
	Chida et al. 2009	HR/RR	ln	-	-	Q	Egger-Test; fail-safe-N
	Shor/Roelfs 2013	HR	-	-	S	Q / I <sup>2</sup>	Funnel Plot + Peter-Test; fail-safe-N
	Jim et al. 2015	r	z	robuste Varianzschätz.	-	-	Funnel Plot; cluster-robuste Variante Egger-Test; Sensitivitätsanalyse mit Missing Data (0 eingesetzt)
	Salsman et al. 2015						
	Sherman et al. 2015						
	Wu et al. 2015	OR	ln	-	-	Q / I <sup>2</sup>	Rangkorrelationstest, Egger-Test
3) R/S & X	Donahue 1985	r	z	-	S	-	-
	Baier/Wright 2001	r	-	a	S	-	-
	Mahoney et al. 2001	r	z	-	-	Q	Fail-safe-N
	Saroglou 2002	r	z	-	S	Q	-
	Saroglou et al. 2004	r	z	-	S	Q	-
	Ano/Vasconellas 2005	r	z	ignoriert	-	Q	Fail-safe-N
	Smith et al. 2007	d	SMD	a	S	Q	Funnel-Plot; fail-safe-N; Trim-and-Fill
	Zuckerman et al. 2013	r	-	a	S	Q	Moderatortest; fail-safe-N
4) X & SWB	Okun et al. 1990	m	-	-	-	Q	-
	DeNeve/Cooper 1998	r		a & Shifting unit of analysis	S	Q	-
	Pinquart/Sörensen 2000	r	z	Samplegröße pro Effekt angepasst	S	Q	-
	Fischer/Boer 2011	r	-	-	S	Q	-
	Weis/Speridakos 2011	d	-	a	Inverse Varianz quadriert	Q	Moderatortest
	Capaldi et al. 2014	r	z	-	S	Q / I <sup>2</sup>	Moderatortest

**Anmerkungen:** Effektstärken: r = Pearsonschen Korrelationskoeffizient r; OR = Odds Ratio; RR = Risikodifferenz; m=mittelwertbasiert (nicht genauer beschrieben); HR=Hazard-Ratio. Transformation: z=Fishers z-Transformation; LN=natürlicher Logarithmus; Log=Logarithmus basiert; SMD=stand. Mittelwertdifferenz. Abhängigkeit: a=bei mehreren Messungen/Effektstärken pro Primärstudie den Mittelwert bilden. Gewichtung: s=Gewichtung nach Stichprobengröße.

In Bezug auf den Test des sogenannten Publikationsbias (bzw. des „file-drawer“-Problems) gibt es eine Vielzahl von Untersuchungsmöglichkeiten (siehe 5.5.1). Immerhin 14 der betrachteten Studien haben keine Überprüfung vorgenommen. 20 Studien wenden mindestens eines der zur

Verfügung stehenden Verfahren an. Von diesen nutzen jedoch acht nur eine Methode zur Überprüfung. Immerhin zwölf Studien setzen zwei Verfahren zur Überprüfung ein. Drei Studien nutzen eine Moderatoranalyse.<sup>54</sup> Die graphische Analyse mittels des sogenannten Funnel-Plots nutzen acht Meta-Studien. Die Diagnose mithilfe einer Regressionsanalyse, der sogenannte Egger-Test sowie Variationen hiervon, verwenden sieben Studien. Weiterhin führen zwei Studien einen Test mittels der „rank-correlation“ Methode durch. Rosenthals fail-safe-N berechnen zwölf der untersuchten Meta-Studien. Deutlich seltener kommt die „trim-and-fill“-Methode zum Einsatz. Sie wird lediglich von drei Studien angewendet. Eine einzige Studie führt eine Varianzanalyse zwischen publizierten und nicht publizierten Studien durch. Die empirische Praxis bei der Publikationsbias-Diagnose ist sehr gut, da viele der betrachteten Studien mehrere Diagnoseverfahren nutzen.

Als letzten Bereich werden die Moderatorvariablen, die in den berücksichtigten Meta-Studien analysiert wurden, betrachtet. Dazu sind die am häufigsten in dem betrachteten Forschungsfeld eingesetzten Moderatorvariablen in Tabelle 10 zusammengefasst. Dabei geht es jedoch nicht um mögliche Erklärungskraft und Signifikanz der Moderatoren, sondern vielmehr darum, welche Moderatoren berücksichtigt werden. Die Moderatorvariablen wurden dazu in zwei Bereiche eingeteilt. Der erste Bereich beinhaltet Moderatoren, die auf inhaltlicher Ebene einen Einfluss auf die Beziehung der Variablen nehmen. Der zweite Bereich hingegen besteht aus Moderatoren, die als Studienmerkmale bezeichnet werden können. Lediglich sieben der betrachteten 34 Studien nehmen keine Moderatoranalyse vor.

Sechs Studien verwenden die schwer operationalisierbare Studienqualität als Moderator (Card 2012: 68ff.; Valentine 2009: 129ff.). Neun Studien nutzen die Stichprobengröße als Moderatorvariable. Ebenso viele Studien berücksichtigen das Publikationsjahr der Primärstudien als potenziellen Moderator. Der Publikationsstatus, also die Erscheinungsform einer Studie, wird von acht Studien als Moderator verwendet.

Die Art der R/S-Messung bzw. der jeweilig interessierenden Variable wird 18-mal als Moderator berücksichtigt. Hierbei ist die Art und Weise, wie diese Moderatorvariable berücksichtigt wird, besonders interessant. Witter et al. (1985) nehmen eine Einteilung in religiöse Aktivität und Religiosität vor. Sawatzky/Ratner/Chiu (2005) unterscheiden differenzierter, da sie vier Einteilungen bezogen auf die Art der Spiritualität vornehmen. Yonker/Schnabelrauch/Dehaan (2012) unterscheiden fünf Arten von Religiosität. Bei ihnen wird unterschieden: Besuch des

---

<sup>54</sup> Weitere Studien haben zwar eine entsprechende Variable („Publikationsstatus“) erhoben, diese allerdings nicht zur Diagnose des Problems verwendet: DeNeve/Cooper (1998); Goyal et al. (2014); Okun/Olding/Cohn (1990); Prati/Pietrantonio (2009).

Gottesdiensts, religiöses Verhalten, Salienz der Glaubensüberzeugungen, religiöse Suche und eine Mischkategorie, die aus verschiedenen Aspekten besteht.

**Tabelle 10: Moderatoruntersuchung**

Bereich	Studie	Inhaltliche Moderatoren						Erhebungsspez. Moderatoren								
		Alter	Geschlecht	Rasse/Ethnie	Herkunft	Bildung	Konfession	Makromerkmale	Studienqualität	Samplegröße	Publikationsjahr	Publikationsstatus	R/S-Messung	SWB-Messung	Forschungsdesign	Stichprobenart
1) R/S & SWB	Witter et al. 1985	x	x						x		x		x	x		
	Sawatzky et al. 2005	x	x	x			x						x	x		x
	Yonker et al. 2012	x	x	x	x					x			x		x	
	Eger/Maridal 2015															
2) R/S & Gesundheit	Bergin 1983															
	McCullough et al. 2000	x	x										x		x	x
	Hackney/Sanders 2003												x	x		
	Smith et al. 2003	x	x	x						x		x	x	x	x	x
	Chida et al. 2009								x	x			x	x	x	x
	Shor/Roelfs 2013	x	x		x				x	x	x					
	Jim et al. 2015	x	x	x	x										x	
	Salsman et al. 2015	x	x	x	x										x	
	Sherman et al. 2015	x	x	x	x										x	
	Wu et al. 2015	x			x								x		x	x
3) R/S & X	Donahue 1985															
	Baier/Wright 2001		x		x	x	x			x						x
	Mahoney et al. 2001													x	x	x
	Saroglou 2002												x			
	Saroglou et al. 2004							x								
	Ano/Vasconellas 2005															
	Smith et al. 2007	x	x											x	x	
	Zuckerman et al. 2013		x										x	x	x	
4) X & SWB	Okun et al. 1990	x								x		x		x	x	x
	DeNeve/Cooper 1998	x	x	x					x	x	x	x	x	x	x	x
	Pinquart/Sörensen 2000	x	x						x		x		x	x		x
	Fischer/Boer 2011															
	Weis/Speridakos 2011	x										x	x	x	x	x
	Capaldi et al. 2014	x	x								x	x	x	x		

**Anmerkungen:** Spezifische auf die jeweilige Forschungsfrage oder auf die Methodik (Eigenschaften eines Experiments etc.) bezogene Moderatoren werden in dieser Übersicht nicht dargestellt. Die Moderatoren wurden für diese Übersicht zu Kategorien zusammengefasst.

Ähnliche Einteilungen lassen sich auch bei den anderen berücksichtigten Studien finden. McCullough et al. (2000) unterteilen den Religiositätsmoderator ebenfalls in vier Bereiche: öffentlich, privat, eine Kombination aus beidem und unklar. Hackney/Sanders (2003) unterscheiden die Dimensionen: institutionelle Religiosität, ideologische Religiosität sowie persönliche Hingabe. Deutlich differenzierter sind etwa Smith/McCullough/Poll (2003) und Chida/Step-toe/Powell (2009) vorgegangen. Sie unterscheiden sieben bzw. sechs unterschiedliche Dimen-

sionen. Smith/McCullough/Poll (2003) unterscheiden religiöses Verhalten, religiöse Einstellungen, religiöse Orientierung, religiöses Coping, religiöses Wohlbefinden, Gott-Konzeption und multidimensionale Religiositätsmessung.

Auch die Art des Messinstrument subjektiven Wohlbefindens wird als Moderatorvariable berücksichtigt. Witter et al. (1985) unterscheiden nach Glück, Lebenszufriedenheit, Moral und globalen Messungen von Lebensqualität bzw. Wohlbefinden. Eine einfache Einteilung in evaluatives und affektives Wohlbefinden nutzen Eger/Maridal (2015). Yonker/Schnabelrauch/Dehaan (2012) nehmen sowohl positive als auch negative Konzepte in die Betrachtung auf. Sie unterscheiden dabei Wohlbefinden, Depression, Angst, Selbstvertrauen und Wohlbefinden/Glück. Sawatzky/Ratner/Chiu (2005) differenzieren zwischen sechs Dimensionen: multiplen Dimensionen, krankheitsspezifischen Messungen, globalen Ein-Item-Messungen, Wohlbefinden, Lebenszufriedenheiten und Anderen.

Unter der Kategorie „Forschungsdesign“ wird eine Vielzahl von Moderatoren zusammengefasst, da diese oftmals sehr speziell auf die jeweilige Untersuchungsform angepasst sind. 15 Studien verwenden einen solchen Moderator für die Analyse. Ähnlich verhält es sich mit der Art der Stichprobe. Dabei geht es jedoch nicht um demographische Aspekte (Geschlecht, Alter, Rasse, etc.), sondern vielmehr darum, ob es sich um etwa eine studentische, klinische oder repräsentative Stichprobe handelt.

In Bezug auf die Moderatoren bleibt festzuhalten, dass viele Studien Moderatoranalysen für die statistische Erklärung der Variabilität der Effektstärken durchführen. Allerdings stellen die meisten Autoren fest, dass auch nach der Moderatorenanalyse Heterogenität besteht und somit statistische Varianz unerklärt bleibt. Auffällig ist, dass die Wahl der Moderatoren kaum hypothesengestützt erfolgt, sondern vielmehr auf pragmatische Art möglichst viele Studieneigenschaften genutzt werden.

#### **4.1.6 Zusammenfassung der methodologischen Praxis**

Die Dokumentation des meta-analytischen Vorgehens, besonders im Hinblick auf die Replizierbarkeit der Meta-Analysen, ist bei einigen Studien eher unvollständig. Dies zeigt sich etwa an der spärlichen Dokumentation der Literatursuchstrategien oder in den unzureichend klar definierten Inklusions- und Exklusionsbedingungen sowie an der begrenzten Übersicht über die verwendeten statistischen Verfahren. Die verwendeten statistischen Methoden und Verfahren entsprechen beim Großteil der Studien den herkömmlichen Analysemethoden. Auffällig ist ebenfalls, dass viele der betrachteten Meta-Studien das Problem der abhängigen Effektstärken

nicht ausreichend berücksichtigen. Dementsprechend sind die Ergebnisse dieser Studien möglicherweise verzerrt und die Ergebnisse unbrauchbar. Keine der gefundenen Meta-Analysen nutzt ein modellbasiertes meta-analytisches Verfahren.

Zusammenfassend betrachtet lässt sich die meta-analytische Forschungspraxis aus zweierlei Gründen kritisieren. Zum einen ist das zugrundeliegende statistische Analysemodell im Hinblick auf die komplexen theoretischen Entwürfe zwischen den Konzepten unzureichend, weil es keine Erklärungsmechanismen berücksichtigen kann. Zum anderen sind die Dokumentation des meta-analytischen Vorgehens sowie die Berücksichtigung der methodologischen Empfehlungen in vielen Studien spärlich, so dass manche Studien nicht nachvollzogen bzw. repliziert werden können. Die Erkenntnisse aus der deskriptiven Betrachtung der methodologischen Praxis bildet nun die Grundlage für die Festlegung der eigenen Studierhebung im nächsten Schritt.

## **4.2 Selektionskriterien: Inklusion und Exklusion**

Vor der Festlegung der eigentlichen Suchstrategie geht es zunächst darum, zu entscheiden, welche Grundgesamtheit der geplanten Meta-Studie zugrunde liegt. Dazu folgen nun die zentralen Inklusions- bzw. Exklusionskriterien, die den Rahmen der Studie bestimmen (siehe Tabelle 11). Prinzipiell werden nur Studien in die Meta-Analyse eingeschlossen, die eine für die Untersuchung relevante Beziehung empirisch-korrelativ innerhalb eines Landes untersuchen. Die Grundgesamtheit der Primärstudien muss jedoch im Hinblick auf das Erscheinungsdatum der Studien eingeschränkt werden, da eine große Anzahl von Studien zum Untersuchungsgegenstand vorliegen. Die Einschränkung lässt sich inhaltlich dadurch begründen, da sich das Verständnis von Religiosität und insbesondere von Spiritualität über die Zeit hinweg wandelt. Um eine Vergleichbarkeit der Ergebnisse gewährleisten zu können, sollte deshalb der Entstehungszeitraum der Studien ähnlich sein. Aus den angeführten Gründen werden für die vorliegende Meta-Analyse lediglich diejenigen Studien berücksichtigt, die ab dem Jahr 2000 erschienen sind.<sup>55</sup> Ebenfalls aus forschungspragmatischen Gründen und wie in der empirischen Praxis üblich, werden nur Studien eingeschlossen, die in englischer oder deutscher Sprache verfasst sind.

---

<sup>55</sup> Dies bezieht sich jedoch nur auf das Erscheinungsdatum der Publikation und nicht auf den Erhebungszeitraum der Daten. Dieser kann auch vor dem Jahr 2000 liegen. Dies hat den erfreulichen Vorteil, dass in der Regel neuere Studien in Hinblick auf die Dokumentation der verwendeten Untersuchungsmethoden erfahrungsgemäß präziser sind als ältere Studien.

Wie bereits deutlich geworden, ist ein überschneidungsfreies (disjunktes) Forschungsdesign von zentraler Bedeutung, um valide Ergebnisse in diesem Forschungsfeld zu erzielen (Koenig/King/Carson 2012: 143). Deshalb bezieht sich das erste Inklusionskriterium auf das Vorliegen einer für die Meta-Analyse relevanten Beziehung in quantitativ-empirischer Form mittels entsprechender Messinstrumente. Als relevante Beziehungen gelten zum einen die direkten Beziehungen zwischen Religiosität/Spiritualität (UV) und Lebenszufriedenheit (AV) und zum anderen zwischen Mediatorvariablen und UV oder AV. Hierbei muss sichergestellt werden, dass alle im Sinne der breiten konzeptionellen Definition passenden Konzepte berücksichtigt werden, aber dennoch tautologische, weil sich überschneidende Konzepte, aus der Analyse ausgeschlossen bleiben.

Beim ersten Konzept, der Religiosität, sind das unter anderem das sogenannte religiöse/spirituelle oder existenzielle Wohlbefinden, das in der Literatur mitunter als Teilaspekt von Wohlbefinden diskutiert wird, sowie bestimmte positive psychologische Zustände („meaning and purpose“), die nicht explizit auf das Transzendente bezogen sind und zum Teil die eigentlichen Mediatorvariablen darstellen.

Für das habituelle Wohlbefinden dürfen nur Messinstrumente im Form von Lebenszufriedenheit verwendet werden. Konzepte aus dem Bereich des psychologischen Wohlbefindens werden aufgrund von „tautologischen Verunreinigungen“ (etwa Sinn des Lebens) sowie Messungen von subjektivem Wohlbefinden, die sich lediglich auf die körperliche oder geistige Gesundheit oder andere objektive Messinstrumente stützen, ausgeschlossen. Genauso können diejenigen Studien nicht berücksichtigt werden, die verunreinigte Messinstrumente zur Erfassung von Lebensqualität nutzen.

Das nächste Inklusionskriterium bezieht sich darauf, dass die untersuchte Beziehung im jeweiligen Forschungsbericht auch dokumentiert wurde. Da wir es im vorliegenden Fall mit Zusammenhängen zwischen metrischen Variablen (bzw. quasi-metrischen) und somit korrelativen Daten zu tun haben, wird der Pearsonsche Produkt-Moment-Korrelationskoeffizient ( $r$ ) gesucht – partielle Korrelationen werden hingegen nicht berücksichtigt. Es wird jedoch keine herkömmliche, sondern eine modellbasierte Meta-Analyse durchgeführt und dafür ist es notwendig, eine Korrelationsmatrix aller notwendigen bivariaten Korrelationen zu erheben. Das heißt also, dass die bivariaten Korrelationen zwischen allen Variablen, die für die jeweilige Studie im Untersuchungsmodell berücksichtigt werden sollen, in der jeweiligen Studie auch berichtet werden müssen. Fehlende Werte (nicht berichtete Beziehungen des Untersuchungsmodells) stellen für das TSSEM-Verfahren kein Problem dar (siehe ausführlich in Kapitel 5).

**Tabelle 11: Selektionskriterien**

Inklusionskriterien	Exklusionskriterien
Korrelative Untersuchung einer relevanten Beziehung	Qualitative und theoretische Arbeiten sowie Meta-Analysen und Literaturreviews
Publikationsdatum der Literaturstelle $\geq$ 2000	Messung von Wohlbefinden nur über Aspekte von Gesundheit oder andere objektive Kriterien bzw. über Affekt
Sprache: Deutsch oder Englisch	Nur Verwendung tautologischer Konzepte zur Untersuchung von R/S, SWB oder Mediatoren
Mindestens eine der untersuchten relevanten Beziehungen muss mit disjunkten Messinstrumenten gemessen werden	Multinationale Studien, die keine separaten Korrelationen nach Ländern ausweisen

Ausgeschlossen werden entsprechend alle Arbeiten, die den formulierten Einschlusskriterien nicht entsprechen. Explizit sind das alle theoretischen und qualitativ-empirischen sowie meta-analytischen oder Review-Arbeiten. Darüber hinaus werden alle Studien ausgeschlossen, die zwar die interessierende Beziehung untersucht, aber zur Messung von habituellem Wohlbefinden lediglich Gesundheitsindikatoren oder andere objektive Kriterien genutzt haben. Darüber hinaus müssen auch multinationale Studien, die für die eingeschlossenen Länder keine separaten Korrelationen ausgewiesen haben, aus der Analyse ausgeschlossen werden. Dies ist unvermeidbar, da ausgewählte Moderatoranalysen mittels Makroindikatoren auf Länderebene durchgeführt werden.

### 4.3 Literatursuchstrategie

An die Definition der Selektionskriterien anknüpfend, folgt nun die Festlegung der Literatursuchstrategie. Diese umfasst zwei Bausteine, die sich an den Empfehlungen der meta-analytischen Literatur orientieren (Card 2012; Reed/Baxter 2009; Rothstein/Hopewell 2009; Urban/Fiebig 2015; White 2009). Eine Übersicht über die entwickelte Suchstrategie ist in Tabelle 12 zu sehen.

**Tabelle 12: Suchstrategie**

	Baustein
1	Bekannte Literaturreviews und Meta-Analysen
2	Suche in Zitations- und Fachdatenbanken

Als ersten Schritt werden die bekannten Meta-Analysen, die Literaturliste des „Handbook of Religion and Health“ (Koenig/King/Carson 2012) sowie weitere bekannte Literaturreviews einbezogen, die nach dem Jahr 2000 erschienen sind. In Tabelle 13 ist eine Übersicht der genannten Meta-Studien und Reviews zu sehen, die bereits zum Beginn der Suche bekannt sind. Da sich in diesem Forschungsbereich die Konzepte oftmals ähneln, werden in dieser Übersicht sowohl Meta-Studien, die die Beziehung zwischen unabhängiger und abhängiger Variable direkt im Fokus haben, als auch Studien die den Zusammenhang von R/S und psychischer Gesundheit untersuchen, berücksichtigt. Beim bisherigen Literaturstudium ist deutlich geworden,

dass Wohlbefinden, Lebenszufriedenheit und psychischer Gesundheit teilweise synonym verwendet wird, so dass ein Screening dieser Literatur hilfreich sein sollte.

**Tabelle 13: Bekannte Literaturreviews und Meta-Studien**

Bereich	Studien
Literaturreviews	Visser/Garssen/Vingerhoets 2010
	Koenig/King/Carson 2012
	Schreiber/Brockopp 2012
	Rizvi et al. 2016
Meta-Analysen	Hackney/Sanders 2003
	Smith et al. 2003
	Sawatzky et al. 2005
	Yonker et al. 2012
	Eger/Maridal 2015
	Salsman et al. 2015

Nach der Durchsicht bereits bekannter Meta-Studien und Reviews folgt im zweiten Schritt die Suche mithilfe fünf unterschiedlicher elektronischer Datenbanken (Reed/Baxter 2009). Da es sich bei dem Untersuchungsthema um ein Querschnittsthema aus den Forschungsdisziplinen Medizin, Psychologie und Sozialwissenschaft handelt, werden entsprechend der Disziplinen drei unterschiedliche Datenbanken genutzt (siehe Tabelle 14). Darüber hinaus wird eine spezielle Datenbank zur Suche nach grauer Literatur verwendet, da in den großen Literaturdatenbanken fast ausschließlich begutachtete Journals erfasst werden und deren Suchergebnisse dementsprechend anfällig für den Publikationsbias sind (Rothstein/Hopewell 2009). Die Auswahl der Datenbanken ist inhaltlich über die Interdisziplinarität der Fragestellung und der damit einhergehenden möglichst großen Abdeckung der Forschungslandschaft in unterschiedlichen Bereichen begründet, aber auch durch Zugänglichkeit zu den entsprechenden Ressourcen eingeschränkt, da der Zugang zu diesen Datenbanken in der Regel kostenintensiv ist.

Bei der ersten verwendeten Datenbank handelt es sich um die freizugängliche medizinische Fachdatenbank PubMed.<sup>56</sup> Für den psychologischen Bereich steht die PSYNDEX Datenbank zur Verfügung.<sup>57</sup> Für den sozialwissenschaftlichen Bereich wird zum einen SOWIPORT und zum anderen der Social Science Citation Index (SSCI) herangezogen. Zu den Datenbanken, die auch Hochschulschriften sowie andere Forschungsberichte indexieren, zählen SOWIPORT und BASE. Bei SOWIPORT handelt es sich nicht nur um eine sozialwissenschaftliche Zitationsda-

<sup>56</sup> Zugänglich unter: <https://www.ncbi.nlm.nih.gov/pubmed/advanced>

<sup>57</sup> Für Angehörige der Universität Stuttgart als Nutzer der Württembergischen Landesbibliothek über den Anbieter EBSCOhost zugänglich.

tenbank, sondern auch um ein Fachportal, dass eine Vielzahl von weiteren sozialwissenschaftlichen Datenbanken<sup>58</sup> lizenziert hat (Urban/Fiebig 2015: 190). Die wichtigste Zitationsdatenbank für diesen Suchabschnitt stellt der Social Science Citation Index als Teil der von Thomson Reuters bereitgestellten Rechercheoberfläche „Web of Science“ dar.<sup>59</sup> Bei der Bielefeld Academic Search Engine (BASE)<sup>60</sup> handelt es sich um eine der weltweit größten wissenschaftlichen Suchmaschinen für Web-Dokumente mit einem hohen Anteil an grauer Literatur.

**Tabelle 14: Literaturdatenbanken nach Disziplin**

Forschungsdisziplin	Datenbankname
Medizin	PubMed
Psychologie	PSYINDEX
Sozialwissenschaft	SOWIPORT
	Web of Science: Social Science Citation Index
Graue Literatur	BASE (Bielefeld Academic Search Engine)

Die Verwendung von SOWIPORT und BASE sind zur Vermeidung von Ergebnisverzerrungen aufgrund des „file-drawer-Problems“ von besonderem Interesse (Rothstein/Hopewell 2009).

Im Anschluss an die Übersicht der zur Literatursuche eingeplanten Zitationsdatenbanken folgt nun die Festlegung der Suchstrings. Diese Suchstrings können auch als operationale Definition der zugrunde liegenden Konzepte gedeutet werden (Cooper 2009: 20ff.), da mit deren Hilfe die abstrakten theoretischen Konzepte wie Religiosität operational präzisiert werden.

Betrachtet man die in Kapitel 4.1.1 gefundenen Suchstrings, wird sogleich deutlich, dass die Suchbegriffe den zugrunde gelegten und in der Regel breiten Konzeptualisierungen entsprechen. Diese Suchstrings müssen in einem nachfolgenden Schritt auf Plausibilität hinsichtlich der Konzepte überprüft werden. Für die abhängige Variable wird nach allen Begriffen, die mit subjektiven Wohlbefinden bzw. Lebenszufriedenheit in Verbindung stehen gesucht, da diese Konzepte oftmals synonym zueinander verwendet werden. Dies hat sich auch im Review der methodologischen Praxis in Kapitel 4.1. gezeigt. In der nachfolgenden Tabelle 15 ist eine Übersicht über die verwendeten Suchstrings zu sehen. Diese Suchbegriffe werden lediglich in englischer Sprache verwendet, da alle Datenbanken mit englischen Keywords arbeiten und mit den Möglichkeiten von Booleschen-Operanden sowie unter Verwendung der Rechts-Trunkierung verknüpft.<sup>61</sup>

<sup>58</sup> Unter: <http://sowiport.gesis.org/Database> findet sich die jeweils aktuelle Fassung der per Nationallizenz für SOWIPORT lizenzierten Datenbanken. Das Portal wurde allerdings mit Ablauf des Jahres 2017 geschlossen.

<sup>59</sup> Der SSCI (ebenso wie der Science Citation Index Expanded) ist zur Erhebungszeit als Angehöriger der Universität Stuttgart über das Datenbank-Infosystem zugänglich.

<sup>60</sup> Frei im Internet zugänglich unter: <https://www.base-search.net/Search/Advanced>.

<sup>61</sup> Die Literaturdatenbanken erlauben es in der Regel die verschiedenen Suchbegriffe mittels booleschen Operanden (AND, OR, NOT etc.) miteinander in Beziehung zu setzen. Eine weitere nützliche Funktion dieser Datenbanken besteht in der sogenannten Trunkierung („wildcard“) mittels eines Sternchen oder ähnlicher Zeichen (\$) (Urban/Fiebig 2015: 83). Diese Funktion erlaubt es etwa die Begriffe „religion“, „religiosity“ und „religiousness“

**Tabelle 15: Suchstrings**

	<b>Religiosität / Spiritualität</b>	<b>Subjektives Wohlbefinden / Lebenszufriedenheit</b>
Vollständig	religion; religiosity; religiousness; religious; religious belief; religious coping; spiritual; spirituality; faith	happiness; subjective well-being; emotional well-being; well-being wellbeing; life satisfaction; satisfaction with life; affect; positive affect; negative affect; life quality; quality of life
Trunkiert	religio* OR spirtual* OR faith	happiness OR well*being OR life satisfaction OR affect OR life quality

Für jede Fachdatenbank muss ein angepasster Suchstring formuliert werden, der die entsprechenden Eigenheiten der Datenbanken berücksichtigt. Da aus den bisherigen Erkenntnissen bereits deutlich geworden ist, dass durch die beschriebene Suche eine Vielzahl von Studien gefunden werden, wird auf die Verwendung weiterer Suchverfahren verzichtet.

#### 4.4 Screeningverfahren

Im nachfolgenden Abschnitt wird in Anschluss an die Definition der Selektionskriterien sowie die Festlegung der Literatursuchstrategie nun das Screeningverfahren und die dazugehörigen Regeln beschrieben. Dabei wird das Verfahren festgelegt, wie die infolge der Literatursuche gefundene Studienpopulation auf ihre Geeignetheit für die Untersuchung hin überprüft wird. Die Festlegung eines solchen Verfahrens ist notwendig, um nachvollziehbar sicherzustellen, dass möglichst alle relevanten Primärstudien, die mit der geschilderten Suchstrategie gefunden werden und den festgelegten Selektionskriterien entsprechen, bei der Studienkodierung berücksichtigt werden.

Das Screening wird aufgrund der als sehr hoch eingeschätzten Anzahl von potenziell relevanten Studien (> 10.000) in zwei Teile aufgeteilt. Beim „Vor-Screening“ wird zunächst anhand der Titel und Abstracts versucht herauszufinden, ob eine bestimmte Primärstudie auf den ersten Blick für die eigene Meta-Analyse geeignet ist. Im Fokus steht dabei die Frage, ob die entsprechende Untersuchung überhaupt eine fürs Untersuchungsmodell relevante Beziehung untersucht. Falls die Informationen des Abstracts zum Screening nicht ausreichen, wird anschließend der Methoden- und Untersuchungsteil oder falls notwendig, die gesamte Studie gelesen, bevor über den Einschluss endgültig entschieden wird. Wenn sich bei der ersten Lektüre bereits zeigt, dass eine Studie unpassend ist, wird die Studie nicht näher betrachtet und aus der weiteren Untersuchung ausgeschlossen. Das Ziel dieses Vor-Screening besteht darin die Anzahl der potenziell relevanten Studien für die detaillierte Begutachtung zu reduzieren. Beim zweiten Teil

---

durch das verkürzte „relig\*“ zu ersetzen. Damit lassen sich auch Auslassungen in der Mitte eines Wortes modellieren, so dass beispielsweise well\*being sowohl für well-being als auch für wellbeing steht.

des Screenings werden die Studien bezüglich der Selektionskriterien präzise überprüft, um die Studienpopulation für die Kodierung entsprechend einzugrenzen.

Die folgenden Regeln stellen sicher, dass alle Studien während des Screenings gleichbehandelt werden, um die Reliabilität der Zuordnung sicherstellen zu können.

1. Das informelle Vor-Screening wird abgebrochen, wenn ersichtlich ist, dass die Studie nicht relevant ist. Eine Studie ist dann nicht relevant, wenn sie keine für das Erklärungsmodell relevante Beziehung untersucht.
2. Nicht relevant im Haupt-Screening ist eine Studie nur dann, wenn eines der Inklusionskriterien nicht zutrifft und/oder eines der Exklusionskriterien zutrifft.
3. Der Grund des Ausschlusses muss dokumentiert werden. Wenn eine Studie aufgrund mehrerer Selektionskriterien für die Untersuchung nicht passend ist, wird sie jedoch nur für das erste zutreffende Exklusionskriterium gekennzeichnet.

Diese Regeln sind auf alle Studien während des Screenings gleichermaßen anzuwenden und dienen der Sicherstellung der Datenqualität auch im Hinblick auf die Reliabilität. Im nachfolgenden Kapitel 4.5 erfolgt nun die Ableitung eines Verfahrens zur Kodierung der Studien und die Entwicklung eines Kodierschema, mit dem die interessierenden Merkmale (Studiencharakteristika und Effektstärken) aus den relevanten Studien extrahiert werden.

## **4.5 Kodierverfahren**

Die relevanten Kodierungsvariablen ergeben sich aus den analytischen Vorüberlegungen und der methodischen Praxis der Meta-Studien (siehe Kapitel 4.1.4) in diesem Themenfeld sowie den Hinweisen aus den methodischen Standardwerken (Cooper 2010; Cooper/Hedges/Valentine 2009). Bei der Kodierung werden, die für die Meta-Analyse relevanten Informationen aus den Primärstudien extrahiert. Besonders relevant sind hierbei die Effektstärken in Form einer Korrelationsmatrix und ausgewählte Studiencharakteristiken. Im Anschluss an die Kodierung wird eine einfache Form von Reliabilitätskontrolle vorgenommen.

Eine vollständige doppelte Kodierung kann aufgrund der Vielzahl der Studien und der personellen Ressourcenbeschränkung nicht durchgeführt werden. Allerdings wird eine Reliabilitätskontrolle auf zweierlei Art durchgeführt. Zum einen werden alle relevanten Studien nach ihrer Kodierung im Hinblick auf die zentralen Aspekte (Fallzahl und Effektstärken) abermals durchgesehen, um etwaige Fehler festzustellen. Dieses Vorgehen stellt die Intra-Rater-Reliabilität sicher. Um auch die Inter-Rater-Reliabilität sicherzustellen, werden 20 % der Stichproben voll-

ständig von einer Kollegin erneut kodiert und etwaige Diskrepanzen per Diskussion der strittigen Punkte gelöst. Die Koderin wurde zuvor umfassend im Hinblick auf alle relevanten Aspekte der Kodierung geschult.

Nach diesen allgemeinen Vorgaben über das Vorgehen während der Kodierungsphase folgt nun die Beschreibung des Kodierschema. Das entwickelte Schema besteht aus insgesamt sechs Bereichen (siehe Tabelle 16). Das Kodierschema ist dabei mittels Filterführung gestaltet, so dass Bereiche, die für eine bestimmte Studie nicht relevant sind, übersprungen werden können.

**Tabelle 16: Bereiche des Kodierschema**

Bereich	Inhalt
1	Meta-Daten
2	Angaben zur Studie
3	Study-Level-Moderatoren
4	Mediatoren
5	Korrelationsmatrix
6	Kodierüberprüfung

Der erste Bereich beinhaltet eine Anzahl von sogenannten Meta-Daten. Dazu gehören die Publikationsnummer und Stichprobennummer. Diese beiden werden unterschieden, da eine Publikation (etwa ein Journalartikel) aus mehreren unabhängigen Stichproben bestehen kann. Über die Publikationsnummer sind dann alle Stichproben eindeutig einer Publikation zuordenbar. Ebenfalls werden dort ein Koderkürzel, Kodier-Datum und die Versionsnummer des Kodierschemas eingetragen. Im zweiten Bereich des Kodierschemas werden die wichtigsten Studienmerkmale, wie Stichprobengröße, Publikationsjahr der Studie, das Land, in dem die Studie durchgeführt wurde, Publikationsstatus, Journal, Datenquelle, welche Korrelationen und welche Messinstrumente verwendet werden, erfasst. Das Schema ist dabei so angelegt, dass sowohl für R/S, als auch für Lebenszufriedenheit mehrere unterschiedliche Variablen erfasst werden können.

Der dritte Bereich beschäftigt sich mit potenziellen study-level Moderatoren. Dort wird das durchschnittliche Alter der Stichprobe, deren geschlechtliche Zusammensetzung, die Konfessionen der Stichprobe sowie Eigenschaften der Stichprobe (Repräsentativität bzw. Art der Spezialpopulation) und die Art des Untersuchungsmodells erfasst. Im vierten Bereich folgt die detaillierte Erfassung der Mediatorvariablen. Auch hier besteht die Möglichkeit, dass für jeden der sechs Arten von Mediatoren bis zu drei unterschiedliche Operationalisierungen erfasst werden können. Der fünfte Bereich stellt den wichtigsten Teil des Kodierschemas dar, da die Korrelationen bzw. die Korrelationsmatrix erfasst wird. Der sechste Bereich beinhaltet eine subjektive Einschätzung der methodischen Güte der Primärstudie und der Schwierigkeit der Kodierung. Die Endversion des Kodierschema ist im Anhang A zu finden.



## **5 Meta-analytische (statistische) Methodik**

Nachdem im vorangegangenen Kapitel das Forschungsdesign der Studierenerhebung und -kodierung festgelegt wurde, folgt nun die Einführung in die statistische Methodik der modellbasierten Meta-Analyse. Zunächst werden die grundlegenden statistischen Verfahren der Meta-Analyse erklärt.

### **5.1 Grundlegende statistische Konzepte und Verfahren**

Im vorliegenden Abschnitt geht es darum, die klassischen meta-analytischen Verfahren kurz zu erläutern und zusammenhängend die zentralen meta-analytischen Konzepte, die auch bei den nachfolgenden Erläuterungen zu den modellbasierten Ansätzen relevant sind, einzuführen. Anhand der grundlegenden Konzepte werden dann die Vorteile der modellbasierten Verfahren deutlich.

Die Gemeinsamkeit aller meta-analytischen Verfahren des APD-Typs liegt darin, dass keine Rohdaten, sondern Effektstärken die Analyseeinheit darstellen. Unter Effektstärke wird ein standardisierter Index für die Richtung und Stärke eines Zusammenhangs zwischen zwei Variablen verstanden (Card 2012: 87). Bei den klassischen Verfahren geht es darum, die Effektstärken der verschiedenen Primärstudien in eine einheitliche Metrik zu überführen und diese zusammenzufassen. Anschließend wird eine durchschnittliche (kumulierte) Gesamteffektstärke und das Ausmaß der Variabilität der Effektstärke berechnet sowie die Variabilität zwischen den Studien mithilfe von Moderatoranalysen auf bestimmte Studienmerkmale zurückgeführt.

Der Reihe nach werden nun zentrale Aspekte der meta-analytischen statistischen Auswertung diskutiert. Zunächst werden die grundlegenden Modellierungsformen, die in einer Meta-Analyse möglich sind, vorgestellt. Anschließend werden folgende Aspekte diskutiert: Heterogenität und Heterogenitätstests, die möglichen Analyseeinheiten (Effektstärken), die Gewichtung der Effektstärken, die Möglichkeit bestimmte Studienartefakte zu korrigieren und der Umgang mit statistisch abhängigen Effektstärken.

#### **5.1.1 Grundlegende Modellierungsformen**

Zur praktischen Umsetzung der vorgestellten meta-analytischen Modellierungsformen wurden in der Vergangenheit verschiedene Umsetzungsvarianten vorgeschlagen. Die Ansätze (Hedges/Olkin 1985; Lipsey/Wilson 2010; Schmidt/Hunter 2015) basieren auf Prozeduren, die dazu genutzt werden, die Varianz der Studienergebnisse in eine zufällige Stichprobenvarianz und einen systematischen Varianzanteil („between“) zu zerlegen, die anschließend mittels Modera-

toranalysen anhand von Studiencharakteristiken erklärt wird (Hox 2010: 206).<sup>62</sup> Die zugrundeliegende Annahme aller meta-analytischen Modellierungsformen ist die Unabhängigkeit der einzelnen Effektstärken (López-López et al. 2018).

Die grundlegenden Modellierungsformen sind das Fixed-Effects Modell (FE), das Random-Effects Modell (RE) sowie die sogenannte Mixed-Effects Modellierung (ME) (Hedges/Vevea 1998: 486ff.). Die Entscheidung welches Modell genutzt wird, hängt von der Art der Inferenzschlüsse die man ziehen will ab, dem Vorliegen von (unerklärter) Heterogenität, der Teststärke<sup>63</sup> und der Komplexität der damit verbundenen Modellierung (Card 2012: 250).

Die **Fixed-Effects-Modellierung** ist das Grundmodell der Meta-Analyse (Kontantopoulos/Hedges 2009: 280ff.). Dieser Modellierungsform liegt die oftmals nur implizite Annahme zugrunde, dass alle betrachteten Primärstudien der Meta-Analyse aus derselben Grundgesamtheit stammen (Cheung et al. 2012: 131). Es gibt eine einzige „wahre“ Effektstärke für alle Studien und alle Stichproben stammen aus derselben Population. Die Unterschiede der Effektstärken zwischen den Studien kommen lediglich aufgrund des Stichprobenfehlers (zufällige Variation) zustande (Urban/Fiebig 2015: 68; Wagner/Weiß 2014: 1122). In der nachfolgenden Formel ist das grundlegende Fixed-Effects-Modell zu sehen.

$$y_i = \beta_F + e_i$$

$\beta_F$  steht für den gemeinsamen Effekt und der Fehlerterm  $e_i$  steht für den bekannten Stichprobenfehler der Primärstudien. Zu den Fixed-Effects-Modellen gehören somit alle meta-analytischen Modelle, bei denen keine Varianzkomponente (systematische Variation) zwischen den Studien geschätzt wird (Jak 2015: 4). Deshalb wird das FE-Modell auch „common effects“ Modell genannt. Im Gegensatz zu den nachfolgenden Modellen hat das Fixed-Effects Modell in der Regel ein schmaleres Konfidenzintervall (Arends 2006: 13). Diese Modellierung hat jedoch zur Folge, dass die konditionalen („bedingten“) Schlüsse lediglich auf diejenigen Studien beschränkt bleiben, die auch in der Analyse eingeschlossen sind (Cheung 2008: 185). Ein Inferenzschluss darüber hinaus ist unzulässig. Besonders relevant ist eine solche Modellierung, sofern man am mittleren Effekt interessiert ist und es keine statistischen oder analytischen Hinweise für ein bestimmtes Ausmaß an Heterogenität gibt. Wird allerdings fälschlicherweise

---

<sup>62</sup> Neben den, vor allem am Ausmaß der Effektstärken interessierten Varianten, gibt es auch einfachere Verfahren, wie etwa die meta-analytische Auswertung von Vorzeichen oder Signifikanzniveaus, die auch „Vote-Counting-Verfahren“ genannt werden und die vor allem bei vielen fehlenden statistischen Informationen zum Einsatz kommen können (Bushman/Wang 2009: 208ff.; Eisend 2014: 42ff.).

<sup>63</sup> Die Teststärke einer Meta-Analyse basiert auf vielen Faktoren wie der Anzahl der Primärstudien, den Fallzahlen der Primärstudien, der Notwendigkeit für Artefaktkorrekturen, dem Ausmaß der Populationsvarianz und die durchschnittliche wahre Effektstärke der Population (Card 2012: 251).

Homogenität zwischen den Studien angenommen, kann es zu Verzerrungen, wie etwa die Unterschätzung des Standardfehlers, kommen (Eller 2015: 86).

Liegt ein bestimmtes Ausmaß an Heterogenität zwischen den Studien vor, gibt es zwei Möglichkeiten damit umzugehen. Entweder kann ein Random-Effects-Modell geschätzt oder die Heterogenität mittels Moderatoren erklärt werden. Beim zweiten Ansatz besteht die Möglichkeit, Variabilität der Effektstärke über die Studien hinweg zu untersuchen (Card 2012: 198ff.). Heterogenität zwischen den Studien wird dabei als Chance betrachtet, um zu einem besseren Verständnis der Daten und zu nützlicheren Schlussfolgerungen zu gelangen (Arends 2006: 14). Dazu werden die Studiencharakteristiken (bspw. durchschnittliches Alter der Stichprobe, Art der Operationalisierung etc.) der einzelnen Primärstudien kodiert. Anschließend werden diese Variablen mithilfe von Moderatoranalysen dazu genutzt, die Variabilität der Effektstärkekoeffizienten auf die unterschiedlichen Eigenschaften der Studien zurückzuführen. Die Studiencharakteristiken werden dazu in einem linearen Regressionsmodell als Prädiktoren zur Vorhersage der Effektstärken genutzt. Diese Form der Berücksichtigung von Moderatoren in FE-Modellen ist allerdings nur zulässig, wenn fast alle Heterogenität durch die Moderatoren erklärt wird (Jak 2015: 3f.). Bleibt trotz Berücksichtigung der Prädiktoren Restheterogenität bestehen, muss ein Random-Effects-Modell spezifiziert werden.

Das Fixed-Effects-Modell ist ein Spezialfall des mathematisch allgemeineren **Random-Effects-Modells**, das oftmals empirisch und analytisch passender ist (Arends 2006: 13; Card 2012: 254f.).<sup>64</sup> Das RE-Modell erlaubt es, Inferenzschlüsse zu ziehen, die über die in die Analyse eingeschlossenen Primärstudien hinausgehen (Jak 2015: 3). Dies wird als un konditionale Inferenz bezeichnet, da angenommen wird, dass es keinen „wahren“ Populationsparameter gibt, sondern, dass der Wert zwischen den Studien variiert und somit jede Studie eine eigene Populationseffektstärke hat (Urban/Fiebig 2015: 68). Diese Variation kann auf den Stichprobenfehler (zufällige Variation) und Unterschiede in den Studiencharakteristiken (systematische Variation) zurückgeführt werden. In der nachfolgenden Formel ist diese Modellierungsform zu sehen.  $\beta_r$  steht dabei für den durchschnittlichen Populationseffekt.

$$y_i = \beta_r + u_i + e_i$$

Zum einen berücksichtigt das RE-Modell wie das FE-Modell den Stichprobenfehler  $e_i$ , zum anderen die „wahren“ Unterschiede zwischen den Populationseffektstärken ( $u_i$ ). Die beobachtete Verteilung der Effektstärken stellt eine zufällige Auswahl aus der Menge von Studien mit

---

<sup>64</sup> Wenn die Populationsheterogenität „0“ ist, entspricht das RE-Modell dem FE-Modell.

unterschiedlichen Kombinationen der Einflussgrößen dar (Wagner/Weiß 2014: 1122). Das Ergebnis einer solchen Schätzung, ist neben der Schätzung des durchschnittlichen Effektes, die Varianzverteilung der Effektstärken in der Population. Zur Schätzung der Heterogenitätsvarianzen gibt es verschiedene Schätzverfahren. Weit verbreitet sind der ungewichtete Method-of-Moments- (UMM) und der gewichtete Method-of-Moments-Ansatz (WMM)<sup>65</sup>, daneben sind aber auch der Maximum-Likelihood- (ML) und der Restricted (residual) Maximum-Likelihood-Schätzer (REML) möglich (Raudenbush 2009: 308ff.). Heterogenität zwischen den Effektstärken bedeutet, dass vermutlich auch Moderatorvariablen wirksam sind.

Bei einer Random-Effects-Modellierung muss darüber hinaus die Studienanzahl ausreichend hoch sein (Cooper/Hedges 2009a: 566). Während im Fixed-Effects-Modell bereits zwei Studien ausreichend sind, kann im Random-Effects-Modell die Schätzung durch die Berücksichtigung von nur wenigen Studien verzerrt werden (Borenstein et al. 2011: 363). Deswegen sollte bei einer kleinen Anzahl von Studien oder kleinen Fallzahlen, unabhängig von analytischen Überlegungen, stets ein Fixed-Effects-Modell geschätzt werden (Card 2012: 251). Das Random-Effects Modell ist somit das allgemeinere, realistischere und darüber hinaus in Bezug auf die Konfidenzintervalle der kumulierten Effektstärken und deren Signifikanztests auch das konservativere Modell (Eller 2015: 86f.).

Das FE- und RE-Modell unterscheiden sich vorwiegend über die Annahmen zu den „wahren“ Effektstärken, die den Primärstudien zugrunde liegen und in der Inferenz, die sie erlauben (Eller 2015: 86). Es besteht auch die Möglichkeit eine Fixed-Effects Modellierung mit Moderatoranalysen zu ergänzen, wenn dadurch alle Heterogenität erklärt werden kann. Bleibt weiterhin Restheterogenität bestehen, sollte ein Mixed-Effects-Modell, also eine Kombination aus FE und RE, geschätzt werden, um die unerklärte Heterogenität angemessen zu berücksichtigen.

Ein **Mixed-Effects-Modell** liegt vor, wenn trotz Berücksichtigung von Moderatorvariablen weiterhin „Rest-Variabilität“ zwischen den Studien besteht (Jak 2015: 3f.). Diese Modelle haben fixe Regressionskoeffizienten (studienpezifische Eigenschaften) sowie Random-Effects, die die „unerklärte“ studienpezifischen Effekte nach Kontrolle der Prädiktoren (Moderatoren) darstellen (Cheung 2008: 185). Es handelt sich somit um eine Kombination aus der Fixed-Effects-Moderatoranalyse mit der Varianzschätzung des Random-Effects-Modells.<sup>66</sup>

$$y_i = \beta_0 + \beta_1 x_1 + u_i + e_i$$

<sup>65</sup> Definition des Method-of-Moments-Ansatz: “A method of statistical estimation in which the sample moments are equated to their expectations and the resulting set of equations solved for the parameters of interest.” (Cooper/Hedges/Valentine 2009: 578).

<sup>66</sup> Je nach Forschungsbereich spricht man bei dieser Form der Modellierung auch von Meta-Regressionen (Medizin) oder von einem „conditionally random model“ (Arends 2006: 13; Card 2012: 239).

$x_1$  steht für eine Studiencharakteristik einer bestimmten Studie. Die Varianz  $u_i$  steht im Mixed-Effects-Modell für die Residualvarianz nach der Kontrolle von Studiencharakteristiken.  $\beta_0$  und  $\beta_1$  stehen für Intercept und Slope der Schätzung. Für diese Schätzung stehen verschiedene Schätzer zur Verfügung (Viechtbauer et al. 2015: 362ff.).

Moderatoranalysen sind weit verbreitet (Hedges 2016: 210).<sup>67</sup> Die Berücksichtigung der Moderatoren führt in der Regel dazu, dass die Heterogenitätskomponente kleiner wird, aber nicht vollständig verschwindet. Rest-Variabilität ist in sozialwissenschaftlichen Daten kaum verwunderlich, da der Erkenntnisgegenstand komplex und mit einer Vielzahl plausibler Wirkmechanismen und konfundierenden Beziehungen verbunden ist. Aus diesen Gründen wird das Mixed-Effects Modell in den meisten Anwendungsfällen als das beste Modell für die Meta-Analyse betrachtet, da es Moderatorvariablen berücksichtigen kann, aber nicht die strenge Annahme benötigt, dass mit den Moderatorvariablen alle systematische Variation zwischen den Studien erklärt werden kann (van den Noortgate et al. 2013: 590).

Neben den beiden klassischen Herangehensweisen an die meta-analytische statistische Modellierung nach Hedges/Olkin (1985) und Schmidt/Hunter (2015) gibt es darüber hinaus die Möglichkeit der Kombination von Mehrebenenanalyse und Meta-Analyse (Raudenbush/Bryk 1985). Die Mehrebenen-Meta-Analyse (MEMA) macht sich die statistischen Vorteile der Mehrebenenanalyse<sup>68</sup> für die Untersuchung von meta-analytischen Fragestellungen zu eigen (Hox 2010; Urban/Fiebig 2015). Dort lassen sich Fixed- und Random-Effects-Modelle sowie Moderatoreffekte auf Basis von Studiencharakteristiken, in einem einheitlichen Framework schätzen. Die Primärstudien werden in einer MEMA anhand bestimmter studienbezogener Merkmale (bspw. Art der Stichprobengruppen: klinisch, studentisch etc.) in Gruppen eingeteilt. Diese Gruppeneinteilung wird anschließend als Level 2 Einheit in die Modellierung aufgenommen. Somit ist es mittels Random-Effects-Modellierung möglich, die Variation zwischen den Level 2 Einheiten zu berücksichtigen. Um auch Abhängigkeiten zwischen den Effektstärken, etwa aufgrund der Nutzung von mehreren Effektstärken für eine Beziehung pro Studie, berücksichtigen zu können, kann das Mixed-Effects-MEMA-Modell zum Drei-Ebenen-Modell erweitert werden (van den Noortgate et al. 2013: 590).

---

<sup>67</sup> Allerdings kann es bei Moderatoranalysen auch zu schwerwiegenden Problemen kommen (Card 2012: 222ff.). Liegen zwischen den verschiedenen Moderatoren (Studiencharakteristiken) hohe Korrelationen vor, kann es wie bei herkömmlichen Regressionsschätzungen zu Verzerrungen aufgrund von Multikollinearität kommen.

<sup>68</sup> Bei dem Verfahren der Mehrebenenanalyse (MLA) handelt es sich um ein multivariates statistisches Verfahren, das es, ähnlich wie bei einer herkömmlichen Regressionsanalyse, ermöglicht den Einfluss mehrerer unabhängiger Variablen auf eine abhängige Variable simultan zu testen. Zusätzlich zur Regressionsanalyse erlaubt es die MLA unterschiedliche hierarchische Analyseebenen (etwa Befragte und regionale Einheiten wie Kommunen) zu unterscheiden und in der Analyse entsprechend statistisch zu berücksichtigen (Bickel 2007; Hox 2010).

Die Interpretation der meta-analytischen Modelle unterscheidet sich kaum von herkömmlichen Regressions- oder Mehrebenenmodellen, jedoch muss bedacht werden, dass es sich um Phänomene auf Studien- und nicht Individualebene handelt (Cheung/Vijayakumar 2016: 125). Die Gefahr eines ökologischen Fehlschlusses, also dem falschen Schluss von der Aggregatebene auf Individualebene, besteht, aufgrund fehlender Individualdaten in einer APD-Meta-Analyse (Cheung 2015: 99). Besondere Vorsicht ist deshalb bei der Interpretation von Moderatoreffekten, die auf Studiencharakteristiken beruhen, geboten, denn Befunde der Studienebene können nicht auf Individualebene übertragen werden.

### 5.1.2 Zentrale Aspekte

In den nachfolgenden Abschnitten werden spezifische Einzelaspekte der statistischen Umsetzung von Meta-Analysen, unabhängig von der Umsetzungsform, diskutiert. Es geht zunächst darum, welche statistischen Möglichkeiten es gibt, darüber zu entscheiden, ob ein gewisses Ausmaß an **Heterogenität** zwischen den Effektstärken der Studien vorliegt. Ob ein Random-Effects oder ein Fixed-Effects-Modell analysiert wird, sollte vorwiegend anhand der gewünschten Inferenz und der Forschungsfrage entschieden werden und nicht nur aufgrund von statistisch vorliegender Heterogenität (Cheung 2015: 93f.; Hedges 2009: 38). Die grundlegende Frage, die darüber entscheidet, ob ein Fixed-Effects oder ein Random-Effects-Modell geschätzt wird, lautet: Ist die Befundstreuung allein auf den Stichprobenfehler zurückzuführen oder geht sie über diesen hinaus?

Neben der analytischen Begründung kann ein **formaler statistischer Heterogenitätstest** durchgeführt werden. Zur quantitativen Bestimmung der Effektstärkenhomogenität der Primärstudien wird die Varianz berechnet, die sich infolge des Stichprobenfehlers jeder Primärstudie ergibt und anschließend getestet, ob diese Varianz für den Großteil der Gesamtvarianz der Effektstärken verantwortlich ist. Dies wird mit der Chi-Quadrat-basierten Homogenitätsstatistik  $Q$  geprüft (Shadish/Haddock 2009: 262ff.). Zur Durchführung des Signifikanztests muss zunächst das gesamte Ausmaß der Heterogenität aller Effektstärken berechnet werden (Card 2012: 185). Dieser Test gibt an, ob Heterogenität statistisch signifikant vorliegt und testet, ob der berechnete Wert  $Q$  für eine bestimmte Irrtumswahrscheinlichkeit den kritischen Wert übersteigt. Ist dies der Fall, übersteigt die Varianz der Effektstärken die zufällig erwartbare Varianz und somit liegt keine Homogenität zwischen den Studien vor. Die  $Q$ -Statistik hat jedoch eine geringe Teststärke und wird, sobald viele Studien der Berechnung zugrunde liegen, schnell signifikant (Cheung/Vijayakumar 2016: 124). Auch trifft die  $Q$ -Statistik keine Aussage über das Ausmaß der vorliegenden Heterogenität (Cheung 2015: 92).

$I^2$  ist eine nützliche Ergänzung zum  $Q$ -Wert, da es über das Ausmaß an Heterogenität Auskunft gibt (Borenstein et al. 2017: 5ff.). Diesem Koeffizienten liegt die Variabilitätszerlegung der Effektstärken zugrunde (Shadish/Haddock 2009: 263f.). Die Variabilität der Effektstärken zwischen den Studien (between Variabilität) wird mit der Gesamtvariabilität (between + within) ins Verhältnis gesetzt. Die zugehörige Regel besagt, dass ein Wert von 25 % als niedrig, ein Wert von 50 % als moderat und ein Wert von 75 % als hoch angesehen wird (Eisend 2014: 48; Urban/Fiebig 2015: 68). Ein Datensatz ist demnach heterogen, umso höher das Variabilitätsverhältnis ist. Ist der Wert hingegen klein, liegt umgekehrt Homogenität zwischen den Effektstärken vor. Bei dem Kennwert handelt es sich lediglich um eine relative Messung der Heterogenität. Wird der Stichprobenfehler kleiner, wird der Wert automatisch größer (Cheung/Vijayakumar 2016: 125). Mit der Spannbreite der Effekte hat  $I^2$  dementsprechend nichts zu tun (Borenstein et al. 2017: 12ff.). Im Anschluss an den formalen Heterogenitätstest und in Übereinstimmung mit den analytischen Vorüberlegungen zum angestrebten Inferenzschluss wird entschieden, ob eine Fixed-Effects, Random-Effects Modellierung oder eine Mixed-Effects-Modellierung erfolgt.

Die **Analyseeinheit** einer jeder Art von Meta-Analyse besteht in der sogenannten Effektstärke (Klein/Kopp/Rapp 2013: 224; Kopp/Lois 2014: 209). Je nach Skalenniveau der relevanten Variable stehen verschiedene Effektstärkekoeffizienten zur Verfügung. Es lassen sich drei Familien von Effektstärken unterscheiden (Card 2012: 87ff.; Urban/Fiebig 2015: 48ff.)<sup>69</sup> Effektstärken auf Basis von Mittelwertdifferenzen (und Standardabweichungen) stellen die sogenannte  $d$ -Gruppe dar (bspw. Cohen  $d$ ; Hedge  $g$ , unstandardisierte (rohe) Mittelwertdifferenz). In dieser Gruppe werden Vergleiche auf Basis von (Treatment-)Mittelwerten durchgeführt. Die zweite Gruppe von Effektstärken sind diejenigen, die auf binären Daten (Kreuztabellen) basieren. Dazu gehören das relative Risiko („risk-ratio“), das Chancenverhältnis („odds-ratio“), die Risikodifferenz („risk difference“) und der Risikoquotient. Die dritte Gruppe stellen diejenigen Effektstärken dar, die auf Produkt-Moment-Korrelationen basieren (bspw. Phi, Rho, Pearsonsche Korrelationskoeffizient  $r$ ). Sie stehen im Zentrum dieser Arbeit. Ihre ohnehin standardisierten Werte lassen sich zwischen den Studien leicht vergleichen und zusammenfassen (Cheung 2015: 55).<sup>70</sup> Da Korrelationen jedoch von den Varianzen der Untersuchungsvariablen

---

<sup>69</sup> Eine genaue Übersicht über die unterschiedlichen Effektstärkegrößen und deren Berechnung findet sich an anderer Stelle (Borenstein et al. 2011: 21ff.; Eisend 2014: 3ff.; Urban/Fiebig 2015: 50ff.). Dort sind ebenfalls die Transformationsregeln zwischen den verschiedenen Effektstärken zu finden. Ebenso finden sich dort Hinweise für Effektstärken auf Basis von univariaten Untersuchungsergebnissen (Mittel- und Modalwerte).

<sup>70</sup> Ist der  $r$ -Koeffizient nach Pearson in einer Primärstudie nicht angegeben, kann er aus anderen statistischen Informationen berechnet werden. Eine Berechnung ist etwa aus Inferenzstatistiken, deskriptiven Daten, Signifikanzlevel und dem Ergebnissen des Omnibustests möglich (Card 2012: 96ff.).

abhängen, empfehlen manche Autoren die Transformation in z-Werte nach Fisher und die Analysen mit den transformierten Werten durchzuführen (Borenstein et al. 2011: 41ff.; Hedges 2009: 46). Anschließend soll für die Ergebnispräsentation jedoch eine Rücktransformation in r vorgenommen werden, um die Ergebnisse inhaltlich interpretierbar zu machen (Card 2012: 89).<sup>71</sup> Diese Transformationen werden auch als Varianzstabilisierungstransformationen bezeichnet (Cheung 2015: 50).

Eine **Gewichtung** der einzelnen Effektstärken ist bei der Aggregation möglich. Dies wird durch die Berücksichtigung eines Gewichtungsfaktors vorgenommen (Card 2012: 176ff.; Shadish/Haddock 2009: 258ff.). Die Grundidee ist, dass Studien mit einer größeren Stichprobenvarianz (und kleinerer Fallzahl) weniger zum gewichteten mittleren Effekt beitragen (Jak 2015: 3). Eine Gewichtung ist etwa durch die Berücksichtigung der unterschiedlichen Stichprobengrößen der einzelnen Primärstudien möglich. Aber auch andere Gewichtungen, etwa die inverse Effektstärkenvarianz oder das Ausmaß von Messverzerrungen, sind denkbar (Eisend 2014: 44). Die Inverse der Stichprobenvarianz der Effektstärken gilt als optimaler Gewichtungsfaktor (Matt/Cook 2009: 547).

Bei der Durchführung empirischer Studien kann es sowohl bei der Operationalisierung als auch bei der Datenerhebung zu Verzerrungen kommen, die einen systematischen Einfluss auf die Zusammenhangsmaße haben und die studienübergreifende Vergleichbarkeit einschränken. Da alle Messungen anfällig für Messfehler sind, entsprechen die beobachteten Korrelationskoeffizienten nicht den wahren Korrelationen (Cheung 2014a: 39). Sind die Ursachen für die Verzerrungen bekannt, können diese mit den entsprechenden **Artefaktkorrekturverfahren** korrigiert werden. So ist es etwa möglich, Korrekturen in Bezug auf die Reliabilität von Messinstrumenten, die Validität der gemessenen Variablen und nachträgliche künstliche Dichotomisierung von kontinuierlichen Variablen vorzunehmen (Card 2012: 126ff.).<sup>72</sup>

Allerdings gibt es Kritik an dieser Praxis. Rosenthal (2010) und andere Autoren sprechen sich gegen Artefaktkorrektur aus, da sich die korrigierten Werte von den typischen wissenschaftlichen Funden unterscheiden und somit weniger aussagekräftig sind, als die unkorrigierten realistischen Ergebnisse (Cheung 2014a: 38). Diese Verfahren führen zu methodologischen Problemen, wie etwa größere Effektstärken und Veränderungen der Standardfehler (Hox 2010:

---

<sup>71</sup> In der meta-analytischen Literatur gibt es eine Kontroverse über die Verwendung von Korrelationen (Befürworter: Schmidt/Hunter (2015)) oder von Fisher z-Werten (Befürworter: Hedges/Olkin (1985)).

<sup>72</sup> In Schmidt/Hunter (2015) ist eine ausführliche Diskussion der Möglichkeiten zur Artefaktkorrektur zu finden. Sie zählen insgesamt elf Artefakte auf, die mit den entsprechenden Methoden korrigiert werden können.

218). Bei der Anwendung dieser Verfahren kommt es deshalb darauf an, ob man an einer theoretischen oder an einer mittels empirischer Daten beobachteten Beziehung interessiert ist (Cheung/Hong 2017).

Eines der größten Probleme, mit dem jede Form der Meta-Analyse zu kämpfen hat, sind **statische Abhängigkeiten zwischen Effektstärken**. Die Verletzung der grundlegenden Unabhängigkeitsannahme führt zu Verzerrungen (Matt/Cook 2009: 546ff.; Riley 2009: 807; van den Noortgate et al. 2013: 589). So kann es zu verzerrten Effekten, verzerrten Standardfehlern, zu schmalen Konfidenzintervallen und zur falschen Zurückweisung der Null-Hypothese (Fehler 1. Art) kommen (Becker 2000: 523; Mavridis/Salanti 2013: 133; van den Noortgate et al. 2013: 577). Diese Abhängigkeiten kommen auf verschiedene Arten zustande. Sie können entstehen, wenn mehrere relevante Variablen in der gleichen Stichprobe erhoben wurden oder aufgrund des Vergleichs der Wirksamkeit von unterschiedlichen „treatments“ (bspw. unterschiedliche Medikamente) in verschiedenen Gruppen im Hinblick auf eine gemeinsame Kontrollgruppe (Eisend 2014: 80; Gleser/Olkin 2009: 358). Oftmals werden verschiedene Messinstrumente zur Erfolgskontrolle des „treatments“ genutzt (bspw. körperliche und psychische Gesundheit). Zur Bekämpfung dieses Problems liegen verschiedene ad-hoc Verfahren (Mittelwertbildung, Entscheidungsregel, Zufallsauswahl und Ignorieren), aber auch anspruchsvollere Modellierungsformen vor. Weitere Ausführungen zu diesem Aspekt finden sich in Kapitel 5.4.

Die hier aufgeführten klassischen Methoden, die sich in vielen Forschungsbereichen bewährt haben, sind oftmals zu vereinfachend zur Zusammenfassung aktueller empirischer Studien (Bergh et al. 2014: 478). Denn mit den herkömmlichen Verfahren können umfassende theoretisch-analytisch hergeleiteten Modelle bzw. Theorien mit komplexen Variablenbeziehungen (Mediatoren), wie sie in den Sozialwissenschaften oder den Gesundheitswissenschaften existieren, nicht getestet werden. Eine solche Möglichkeit bietet bislang nur die modellbasierte Meta-Analyse. Diese macht realistischere Annahmen darüber, wie ein Effekt zustande kommt, weil die dahinterliegenden Mechanismen spezifiziert und getestet werden (Shadish/Sweeney 1991: 885). Die theoretisch-analytischen Modelle können dabei komplexer als in Primärstudien sein, da es für das Verfahren nicht relevant ist, dass in jeder Studien alle relevanten Variablenbeziehungen untersucht wurden (Jak 2015: 12).

## 5.2 SEM-basierte Meta-Analyse

In diesem Abschnitt wird die sogenannte SEM-basierte Meta-Analyse erläutert, um diese von MASEM abgrenzen zu können. Dabei handelt es sich um kein modellbasiertes Verfahren, sondern es werden herkömmliche meta-analytische Modelle als Strukturgleichungsmodell ausgedrückt. In einer SEM-basierten Meta-Analyse werden also keine latenten Variablen oder Pfadmodelle wie bei einer MASEM analysiert, sondern lediglich die Flexibilität des SEM-Ansatzes sowie der zugehörigen EDV-Programme genutzt (Card 2012: 219). Der größte Vorteil der SEM-basierten Meta-Analyse liegt darin, dass nicht nur Korrelationskoeffizienten, sondern alle bekannten Effektstärken eingesetzt werden können.

In der SEM-basierten Meta-Analyse wird das Maximum-Likelihood-Schätzverfahren angewandt (Cheung 2015: 89). Im Fixed-Effects-Modell der SEM-basierten Meta-Analyse wird die Varianz des Messfehlers als die bekannte Stichprobenvarianz festgelegt, damit in diesem Modell nur ein frei zu schätzender Parameter existiert (Cheung 2015: 100). Die Effektstärken und die zugehörigen Prädiktoren (Moderatoren) werden reskaliert (Card 2012: 219ff.). Dabei wird jede Studie zum gleich gewichteten Fall in der Analyse, da die Gewichtung bei der Transformation berücksichtigt wird. Eine Gewichtung während der Analyse ist dementsprechend nicht notwendig.<sup>73</sup>

Im Random-Effects-Modell hingegen werden die studienspezifischen Random-Effects als latente Variable behandelt (Cheung 2015: 101). Die Umsetzung erfolgt ähnlich wie beim Fixed-Effects-Modell, allerdings wird die Effektstärke, die vom Intercept-Pfad vorhergesagt wird, als Random-Slope modelliert (Card 2012: 246). Das RE-Modell der SEM-basierten Meta-Analyse ist ein einfaktorielles Ein-Indikator-CFA-Modell, bei dem der Messfehler fixiert wird. Auch ein Mixed-Effects-Modell kann im Kontext der SEM-basierten Meta-Analyse umgesetzt werden. Dazu werden die transformierten (gewichteten) Studiencharakteristiken (Moderatorvariablen) als fixierte Prädiktoren ins Modell aufgenommen. Der Pfad von der Moderatorvariable zur Effektstärke evaluiert dann den Moderatorzusammenhang.

---

<sup>73</sup> Zur Umsetzung dieser Modellierungsstrategie gibt es zwei unterschiedliche Möglichkeiten. Entweder können die Effektstärken so transformiert werden, dass sie eine gemeinsame bekannte Stichprobenverteilung mit einer Varianz von 1 haben oder man kann eine Definitionsvariable dazu benutzen, die bekannte Stichprobenvarianz als den bekannten Stichprobenfehler festzulegen. Um eine durchschnittliche mittlere Effektstärke zu erhalten, wird ein Modell geschätzt, in dem eine Regression von der transformierten Effektstärke auf den transformierten Intercept geschätzt wird. Der Intercept stellt dabei eine Konstante von 1.0 dar. Zur Umsetzung wird zusätzlich die Varianz der transformierten Effektstärke auf 1 sowie der Indikator-Intercept auf 0 gesetzt. Somit wird die durchschnittliche Effektgröße durch den Regressionskoeffizienten des transformierten Intercepts repräsentiert.

Auch der Restricted (residual) Maximum-Likelihood-Schätzer (REML) kann für die Schätzung von SEM-basierten Meta-Analysen verwendet werden, da der ML-Schätzer eine negative Verzerrung bezüglich der geschätzten Varianzkomponenten aufweist und unverzerrte Schätzergebnisse nur bei großen Fallzahlen liefert (Cheung 2015: 279ff.). Allerdings sind die Fallzahlen, also die Anzahl der Primärstudien, in Meta-Studien oftmals gering, so dass hier Verzerrungen erwartet werden können. Die Verwendung des REML-Schätzers hat den Nachteil, dass Modellvergleiche in einem Fixed-Effects Modell nicht möglich sind. Ebenfalls muss auf die Möglichkeiten des ML-Schätzers, wie etwa die Verwendung von robusten Statistiken, verzichtet werden. Aus diesem Grund empfiehlt Cheung (2015) die Verwendung von REML nur, wenn die Fallzahl sehr klein ist oder die Varianzkomponente im Forschungsdesign eine wichtige Rolle spielt.

SEM-Programme wie *Mplus* können in einer SEM-basierten Meta-Analyse direkt mit fehlenden Werten in den Moderatoren umgehen, indem sie den Full-Information-Maximum-Likelihood -Schätzer<sup>74</sup> verwenden und mithilfe von Definitionsvariablen die Prädiktoren (Moderatoren) ebenfalls zu abhängige Modellvariablen machen (Card 2012: 319ff.; Cheung 2015: 292ff.). Gerade in einer Meta-Studie ist dies ein kaum zu überschätzender Vorteil, da der Verlust von einigen Fällen (Studien) extreme Auswirkungen auf die Teststärke und somit auf die Schlussfolgerungen der Meta-Analyse haben kann.

### **5.3 Meta-Analytic-Structural-Equation-Modelling-Ansätze (MASEM)**

Die Grundidee der MASEM-Verfahren liegt in der Zusammenfassung und Analyse eines Modells mittels Strukturgleichungsmodell (Becker 2001: 191ff.; Cheung/Sun/Chan 2018: 2). Zunächst stellt sich die Frage, wieso zur multivariaten Meta-Analyse nicht direkt die Ergebnisse von multivariaten Regressionsanalysen zur Zusammenfassung von Studienergebnissen genutzt werden. In Becker/Wu (2007) und Wu/Becker (2013) werden einige der bisherigen Ansätze zusammengefasst und verglichen. Dabei hat sich gezeigt, dass die bisherigen Ansätze nicht geeignet sind oder zumindest problembehaftet sind, da etwa die Kovarianzen zwischen den Slopes vorliegen müssen (Becker 2009: 393f.; Becker/Wu 2007: 426). Nahezu keine empirische Studie liefert Informationen in diesem Umfang. In Abbildung 11 ist eine schematische Übersicht, über die in dieser Arbeit diskutierten, auf Basis von Korrelationen durchführbaren, MASEM-Verfahren zu sehen.

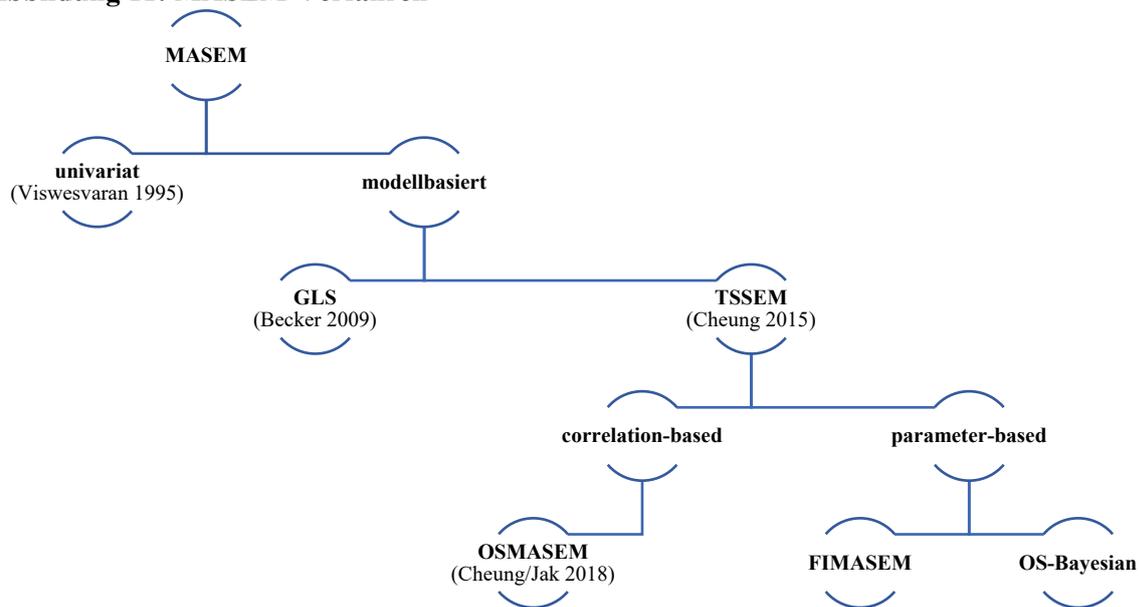
Die MASEM-Analyse gliedert sich, unabhängig vom konkreten Ansatz, in zwei Phasen (Jak 2015: 16). In der ersten Phase wird eine kumulierte Korrelationsmatrix erstellt, auf deren Basis

---

<sup>74</sup> Siehe dazu Enders (2010: 86).

in der zweiten Phase ein Strukturgleichungsmodell getestet wird. Dabei wird die kumulierte Korrelationsmatrix jeweils als beobachtete Kovarianzmatrix analysiert. Dieses Vorgehen ist aus mathematischer Sicht zunächst einmal problematisch, da die statistischen Grundlagen der Strukturgleichungsmodellierung auf der Analyse von Kovarianzmatrizen beruhen (Little/Kline 2016: 13ff.). Eine MASEM-Umsetzung auf deren Basis ist jedoch aufgrund von unterschiedlichen Skalierungen der Messinstrumente zwischen den Studien oder fehlenden statistischen Informationen, um die Kovarianzen berechnen zu können, selten möglich (Cheung/Cheung 2016: 140f.; Sheng et al. 2016: 187).

**Abbildung 11: MASEM-Verfahren**



**Anmerkung:** Eigene Abbildung in Anlehnung an Reyes-Gómez/Obeidat (2016) und ohne die verschiedenen Modifikationen des GLS-Ansatzes.

In der zweiten Phase wird die kumulierte Korrelationsmatrix, die alle interessierenden Variablen enthält, als Datengrundlage für den Test des abgeleiteten Strukturgleichungsmodells genutzt („Modellfitting“). MASEM kombiniert somit die Grundgedanken der Strukturgleichungsmodellierung mit denen der Meta-Analyse, indem Studien anhand der Korrelationsmatrizen zusammengefasst werden, um daraus möglichst generalisierbare Schlüsse auf einem äußerst starken empirischen Fundament ziehen zu können (Cheung 2014a: 29).

Verschiedene Aspekte sind bei der Auswahl eines konkreten Analyseansatzes entscheidend. Landis (2013) zählt etwa drei zentrale Entscheidungspunkte auf, die berücksichtigt werden müssen (Landis 2013: 254ff.):

- Wie gehe ich mit fehlenden Werten um?
- Welche Stichprobengröße wird in der Analyse verwendet?
- Welcher Ansatz soll bei der Umsetzung der MASEM-Analyse verwendet werden?

Eine ähnliche Aufzählung findet sich auch beim Literaturreview von Sheng et al. (2016), die ihren Review auf Basis der folgenden Entscheidungspunkte durchgeführt haben:

- Wie soll die Input-Matrix gefüllt werden?
- Listenweiser Fallausschluss vs. Paarweiser Fallausschluss?
- Umgang mit leeren Zellen/fehlenden Werten?
- Welcher Analyseansatz soll verwendet werden?
- Welche Stichprobengröße soll für die Analyse genutzt werden?
- Was passiert bei nicht positiv definiten Matrizen?
- Werden Moderatorvariablen untersucht?
- Kovarianz vs. Korrelation?

Die nachfolgenden Ausführungen werden eine klare Antwort auf diese Fragen geben. In einer modellbasierten Meta-Analyse ist es besonders wichtig, dass nur Korrelationen zur Beantwortung der Forschungsfrage genutzt werden, die auch angemessen, das heißt der Operationalisierung entsprechend, sind (Sheng et al. 2016: 199ff.). Ebenfalls sollte bei der Erstellung der (durchschnittlichen) Korrelationsmatrix besondere Sorgfalt walten. Auch die verbreitete Praxis auf bereits bestehende meta-analytische Ergebnisse (Korrelationen) zurückzugreifen, muss kritisch betrachtet werden, da dieses Vorgehen mathematisch nicht gesichert ist.

### 5.3.1 Univariater $r/z$ -Ansatz

In einem ersten Schritt wird nun der älteste und am weitesten verbreitete Ansatz diskutiert. Dieser Ansatz baut konzeptionell auf Studien auf, die bereits die Kombination von Strukturgleichungsmodellierung und Meta-Analyse zur kausalen Modellierung eingesetzt haben (Brown/Peterson 1993; Peters/Hartke/Pohlmann 1985), welche wiederum auf den vorgeschlagenen meta-analytischen Prozeduren von Hunter/Schmidt/Jackson (1982) basieren. Dabei handelt es sich um den von Viswesvaran/Ones (1995) vorgeschlagenen univariaten  $r/z$ -Ansatz. Dieser bleibt jedoch auf die bivariate Zusammenfassung der einzelnen Beziehungen beschränkt und beruht auf keiner konsistenten statistischen Theorie (Cheung/Cheung 2016: 141).

Beim **univariaten  $r/z$ -Ansatz** (Viswesvaran/Ones 1995) werden die Korrelationskoeffizienten einzeln auf Grundlage der bivariaten Beziehungen aus allen Primärstudien zusammengefasst, als ob sie voneinander unabhängig wären (Cheung 2015: 218; Jak 2015: 16). Mit herkömmlichen meta-analytischen Methoden wird eine kumulierte Korrelationsmatrix aus den einzelnen Korrelationskoeffizienten gebildet. Anschließend dient diese Matrix als Grundlage für den Test eines Strukturgleichungsmodells. Viswesvaran/Ones (1995) interpretieren die erste Phase von MASEM als Messmodell und die zweite Phase als Strukturmodell (siehe Tabelle 17).

Wie zu sehen ist, entspricht die erste Phase (Messmodell) fast vollständig den Arbeitsschritten einer Meta-Analyse. Viswesvaran/Ones (1995) folgen dabei dem Ansatz von Schmidt/Hunter

(2015) und fassen die Primärstudien zunächst mittels vereinfachter psychometrischer Meta-Analyse zusammen (Cheung/Chan 2005b: 41). Dieser Ansatz wird als univariater  $r$ -Ansatz bezeichnet, da die untransformierten Korrelationskoeffizienten verwendet werden (Cheung 2015: 218). Ein Populationswert (mittlere Effektstärke) wird zunächst für jeden einzelnen Korrelationskoeffizient getrennt geschätzt und eine Gewichtung, etwa mit der inversen geschätzten Stichprobenvarianz (quadrierter Standardfehler), vorgenommen (Jak 2015: 16; Tang/Cheung 2016: 476).

**Tabelle 17: Schritte einer MASEM nach Viswesvaran/Ones 1995**

**Erste Phase: MASEM-Messmodell**

- 
1. Identifikation relevanter Variablen und Beziehungen
  2. Identifikation der unterschiedlichen Messungen und Operationalisierungen
  3. Identifikation und Zusammentragen der relevanten Primärstudien
  4. Durchführung der (psychometrischen) Meta-Analyse und Erstellung der kumulierten Korrelationsmatrix
  5. Überprüfung mittels Faktoranalyse, ob die konzeptionell verschiedenen Messungen und Operationalisierungen einer gemeinsamen Variable zugeordnet werden können
- 

**Zweite Phase: MASEM-Strukturmodell**

- 
1. Berechnung von zusammengesetzten Korrelationen für Variablen, die mittels unterschiedlicher Operationalisierungen gemessen wurden
  2. Schätzung des Pfadmodells auf Grundlage der kumulierten Korrelationskoeffizienten
- 

**Anmerkungen:** In Anlehnung an Viswesvaran/Ones (1995: 867); Übersetzung nach Eller (2015: 92).

Dieses Vorgehen berücksichtigt nicht die vorliegenden Abhängigkeiten der Korrelationen innerhalb von Studien und verletzt somit die statistische Unabhängigkeitsannahme. Dies führt zu Verzerrungen und kann im schlimmsten Fall dazu führen, dass die Schlüsse daraus falsch sind. Ähnlich wie bei den herkömmlichen Verfahren der Meta-Analyse wird empfohlen, Fishers  $z$ -Transformation durchzuführen, bevor die Korrelationskoeffizienten zusammengefasst werden (Cheung/Chan 2005b: 41). Zur Anwendung kommt dabei das Vorgehen nach Hedges/Olkin (1985). Die Unterschiede zwischen den beiden Ansätzen sind jedoch gering, im Random-Effects-Modell weist der  $r$ -over- $z$  Ansatz nach Hedges/Olkin (1985) eine leicht höhere Leistungsfähigkeit auf.<sup>75</sup>

Unabhängig von der Methode zur Zusammenfassung wird anschließend eine Korrelationsmatrix gebildet. Diese wird zum Test eines Strukturgleichungsmodells verwendet. Für den Test des Strukturgleichungsmodells wird eine einzige Stichprobengröße benötigt. Das Problem beim univariaten Ansatz besteht darin, zu entscheiden, welche Stichprobengröße bei der Analyse

---

<sup>75</sup> Abschließend ist die Frage jedoch noch nicht beantwortet, welcher der beiden univariaten Ansätze eine höhere Performanz aufweist (Cho 2015: 15; Jak 2015: 16f.).

verwendet werden soll (Viswesvaran/Ones 1995: 877). Dazu stehen verschiedene Methoden, wie das arithmetische Mittel, der harmonische Mittelwert, der Median oder die totale Stichprobengröße, zur Verfügung. Die Zellen der Matrix basieren somit auf unterschiedlichen Fallzahlen, da üblicherweise nicht alle Studien für alle Zellen Werte liefern. Der Fehler 1. Art der Chi-Quadrat-Statistik, die Fit-Indizes, die Teststärke und die Standardfehler der Parameterschätzungen hängen jedoch von der jeweils ausgewählten Stichprobengröße ab, so dass unterschiedliche Stichprobengrößen zu unterschiedlichen Ergebnissen und Schlussfolgerungen führen (Cheung 2015: 220).

### 5.3.2 GLS-Ansatz

Zunächst wird nun der multivariate GLS-Ansatz (Becker 1992, 1995, 2009; Becker/Schram 1994) konzeptionell vorgestellt.<sup>76</sup> Beim GLS-Ansatz handelt es sich um ein zweistufiges multivariates Verfahren, das im Gegensatz zum univariaten Ansatz, die einzelnen Elemente der Korrelationsmatrizen der Primärstudien als voneinander abhängig betrachtet. Zunächst wird eine kumulierte Korrelationsmatrix über alle Studien hinweg geschätzt, mit deren Hilfe anschließend spezifische Muster aufgedeckt werden. Abschließend wird die Variabilität zwischen den Studien mittels Moderatoranalysen erklärt. Mathematisch basiert der GLS-Ansatz auf dem Generalized-Least-Square-Ansatz<sup>77</sup> nach Raudenbush/Becker/Kalaian (1988). Es wird nicht nur die Stichprobenvarianz innerhalb jeder Studie zur Gewichtung der Korrelationskoeffizienten verwendet, sondern auch die Stichprobenkovarianz (Jak 2015: 17).

Der GLS Ansatz lässt sich systematisch in mehreren aufeinanderfolgenden Schritten beschreiben (Card 2012: 290ff.). Für die Umsetzung ist eine ausreichend große Datengrundlage notwendig (Becker 2009: 393). Die Daten müssen so geordnet werden, dass die empirischen Informationen der Studien in zwei Matrizen organisiert sind (Jak 2015: 18). In der ersten Matrix befindet sich ein Vektor mit den beobachteten Korrelationen. Die zweite Designmatrix ist eine Indikatormatrix (Dummyvariable), in der sich die Informationen darüber befinden, welche Korrelationen in der jeweiligen Primärstudie untersucht werden (Becker 2009: 386f.). Der Ansatz kann dabei mit fehlenden Korrelationen umgehen.

---

<sup>76</sup> Für eine mathematisch detaillierte Darstellung siehe die Ausführungen von Becker (Becker 2009: 377ff.) und für eine an einem Beispiel leicht nachvollziehbare Darstellung siehe Card (2012).

<sup>77</sup> Unterschiede zwischen den Schätzern im SEM- und Meta-Analysen-Kontext: "In the literature of meta-analysis and regression analysis, WLS and GLS are used to handle the error structures with a diagonal and a block-diagonal variance-covariance matrix, respectively. In other words, WLS is a special case of GLS with uncorrelated residuals. In the context of SEM, GLS is a special case of WLS. The GLS estimation method is used with a normality assumption on the data, whereas the WLS estimation method can be used for data with arbitrary distributions." (Cheung 2015: 25).

Im nächsten Schritt müssen die Varianzen und Kovarianzen der Effektstärkeschätzungen berechnet werden. Anschließend wird eine kumulierte Fixed-Effects-Korrelationsmatrix geschätzt, die alle empirischen Informationen beinhaltet. Daran anknüpfend wird, wie auch in der herkömmlichen Meta-Analyse üblich, ein statistischer Homogenitätstest durchgeführt, der evaluiert, ob Heterogenität zwischen den Effektstärken vorliegt. Variiert eine der Effektstärken systematisch zwischen den Studien, sollte im Einklang mit analytischen Überlegungen, eine Random-Effects-Korrelationsmatrix geschätzt werden. Dazu wird eine Schätzung der between-Studienvarianz für jede Effektstärke benötigt. Somit wird eine Gewichtung mit der Stichprobenvarianz und der between-Stichprobenvarianz vorgenommen. Zur Umsetzung dieser Schätzung wird im GLS-Ansatz der sogenannte „Method of Moments“-Ansatz verwendet, der auf der  $Q$ -Statistik des Fixed-Effects-Modells basiert (Jak 2015: 29). Anschließend kann ein multivariates Modell, bspw. ein Pfadmodell mit der kumulierten Korrelationsmatrix getestet werden (Becker 2009: 391f.). Ein solches Modell liefert Regressionskoeffizienten für die Prädiktoren im untersuchten Modell und Schätzer für die Varianzaufklärung des spezifizierten Modells. Kategoriale Moderatorvariablen können im finalen Schritt untersucht werden. Dafür werden separate Modelle für die unterschiedlichen interessierenden Gruppen (bspw. klinische Stichproben vs. studentische Stichproben) spezifiziert und miteinander verglichen.

Verschiedene Modifikationen des Ansatzes wurden bislang vorgeschlagen, da die Schätzungen des ursprünglichen GLS-Ansatzes instabil sind (Jak 2015: 18). Becker/Fahrbach (1994) haben etwa vorgeschlagen, im GLS-Ansatz Korrelationen auf Basis der  $z$ -transformierten Korrelationen für die Analyse zu verwenden (Becker 2009: 388ff.). Ein Problem bei der Verwendung von  $z$ -Werten ist jedoch, dass die Interpretation der Parameterschätzungen wenig intuitiv ist. Dieses Problem könnte umgangen werden, wenn nach der Analyse die  $z$ -Werte zurück in einen Korrelationsvektor transformiert werden (Cheung 2015: 243).<sup>78</sup> Allerdings besteht Cheung/Chan (2005b) zufolge der große Nachteil dann darin, dass die asymptotische Kovarianzmatrix der  $z$ -Werte nicht ohne weiteres als Gewichtungsmatrix in der 2. Analysephase verwendet werden kann. Deshalb wird dieser Ansatz bislang kaum verfolgt und muss bezüglich der Performanz noch weiter überprüft werden (Zhang 2011: 13). Der modifizierte GLS-Ansatz nach Becker/Fahrbach (1994) hat laut Simulationsstudien eine höhere Leistungsfähigkeit und damit einhergehend weniger verzerrte Schätzungen als der ursprüngliche GLS-Ansatz (Cho

---

<sup>78</sup> Dieses Vorgehen ist ähnlich wie das von Hedges/Olkin (1985) vorgeschlagene Vorgehen für die herkömmliche Meta-Analyse, bei dem die Analyse zunächst auch mit den  $z$ -Werten vorgenommen wird, diese anschließend jedoch für die Interpretation zurück in  $r$ -Werte transformiert werden.

2015: 20). Bei diesem Ansatz wird der Mittelwert der relevanten Korrelationen für die Schätzung der Populationskorrelationen verwendet (Furlow/Beretvas 2005: 229). Dies führt, den Simulationsergebnissen von Zhang (2011) nach, zu weniger verzerrten Parameterschätzungen und Standardfehlern als in der ursprünglichen GLS-Variante und der Fehler 1. Art wird besser kontrolliert. Eine weitere Modifikation hat Cheung (2001) vorgeschlagen. Dabei wird der Ansatz nach Becker/Fahrbach (1994) um die Gewichtung jeder Korrelation mit der zugehörigen Stichprobengröße erweitert. Dieser Ansatz ist nach der Simulationsstudie von Yuan/Kano (2018) mathematisch äquivalent und liefert empirisch die gleichen Ergebnisse wie der TSSEM und der ML-MASEM-Ansatz. Diese Erweiterungen werden in der Praxis jedoch kaum eingesetzt und es wird oftmals die ursprüngliche Variante verwendet.

Furlow/Beretvas (2005) haben eine zusätzliche Erweiterung des Cheung (2001) Ansatzes vorgeschlagen, in der eine inverse Varianzgewichtung angewandt wird. Dieses Verfahren wird als W-COV GLS-Ansatz bezeichnet und hat im Vergleich zu allen anderen vorgestellten GLS-Modifikationen insgesamt die höchste Leistungsfähigkeit (Furlow/Beretvas 2005: 251ff.; Zhang 2011: 55ff.).

Die vorgeschlagenen Modifikationen liefern präzisere Schätzergebnisse als der ursprüngliche GLS-Ansatz, sowie ähnlich präzise Parameterschätzungen und Standardfehler wie der TSSEM-Ansatz. Darüber hinaus haben die vorgestellten Modifikationen hinsichtlich des Fehlers 1. Art beim Homogenitätstest sogar eine höhere Leistungsfähigkeit. Warum diese Modifikationen des GLS-Verfahrens dem TSSEM-Ansatz nicht vorgezogen werden sollten, ergibt sich aus dem folgenden Abschnitt, indem die gegen den univariaten  $r/z$ -Ansatz und dem GLS-Verfahren vorgebrachten Einwände systematisiert werden. Aus der Diskussion dieser Schwächen ergibt sich der praktische Nutzen des TSSEM-Ansatzes für die empirische Forschung.

### **5.3.3 Systematisierung der Probleme von univariatem und GLS-Ansatz**

Zunächst werden die Probleme des univariaten  $r/z$ -Ansatzes aufgeführt, bevor daran anknüpfend die GLS-spezifischen Schwächen diskutiert werden. In Tabelle 18 ist eine Übersicht der wesentlichen Schwächen beider Ansätze zu sehen.

Das folgenreichste Problem des univariaten MASEM-Ansatzes liegt darin, dass die Korrelationsmatrizen so analysiert werden, als ob es beobachtete Kovarianzmatrizen wären (Cheung/Cheung 2016: 142). Die Schätzung der Parameter in Strukturgleichungsmodellen basiert auf der Stichprobentheorie von Kovarianzmatrizen. Werden Strukturgleichungsmodelle jedoch mit Korrelationsmatrizen getestet, kann dies zu Verzerrungen der Standardfehler und der Chi-Quadrat-Statistik führen. Eine weitere große Schwäche des univariaten Ansatzes ergibt

sich daraus, dass die kumulierte Korrelationsmatrix erstellt wird, indem die Homogenität der einzelnen Zellen (Korrelationen) getrennt voneinander getestet wird (Steinmetz/Isidor 2013: 2f.). Die Korrelationen werden so behandelt, als ob sie unabhängig voneinander sind. Es können somit lediglich bivariate Beziehungen fehlerfrei untersucht werden. Einer der größten Vorteile von SEM, nämlich die simultane Schätzung von mehreren Beziehungen, bleibt somit völlig ungenutzt.

Eine weitere Schwierigkeit entsteht aus der Gegebenheit, dass die gefundenen Primärstudien in der Regel nicht alle Korrelationen für die Matrix liefern (Card 2012: 285). Daraus ergeben sich zwei Probleme. Die Korrelationskoeffizienten der Korrelationsmatrix basieren auf einer unterschiedlichen Anzahl von Primärstudien, da die Matrix üblicherweise durch Mittelwertbildung mit paarweisem Fallausschluss berechnet wird. In manchen Fällen kann dies zu einer nicht lösba- ren, nicht positiv definiten Matrix führen (Jak 2015: 17). Selbst wenn die Matrix positiv bleibt, sind ihre statistischen Eigenschaften fragwürdig, da die verschiedenen Elemente der Korrelationsmatrix auf verschiedenen Stichproben basieren (Cheung/Chan 2005b: 42). Das zweite Problem besteht darin, dass für die SEM-Analyse in der zweiten Phase der MASEM- Analyse eine Stichprobengröße ausgewählt werden muss. Dazu stehen verschiedene Optionen (bspw. Median, Mittelwert, harmonischer Mittelwert) zur Verfügung. In manchen Teilen des Modells wird deshalb eine zu große und in anderen eine zu kleine Stichprobengröße zugrunde gelegt.

**Tabelle 18: Schwächen der herkömmlichen MASEM-Verfahren**

<b>univariater r/z-Ansatz</b>	<b>GLS-Ansatz</b>
Korrelationsmatrix wird als Datengrundlage für die SEM-Analyse verwendet (verzerrte Standardfehler und Chi-Quadrat-Statistik)	Schlechte Performanz bei niedrigen Fallzahlen innerhalb der Primärstudien
Homogenitätstest der bivariaten Korrelationen in der Korrelationsmatrix erfolgt getrennt voneinander (nur bivariate Beziehungen fehlerfrei)	Nur die Schätzung genau identifizierter Regressionsmodelle ist in der zweiten Phase korrekt möglich; unklar welche Stichprobengröße verwendet werden soll
Korrelationskoeffizienten der Korrelationsmatrix basieren auf einer unterschiedlichen Anzahl von Primärstudien (nicht positiv definite Matrix und unklar welche Stichprobengröße verwendet werden soll)	Stichprobenvariation wird ignoriert/keine Gewichtung in der 2. Analysephase
Stichprobenvariation wird ignoriert/keine Gewichtung in der 2. Analysephase	Keine Möglichkeit indirekte Effekte (Mediatoren) testen zu können und keine Fit-Indizes für die Modell-evaluation

**Anmerkungen:** Eigene Zusammenstellung.

Darüber hinaus gibt es im univariaten Ansatz das Problem, dass die Stichprobenvariation zwischen den Studien (Second-Order-Fehler) im univariaten r/z-Ansatz in der 2. Analysephase, also bei der Berechnung des Strukturgleichungsmodells, ignoriert wird (Landis 2013: 256). Während in der ersten Phase die Stichprobenvariation dadurch berücksichtigt wird, dass die

Studien entsprechend ihrer Präzision gewichtet werden, wird dies in der zweiten Phase des univariaten Ansatzes ignoriert. Im univariaten Ansatz ist eine Gewichtung in der zweiten Phase nicht möglich.

Der (ursprüngliche) GLS-Ansatz kann einige Schwächen des univariaten  $r/z$ -Ansatzes ausgleichen, hat jedoch eigene Probleme (Jak 2015: 19). Die Kritik bezieht sich vorwiegend auf die ursprüngliche Variante des GLS-Verfahrens und nicht auf die modifizierten Varianten, da diese eine ähnliche gute Performanz wie der TSSEM-Ansatz aufweisen (Furlow/Beretvas 2005: 251ff.; Zhang 2011: 54ff.). Allerdings unterliegen alle Varianten bei den Modellierungsmöglichkeiten weiteren Beschränkungen. Es können lediglich genau identifizierte Regressionsmodelle mit der sogenannten „direct computation method“ korrekt analysiert werden (Cheung 2015: 223; Zhang 2011: 24). Sollen jedoch Pfad- oder Faktormodelle mit diesem Ansatz geschätzt werden, kommt es zu ähnlichen Problemen wie beim univariaten Ansatz, da die kumulierte Korrelationsmatrix als beobachtete Matrix analysiert wird. Es bleibt zum einen unklar, welche Stichprobengröße bei der SEM-Analyse verwendet werden soll (Mittelwert, Median, harmonischer Mittelwert, etc.) und zum anderen werden potenzielle Unterschiede in der Präzision der Korrelationskoeffizienten, die aus den unterschiedlichen Stichprobengrößen folgen, nicht berücksichtigt (Jak 2015: 19). Dies hat zur Folge, dass eine Gewichtung der Korrelationsmatrix im ursprünglichen GLS-Ansatz somit nicht vorgesehen ist (Becker 2009: 393).

Eine Simulationsstudie, in der verschiedene multivariate MASEM-Verfahren verglichen wurden, zeigt, dass der herkömmliche Ansatz lediglich dann eine gute Leistungsfähigkeit aufweist, wenn die Fallzahl innerhalb der Primärstudien groß genug ist ( $>200$ ). Ansonsten liefert er beim Homogenitätstest verzerrte Ergebnisse (Zhang 2011: 57). In einer weiteren Simulationsstudie, bei der TSSEM und der ursprünglicher GLS-Ansatz miteinander verglichen wurden, hat sich ein deutlicher Performanznachteil des GLS-Verfahrens gezeigt (Cheung/Chan 2005b: 60f.). In der zweiten MASEM-Analysephase liefert der GLS-Ansatz erst ab sehr großen Stichproben ( $>1000$ ) pro Primärstudie zufriedenstellende Ergebnisse. Dies ist ein nicht unerheblicher Nachteil, da in vielen Wissenschaftsbereichen die Stichprobengrößen in empirischen Untersuchungen oftmals deutlich geringer sind. Darüber hinaus fehlen die Möglichkeiten indirekte Effekte (Mediatoren) testen und Fit-Indizes evaluieren zu können (Becker 2009: 393).

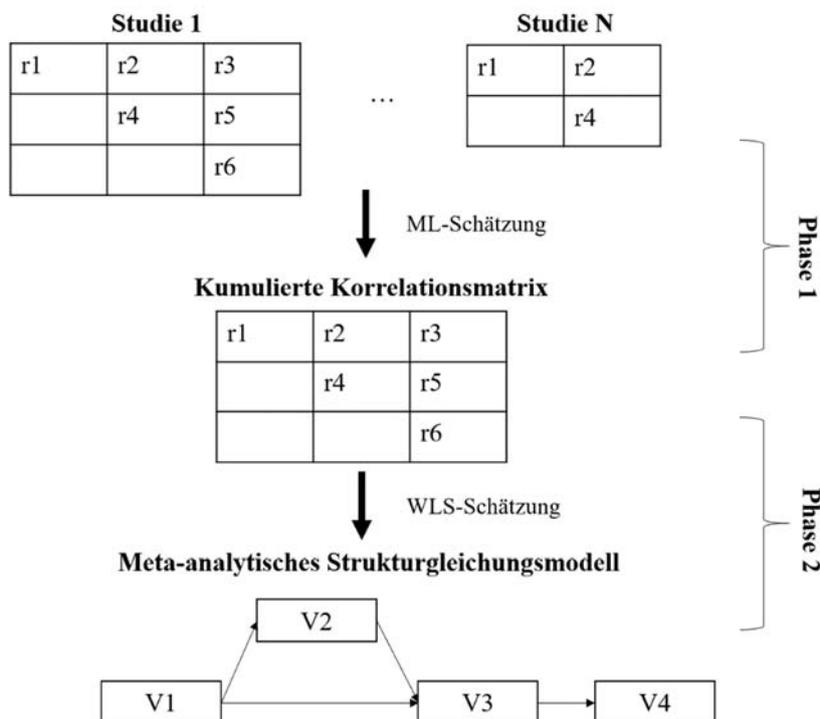
Bei ungleichen Fallzahlen innerhalb der Studien hat eine Simulationsstudie deutliche Performanzvorteile des multivariaten Ansatzes gegenüber den univariaten Ansätzen gezeigt (Alamohoda/Ayatollahi/Bagheri 2017). Bis heute ist der univariate  $r/z$ -Ansatz jedoch in der empirischen Praxis am weitesten verbreitet (Landis 2013: 254; Tang/Cheung 2016: 486). Die Ansätze

mit der besten empirischen Performanz sind der gewichtete GLS-Ansatz sowie das TSSEM-Verfahren (Jak 2015: 34). Darüber hinaus sprechen vor allem die Möglichkeiten, Mediatoren testen zu können und das Vorliegen von Fit-Indizes für die Verwendung des TSSEM-Ansatzes.

#### 5.4 Two-Stage-Structural-Equation-Modelling-Ansatz (TSSEM)

Der Two-Stage-Structural-Equation-Modelling-Ansatz (TSSEM) wurde von Cheung/Chan (2005b) entwickelt und von Cheung (2014a) erweitert. Er zeigt im Vergleich zu den zuvor genannten Ansätzen die beste Performanz und sollte deshalb zur Beantwortung modellbasierter meta-analytischer Fragestellungen verwendet werden (Cheung/Chan 2005b: 58ff.; Jak 2015: 22; Sheng et al. 2016: 198).<sup>79</sup> Die beiden zuvor beschriebenen MASEM-Ansätze nutzen meta-analytische Techniken zur Zusammenfassung der Korrelationsmatrizen (oder Kovarianzmatrizen)<sup>80</sup>, während beim TSSEM-Ansatz sowohl die Pooling-Phase (erste Phase) als auch die Fitting-Phase (zweite Phase) mithilfe von Strukturgleichungsmodellierung, genauer Multigruppen-SEM, durchgeführt wird. In Abbildung 12 ist eine Darstellung des Ablaufs des TSSEM-Verfahrens zu sehen.

**Abbildung 12: Schematische Darstellung TSSEM**



**Anmerkung:** Eigene Abbildung in Anlehnung an Jak/Cheung (2018a).

<sup>79</sup> Bei einem Literaturreview wurde allerdings festgestellt, dass von allen gefundenen MASEM Studien, nur 9 (5,6 % aller gefundenen Studien) den TSSEM-Ansatz zur Analyse verwenden (Sheng et al. 2016: 196). Dies unterstreicht den geringen Bekanntheitsgrad.

<sup>80</sup> Auch im TSSEM-Ansatz können nach Cheung/Chan (2009) Kovarianzen analysiert werden, wenn die gleichen Messinstrumente in allen Studien verwendet werden. Dann können auch Forschungsfragen beantwortet werden, die sich auf die Skalierung der Variablen beziehen (Cheung 2015: 224). Allerdings sind solche Studien aufgrund fehlender empirischer Informationen im sozialwissenschaftlichen Bereich selten durchführbar.

Prinzipiell kann eine TSSEM-Analyse bereits mit zwei vorliegenden Korrelationskoeffizienten durchgeführt werden. Allerdings erhöht sich die Genauigkeit der Schätzung entsprechend mit der Anzahl der Studien sowie deren Stichprobengröße (Yuan 2016: 219). Die mathematischen Modelle von GLS- und TSSEM-Ansatz sind ähnlich. Die Unterschiede liegen lediglich in der Berechnung der Parameterschätzungen und der Standardfehler (Cheung/Hafdahl 2016: 114).

In der **ersten Phase** wird die Korrelationsmatrix aus den Primärstudien als Gruppe in einem Multigruppen-Strukturgleichungsmodell zur Zusammenfassung der Korrelationskoeffizienten und der zugehörigen Kovarianzen genutzt (Cheung/Chan 2005b: 47). Unverzerrte Parameterschätzungen werden mithilfe des Maximum-Likelihood-Schätzers erreicht, wenn die fehlenden Korrelationen „missing completely at random“ (MCAR) oder „missing at random“ (MAR) sind. Das Verfahren kann somit mit fehlenden Korrelationen in den Primärstudien umgehen (Jak/Cheung 2017: 1ff.).<sup>81</sup> Zunächst wird getestet, ob die Korrelationen zwischen den Studien variieren (Homogenitätstest). Dabei wird mithilfe von Multigruppen-SEM ein Modell auf die Korrelationsmatrizen aller Studien angepasst, um zu testen, ob die Korrelationen über die Studien hinweg gleich sind (Jak 2015: 21). Zum formalen Homogenitätstest gibt es verschiedene Möglichkeiten. Es gibt eine multivariate Version des  $Q$ -Tests der auf dem GLS-Ansatz basiert, der allerdings wenig präzise ist, da er die Homogenitätsannahme überzufällig häufig zurückweist oder es gibt die Möglichkeit, für alle Korrelationskoeffizienten einzeln einen angepassten  $Q$ -Test durchzuführen (Jak 2015: 27). Sobald eine der Korrelationskoeffizienten signifikante Heterogenität aufweist, sollte die Homogenitätsannahme der Korrelationsmatrix zurückgewiesen werden. Das Ausmaß an Heterogenität kann mittels  $I^2$ -Wert ausgedrückt werden (Cheung 2015: 235). Dazu werden die bivariaten Elemente der Korrelationsmatrix betrachtet, um einen Eindruck über die Verteilung der Heterogenität zwischen den Korrelationen zu erhalten (Cheung 2014a: 32). Ist der  $I^2$ -Wert klein, liefern beide Modelle (FE und RE) ähnliche Ergebnisse.

Daneben ist auch die Homogenitätsüberprüfung im Fixed-Effects-Modell auch anhand der besseren Modellpassung auf Grundlage der aus der SEM-Analyse bekannten Fit-Indizes möglich. Dabei ist der, üblicherweise in der SEM-Analyse weitverbreitete (Urban/Mayerl 2014: 95f.), inkrementelle „Comperative-Fit-Index“ (CFI) für die Evaluierung der Homogenität zwischen den Korrelationskoeffizienten jedoch nicht geeignet, da dieser sehr stark von der Stärke der Korrelation selbst abhängt und somit bei der Evaluierung der Heterogenität in einer MASEM-

---

<sup>81</sup> Nicht alle Studien müssen alle Variablen beinhalten, sondern es genügt, wenn eine der relevanten Beziehungen berichtet wird (Quinn/Wagner 2018: 10). Wie beim GLS-Ansatz wird jedoch eine Selektionsmatrix (Designmatrix) benötigt, die darüber Auskunft gibt, welche Korrelationen in welcher Studie untersucht wurden.

Analyse keine Rolle spielen sollte (Jak 2015: 35). Besser geeignet sind die absoluten Fit-Indizes „Root-Mean-Square-Error-of-Approximation“ (RMSEA) und das „Standardized-Root-Mean-Squared-Residual“ (SRMR) zur Evaluierung der Homogenität der Korrelationsmatrizen (Cheung 2015: 227).<sup>82</sup> Größere Unterschiede zwischen den Korrelationen führen zu einem größeren SRMR-Wert, was eine höhere Heterogenität zwischen den Studien anzeigt. Wie in der herkömmlichen SEM-Analyse sollten generell immer mehrere Fit-Indizes berichtet werden und die Grenzwerte stellen lediglich einen groben Richtwert dar.

Ist die Modellpassung, in Übereinstimmung mit den analytischen Überlegungen, nicht signifikant unterschiedlich voneinander, kann Homogenität angenommen und das Erklärungsmodell in der zweiten Phase geschätzt werden (Landis 2013: 255). Wird die Homogenitätsannahme allerdings zurückgewiesen, dann darf das Untersuchungsmodell in der zweiten Phase nicht mit der kumulierten FE-Korrelationsmatrix getestet werden (Jak 2015: 20). Bei den bisherigen praktischen MASEM-Anwendungen hat sich jedoch gezeigt, dass bislang meist ein Fixed-Effects-Modell geschätzt wird, obwohl das Random-Effects-Modell meist analytisch überzeugender ist und die Homogenitätsannahme empirisch widerlegt wurde (Cheung/Sun/Chan 2018: 3).<sup>83</sup> Generell gibt es zwei Möglichkeiten mit Heterogenität in MASEM umzugehen. Ursprünglich musste nach Cheung/Chan (2005) die Heterogenität mittels Moderatoranalysen von studienspezifischen Variablen berücksichtigt werden, da ansonsten verzerrte Signifikanztests entstehen (Jak/Cheung 2018b: 2). Dazu können die Studien in relativ homogene Subgruppen eingeteilt und die Gruppenvariablen als Moderatoren zur Erklärung der Varianz genutzt werden.<sup>84</sup>

Seitdem der TSSEM-Ansatz allerdings um die Möglichkeit einer Random-Effects-Modellierung nach Cheung (2014b) erweitert wurde, ist die strikte Homogenitätsannahme nicht mehr notwendig. Das Random-Effects-Modell stellt die bessere Alternative dar und sollte im Hin-

---

<sup>82</sup> Der RMSEA basiert auf der Überlegung, dass Modelle nur Annäherungen an die Realität sind. Allerdings gibt es bislang keine Kennwerte aus Simulationsstudien, ab welchem RMSEA-Grenzwert das Ausmaß an Heterogenität zu groß für die Schätzung eines Fixed-Effects-Modells ist (Jak 2015: 35). Der RMSEA sollte bei einer herkömmlichen SEM-Analyse kleiner 0,05 bzw. mindestens kleiner als 0,08 sein (Byrne 2006: 94ff.). Das SRMR basiert hingegen auf der Differenz zwischen dem beobachteten und den im Modell geschätzten Korrelationskoeffizienten bzw. deren standardisierten Residuen und sollte bei einer herkömmlichen SEM-Analyse ebenfalls unter dem Wert von 0,08 liegen (Urban/Mayerl 2014: 90ff.).

<sup>83</sup> Der methodologische Praxisreview in dem hier relevanten Forschungsbereich hat ebenfalls gezeigt, dass das Random- oder Mixed-Effects-Modell in den meisten Forschungskontexten inzwischen das Standardmodell darstellt (siehe Kapitel 4.1.5).

<sup>84</sup> Zwei verschiedene Formen von Gruppenvariablen können verwendet werden. Zum einen ist es möglich a priori Variablen (kategoriale bzw. kategorisierte Studiencharakteristiken) und zum anderen ad hoc Variablen (mithilfe von Cluster-Analysen oder Mixture-Modelling) zu nutzen (Cheung/Chan 2005a: 973ff.).

blick auf das häufige Ziel einer MASEM, nämlich die Generalisierung wissenschaftlicher Erkenntnisse über die eingeschlossenen Studien hinaus, verwendet werden (Cheung 2015: 233). Die Heterogenität zwischen den Studien wird bei der Schätzung in der ersten Phase berücksichtigt, indem die Varianz und die Kovarianz der Korrelationskoeffizienten auf Studienebene mittels ML-Schätzung geschätzt wird (Jak 2015: 28f.). Die Heterogenität der Effektstärken wird dabei mittels Fehlerterm als „noise“ berücksichtigt.

Die Umsetzung ist einfach, führt jedoch, vor allem bei kleiner Studienanzahl und vielen Variablen, häufig zu Schwierigkeiten bei der Schätzung, da nicht ausreichend empirische Informationen vorhanden sind. In Folge dessen kann die Varianzkomponentenmatrix der Random-Effects fehlerhaft sein. Becker (2009) und Cheung (2014b) empfehlen beim Vorliegen des Problems nur die Varianz zwischen den Studien in die Gewichtungsmatrix (Diagonale der Kovarianzmatrix) an Stelle einer vollständigen symmetrischen Matrix einzufügen, damit nicht alle Elemente der Matrix geschätzt werden müssen (Cheung 2014b: 5). Wenn diese Matrix „0“ ist, entspricht dies genau dem Fixed-Effects-Modell.

In der **zweiten Phase** wird das Untersuchungsmodell auf Basis der kumulierten Korrelationsmatrix aus der ersten Phase überprüft. Dabei sind Modelle mit manifesten und latenten Variablen sowie Mediatoren möglich. Lediglich die Berücksichtigung von (kontinuierlichen und kategorialen) Moderatorvariablen im Modell ist in TSSEM nicht möglich (Cheung/Cheung 2016: 144; Sheng et al. 2016: 190). Dies hängt damit zusammen, dass die Korrelationen zwischen dem Interaktionsterm und den relevanten Modellvariablen der Primärstudien fehlen. Der Vektor aus der kumulierten Korrelationsmatrix und der zugehörigen asymptotischen Korrelationsmatrix, die in der ersten Phase geschätzt werden, sind der Stichprobenkorrelationsvektor in der zweiten Phase (Cheung 2015: 227f.). Als Gewichtungsmatrix dient dem TSSEM die inverse Matrix der asymptotischen Varianzen und Kovarianzen (ACOV) der kumulierten Korrelationsmatrix aus der ersten Phase. Dazu wird die sogenannte „asymptotically distribution free method“ (ADF), die auch „Weighted-Least-Square“ genannt wird, verwendet (Browne 1984: 62ff.; Urban/Mayerl 2014: 106).<sup>85</sup> Dies stellt sicher, dass diejenigen Korrelationskoeffizienten, die mit einer höheren Präzision geschätzt werden, weil sie auf mehr Studien basieren, ein größeres Gewicht bei der Schätzung des SEM-Modells haben (Cheung 2015: 228). Da nun in der

---

<sup>85</sup> Die ADF-Methode wird in LISREL „generally weighted least squares method“ und in dem R-Paket Mx „asymptotic weighted least square“ genannt. In *Mplus* und EQS ist diese Methode (bislang) nicht implementiert (Cheung/Chan 2005b: 47). Somit ist in *Mplus* und EQS bislang keine modellbasierte Meta-Analyse des Typs TSSEM durchführbar.

zweiten Phase bei der Bestimmung der Diskrepanzfunktion die Korrelationselemente der Matrix mit ihrer Präzision gewichtet werden, hat die Wahl der SEM-Stichprobengröße ohnehin keinen Einfluss auf die Chi-Quadrat-Statistik und die geschätzten Standardfehler und deshalb wird die totale Stichprobengröße verwendet (Cheung 2014a: 31). Damit wird beiläufig der Nachteil der ADF-Methode im Vergleich zum ML-Schätzer, die notwendigen hohen Fallzahlen (>1000) für eine unverzerrte Schätzung, ausgeglichen (Landis 2013: 255; Urban/Mayerl 2014: 107).

Es gibt zwei unterschiedliche Varianten, wie in der 2. Analysephase sichergestellt werden kann, dass die Diagonale der Modellkorrelationsmatrix „1“ ist. Dies ist Voraussetzung dafür, dass das Strukturgleichungsmodell überprüft werden kann (Cheung 2015: 230ff.). In der ersten Variante werden die Fehlervarianzen aus der Schätzung ausgeschlossen. Es werden alle Parameter bis auf die Fehlervarianzen geschätzt. Mit diesem Ansatz erhält man Standardfehler, aus denen herkömmliche Wald-based Konfidenzintervalle berechnet werden können. Dies kann für einfache Regressions- oder CFA-Modelle umgesetzt werden. Im zweiten Ansatz werden hingegen nicht-lineare Parameterfixierungen der Fehlervarianzen eingesetzt. Dies ist notwendig, sofern Pfadmodelle mit Mediatoren oder Beziehungen zwischen latenten Variablen im Modell berücksichtigt werden sollen. Dieser Ansatz liefert jedoch keine Standardfehler, da diese verzerrt sein könnten. Deshalb können auch keine Konfidenzintervalle bestimmt werden. Es ist dort erforderlich, Likelihood-based-Konfidenzintervalle (LBCIs)<sup>86</sup> zur Einschätzung der Präzision zu nutzen (Cheung 2017b: 12). Diese werden auch bei kleinen Fallzahlen oder im Random-Effects-Modell empfohlen. Sie basieren auf Likelihood-Funktionen anstatt auf dem Standardfehler und sind dadurch weniger verzerrt (Cheung 2009: 269ff.). Aufgrund der hohen Fallzahl in einer MASEM-Analyse sind die Signifikanzen der Effektstärken ohnehin nicht so bedeutsam und der Fokus sollte auf der Höhe der Effekte liegen (Sheng et al. 2016: 198).

Hält die Homogenitätsannahme zwischen den Korrelationsmatrizen in der ersten Phase nicht, muss in der zweiten Phase korrekterweise ein Random-Effects-Modell spezifiziert werden. Das Vorgehen im Random-Effects-Modell ist in der zweiten Phase identisch mit dem Vorgehen im Fixed-Effects-Modell. Der Unterschied zwischen den beiden Modellierungsformen liegt lediglich darin, dass die zweite Phase entweder auf einer kumulierten Fixed-Effects- oder einer kumulierten Random-Effects-Korrelationsmatrix beruht. Die Varianzkomponente der studienspezifischen Random-Effects spielt bei der Schätzung der Diskrepanzfunktion in der zweiten Phase keine Rolle (Cheung 2014a: 32). Sie wird lediglich in der ersten Phase für die Berücksichtigung

---

<sup>86</sup> Siehe dazu im Einzelnen die Ausführungen bei Neale/Miller (1997) und Cheung (2009).

der Heterogenität der Random-Effects bei der Schätzung der kumulierten Korrelationsmatrix verwendet (Cheung 2014b: 6; Tang/Cheung 2016: 478).<sup>87</sup> Wichtig ist, dass die richtige Gewichtungsmatrix, also diejenige, die mittels Random-Effects-Schätzung erstellt wurde, aus der ersten Phase genutzt wird.

Die Standardfehler der Parameterschätzungen sind im Random-Effects-Modell etwas größer als im Fixed-Effects-Modell und dementsprechend die Konfidenzintervalle breiter (Cheung 2015: 235). Zur Evaluierung der Modellpassung in der 2. Analysephase des TSSEM-Ansatzes werden Fit-Indizes verwendet. Die beiden inkrementellen Fit-Indizes „Comparative Fit Index“ (CFI) und „Tucker-Lewis-Index“ (TLI), die in der Strukturgleichungsmodellierung stark verbreitet sind (Urban/Mayerl 2014: 86ff.), sind für die Evaluierung meta-analytischer Modelle in der 2. TSSEM-Phase weniger geeignet (Cheung 2015: 232f.). Dies hängt vor allem damit zusammen, dass deren Logik einen Modellvergleich zu einem Baselinemodell (Nullmodell) voraussetzt, um eine Verbesserung des Modells anzeigen zu können. Bei Verwendung des WLS-Schätzers kommt es allerdings zu Verzerrungen, so dass für die Evaluierung der Modellpassung vorwiegend der RMSEA und das SRMR geeignet sind, da sie nicht von den Teststatistiken des Nullmodells abhängen. Das Ergebnis dieses Analyseschritts besteht aus den Parameterschätzungen mit Standardfehlern, zugehörigen Konfidenzintervallen, Chi-Quadrat-Teststatistiken und Fit-Indizes.

Das TSSEM-Verfahren nutzt also in der 2. MASEM-Analysephase eine WLS-Schätzung, um die Präzision der Koeffizienten beim Test des Strukturgleichungsmodells mittels Gewichtung zu berücksichtigen (Cheung 2015: 224ff.). Im TSSEM-Ansatz wird, anstatt der zur Analyse von Strukturgleichungsmodellen eigentlich notwendigen Kovarianzmatrix, in der Regel eine Korrelationsmatrix verwendet. Dies kann, wie bereits ausgeführt, zu verzerrten Standardfehlern der Parameterschätzungen und falschen Inferenzstatistiken führen. Der Unterschied zwischen den herkömmlichen MASEM-Verfahren und dem TSSEM-Verfahren besteht jedoch darin, dass bei der Schätzung ausdrücklich berücksichtigt wird, dass Korrelationen anstatt Kovarianzen bei der SEM-Analyse genutzt werden.

Im klassischen TSSEM-Ansatz liegen zwei Möglichkeiten zur Durchführung von Moderatoranalysen vor (Cheung/Hong 2017). Beim ersten Ansatz werden die Studien anhand bestimmter

---

<sup>87</sup> Diese Praxis kritisieren Cheung/Sun/Chan (2018) jedoch und weisen analytisch und mittels einer Simulationsstudie nach, dass es durch das Ignorieren der Heterogenität in der zweiten Phase zu Verzerrungen bei der Schätzung kommen kann, da die Kovariation der Random-Effects in den Populationsparametern zu einer Verzerrung der Pfadkoeffizientenschätzungen und zu einer nicht-normalen Verteilung der Random-Effects-Korrelationen führen kann. Allerdings zeigen sie auch, dass die Verzerrungen lediglich im Bereich der zweiten Nachkommastelle (.01 bis maximal .02) sind.

Studiencharakteristiken gruppiert (siehe auch Fußnote 84) und es werden getrennt voneinander unterschiedliche Strukturmodelle mithilfe eines Multigruppenvergleichs getestet (Urban/Mayerl 2014: 40ff.). Der Nachteil dieses Ansatzes besteht darin, dass kontinuierliche Variablen kategorisiert werden müssen, was mit einem Informationsverlust einhergeht und zu einer kleineren Anzahl von Studien in den Gruppen mit damit verbundenen Konvergenzproblemen führt (Jak/Cheung 2018b: 2ff.; Tang/Cheung 2016: 488). Dies sollte auf sehr wenige, a priori bedeutsame Moderatoren beschränkt bleiben. Es können auch verschiedene kategoriale Moderatoren bei der Subgruppenbildung kombiniert werden, allerdings muss berücksichtigt werden, dass die Studienanzahl je Subgruppe dadurch sehr gering werden und es besonders im Random-Effects-Modell zu Konvergenzproblemen kommen kann (Jak/Cheung 2018b: 13f.).

Die zweite Möglichkeit besteht darin, das Strukturgleichungsmodell auf jede einzelne Studie anzupassen. Die Parameterschätzungen werden dann als Effektstärken in einer Mixed-Effects-Meta-Analyse verwendet. Der Nachteil dieser Methode besteht jedoch darin, dass vollständige Korrelationsmatrizen ohne fehlende Werte benötigt werden. Deshalb ist diese Form der Moderatoranalyse meistens nicht nutzbar.

Zusammengefasst ist der Ablauf einer TSSEM-Analyse folgendermaßen: Zunächst wird eine kumulierte Korrelationsmatrix erstellt. In der zweiten Phase wird der WLS-Schätzer dazu genutzt, ein Strukturgleichungsmodell zu schätzen, das die Präzision der einzelnen Elemente der Korrelationsmatrix berücksichtigt. Die Vorteile des TSSEM gegenüber dem GLS-Ansatz bestehen also darin, dass neben der korrekten Analyse der kumulierten Korrelationsmatrix in SEM-Programmen, Modellvergleiche auf Basis des Chi-Quadrat-Tests möglich sind und Fit-Indizes zur Beurteilung der Güte der Schätzung vorliegen.

#### **5.4.1 Erweiterungen und Spezifika des TSSEM**

Nun werden sieben spezifische Aspekte bzw. Erweiterungen des TSSEM-Ansatzes diskutiert. Zunächst folgen Überlegungen zum Umgang mit fehlenden Werten, die besonders im Fixed-Effects-Modell problematisch sind. Anschließend werden verschiedene Möglichkeiten zur Quantifizierung und Erklärung von Heterogenität zwischen den Studien vorgestellt (Parameter-based-TSSEM, TS-FIMASEM, ML-MASEM und One-Stage-MASEM (OSMASEM)) Abschließend wird der Umgang mit Studienartefakten und das Problem abhängiger Effektstärken diskutiert.

## **(1) Fehlende Werte**

Vor dem Hintergrund, dass in einer MASEM-Analyse typischerweise kaum Studien das vollständige analytische Modell überprüft haben, ist unvollständiges Datenmaterial die Regel (Sheng et al. 2016: 194). Fehlende Werte sind in einer MASEM-Studie auf verschiedenen Ebenen möglich und vor allem im Fixed-Effects-Modell problematisch. Sie können auf Studienebene, Variablenebene und auf Koeffizientenebene vorliegen (Jak/Cheung 2017: 2). Auf Studienebene können Missings vorkommen, da nicht alle Studien, die durchgeführt worden sind, in die Meta-Analyse eingeschlossen werden (Publikationsbias). Fehlende Werte auf Variablenebene kommen zustande, da Variablen, die im meta-analytischen Modell relevant sind, in manchen Studien nicht berücksichtigt wurden (Furlow/Beretvas 2010: 129). Auf Koeffizientenebene kommen fehlende Werte vor, weil es sich etwa um eine Korrelation zwischen Variablen handelt, die nicht im Fokus der Primäruntersuchung stehen. Bei einem Random-Effects Modell sind fehlende Korrelationen kein Problem, da diese herausgefiltert werden. Bei einem Fixed-Effects-Modell hingegen muss für jede fehlende Korrelation eine der beiden damit verbundenen Variablen der Korrelation aus der Analyse ausgeschlossen werden (Jak 2015: 36). Dieser Ansatz wird auch „omitted-variables“(OV)-Ansatz genannt (Jak/Cheung 2017: 3f.).

Jak/Roorda/Koomen (2013) haben dazu eine Alternative, den sogenannten „omitted-correlation“-Ansatz (OC) eingeführt, der für jede fehlende Korrelation einen Parameter ersatzweise ins Modell aufnimmt.<sup>88</sup> Allerdings sind die gefundenen Unterschiede in der Simulationsstudie klein und die fixierten Werte in der Korrelationsmatrix können dazu führen, dass sich der Modellfit des Modells verschlechtert (Jak 2015: 36). Der Standardfehler des OV-Ansatzes steigt mit zunehmender Anzahl von fehlenden Werten (Jak/Cheung 2017: 11). Jak/Cheung (2017) kommen für das Fixed-Effects-Modell zu dem Schluss, dass bei vielen fehlenden Werten der OC-Ansatz verwendet werden sollte.

## **(2) Parameter-based-TSSEM**

Mit dem „parameter-based“-TSSEM, der die Parameterschätzungen bzw. die Funktion von Parameterschätzungen als Effektstärken betrachtet, kann die Heterogenität zwischen den Studien besser berücksichtigt werden. Die Kernannahme ist, dass das vorgeschlagene Strukturmodell

---

<sup>88</sup> Die dahinterliegende Idee besteht darin, dass die fehlenden Korrelationen in einer Studie auf einen beliebigen angemessenen Wert (etwa „0“) gesetzt werden, damit keine Korrelationsmatrix entsteht, die nicht positiv definit ist und die fixierten Werte die Ergebnisse der Analyse nicht beeinflussen können (Jak/Cheung 2017: 4; Jak/Roorda/Koomen 2013: 138f.). Für die Umsetzung wird eine weitere Matrix genutzt, in der „0“ für alle vorhandenen Korrelationen und ein frei zu schätzender Parameter für jede fehlende Korrelation in einer bestimmten Studie steht. Mit diesem Ansatz können, laut Simulationsergebnissen, Modelle mit einer größeren Anzahl von Freiheitsgraden und schmälere Konfidenzintervallen geschätzt werden.

sich in allen Studien bewährt, während die Parameter zwischen den Studien variieren können (Cheung 2015: 240ff.). Dieser Ansatz weist bedeutsame Vorteile auf. Es besteht etwa die Möglichkeit, die Heterogenität der Parameterschätzungen zwischen den Studien zu quantifizieren und kontinuierliche Prädiktoren bei der Parameterschätzung im Modell zu nutzen (Cheung/Cheung 2016: 144). Um diesen Ansatz verwenden zu können, müssen alle Studien alle Variablen des Untersuchungsmodells beinhalten (Cheung/Sun/Chan 2018: 3). Das Untersuchungsmodell darf darüber hinaus lediglich genau identifiziert sein und nur wenige Modelle sollten getestet werden, da bislang kein Modellvergleich möglich ist (Cheung/Cheung 2016: 153). Deshalb ist dieser Ansatz für die empirische Praxis bislang kaum zu gebrauchen.<sup>89</sup>

### **(3) TS-FIMASEM**

Yu et al. (2016) schlagen eine andere Erweiterung des TSSEM-Verfahrens vor. Dieser Ansatz wird TS-FIMASEM („Two-Stage-Full-Information-Meta-Analytic-Structural-Equation-Modelling“) genannt. Cheung (2017a) rechnet diesen Ansatz auch den parameter-based Verfahren zu. Diese Erweiterung ist im Gegensatz zum herkömmlichen TSSEM-Ansatz in der Lage, das Ausmaß an Heterogenität zu quantifizieren und somit auch die Stabilität der Schätzung zu beurteilen. TSSEM erkennt zwar, dass Heterogenität vorliegt, kann diese jedoch nicht quantifizieren (Cheung/Sun/Chan 2018: 5f.). In herkömmlichen Meta-Analysen kann dies auch durch Credibility-Intervalle (CVs)<sup>90</sup> berücksichtigt werden (Yu et al. 2016: 1457f.). Yu et al. (2016) schlagen vor, sowohl die Effektstärken als auch die zugehörige Effektstärkenheterogenität (Standardabweichung) in eigenen Matrizen zusammenzufassen. Im nächsten Schritt werden die beiden kumulierten Matrizen dazu genutzt, eine zufällige Stichprobe von Matrizen zu erstellen. Dabei bilden die Ergebnisse der 1. TSSEM Phase die Grundlage für ein Bootstrapping-Verfahren. Anschließend wird für jede per Bootstrapping erstellte Matrix ein Strukturgleichungsmodell geschätzt. Das Ergebnis davon ist eine Verteilung für jede Modellschätzung (Parameterschätzung und Fit-Indizes). Für die Parameterschätzungen schlagen Yu et al. (2016) ein Credibility-Intervall mit einem Grenzwert von 80 % zur Zusammenfassung der Verteilung vor.

---

<sup>89</sup> Bei einem empirischen Vergleich der beiden Ansätze bei Cheung/Cheung (2016) hat sich gezeigt, dass unter der Voraussetzung, dass ein Regressionsmodell getestet wird und vollständige Daten vorliegen, beide Ansätze ähnliche Ergebnisse liefern. Beim „correlation-based“-Ansatz wurden lediglich kleinere Standardfehler festgestellt.

<sup>90</sup> Ein „credibility intervall“ bezeichnet in einer Meta-Analyse, dass eine bestimmte prozentuale Anzahl (in der Regel 80 %) der Parameterschätzungen innerhalb einer bestimmten Verteilung von Parameterwerten liegt. Es wird mithilfe der Standardabweichung und nicht des Standardfehlers gebildet (Schmidt/Hunter 2015: 228f.). Credibility-Intervalle beziehen sich also vorwiegend auf die Generalisierbarkeit von meta-analytischen Schätzungen, während Konfidenzintervalle sich auf die Präzision und die statistische Signifikanz einer meta-analytischen Schätzung fokussieren (Yu et al. 2016: 1470).

Das Credibility-Intervall repräsentiert also den Bereich, indem mindestens 80 % der Populationsparameter fallen und gibt somit Auskunft über das Vorliegen von Moderatoren (Sheng et al. 2016: 190).

Darüber hinaus schlagen die Autoren vor, die Fit-Indizes auf Basis der per Bootstrapping erstellten Matrizen zu überprüfen (Yu et al. 2016: 1460). So kann quantifiziert werden, wieviel Prozent der Matrizen zu Strukturgleichungsmodellen mit angemessener Modellpassung führen. Breitere Credibility-Intervalle zeigen dabei stärkere Effekte von Moderatoren an. Die Verteilung der Fit-Indizes gibt einen Eindruck, unter welchen Voraussetzungen ein bestimmtes Modell stärker als andere Erklärungsmodelle von der Empirie unterstützt wird. Mit diesem Vorgehen lässt sich das empirisch passendste Modell von dem Modell mit der höchsten Generalisierbarkeit unterscheiden.

Diese Erweiterung des TSSEM-Ansatzes könnte die Generalisierbarkeit der modellbasierten meta-analytischen Ergebnisse erhöhen, da es möglich ist, den Populationsanteil zu finden, der von einer bestimmten Modellschätzung repräsentiert wird. Damit könnten einige Beschränkungen der bisherigen auf Nullhypothesen-Signifikanztests basierenden Praxis überwunden werden (Yu et al. 2016: 1470). Diese Modifikationen hat Cheung (2017a) mittels Simulationen überprüft. Darin zeigt er einige (technische) Schwächen und Fehler des Vorschlags von Yu et al. (2016) auf. Die Ergebnisse der Studie zeigen, dass der Ansatz die Erwartungen nicht erfüllen kann. Wenn Anwender jedoch an der Heterogenität der Parameterschätzungen interessiert sind, können die mittels Bootstrapping erstellten CVs aufschlussreich sein (Cheung 2017a: 28; Yu et al. 2018: 808).<sup>91</sup>

#### **(4) ML-MASEM**

Eine weitere Fortentwicklung des TSSEM-Verfahrens besteht darin, dass der ML-Schätzer in beiden Phasen der Analyse eingesetzt wird. Demzufolge wird in beiden Analysephasen eine ML-Multigruppenanalyse durchgeführt und die Modelle sind somit ineinander verschachtelt. Dieses Verfahren wird ML-MASEM genannt (Jak 2015: 36ff.). Im Gegensatz zum TSSEM-Verfahren besteht ML-MASEM aus insgesamt drei Schritten. Das 0-Modell aus dem Schritt 0 („saturated modell“) dient lediglich als Vergleichsmodell für den Homogenitätstest (Oort/Jak 2016: 159). Der erste Schritt in ML-MASEM ist identisch mit der ersten Phase im TSSEM-Ansatz. Der große Unterschied zum TSSEM ist somit im zweiten Schritt zu finden, da der TSSEM-Ansatz das spezifizierte Modell mit der geschätzten kumulierten Korrelationsmatrix

---

<sup>91</sup> Eine ähnliche Richtung schlägt der One-Stage Bayesian-Approach ein (Ke/Zhang/Tong 2018). Auch dort steht die Modellierung der Parameterheterogenität im Fokus.

testet, während das ML-MASEM Verfahren hingegen das Strukturgleichungsmodell mithilfe der beobachteten Korrelationsmatrizen innerhalb eines Multigruppenmodells testet (Oort/Jak 2016: 159). Der Vergleich von Modell 1 und Modell 0 mithilfe eines Chi-Quadrat-Differenzen-Tests stellt somit einen Test der Homogenitätsannahme dar. Der Vergleich von Modell 2 und 1 ist wiederum ein Test für die Passung des Strukturgleichungsmodells. Lediglich bei kleinen Stichprobengrößen liefert der ML-MASEM-Ansatz signifikant geringer verzerrte Parameterschätzungen als der ursprüngliche TSSEM-Ansatz unter Verwendung des WLS-Schätzers (Oort/Jak 2016: 165). Die Teststatistiken, Konfidenzintervalle, Falsch-Positiv-Raten und Richtig-Positiv-Raten unterscheiden sich zwischen den beiden Verfahren nicht. Die Vorteile dieser Modifikation ergeben sich durch die Möglichkeit ein Schätzverfahren in beiden Analysephasen einsetzen zu können.

### **(5) One-Stage-MASEM (OSMASEM)**

Der Nachteil der ML-Weiterentwicklung bestand darin, dass der Ansatz nur auf Grundlage von Fixed-Effects-Modellen angewendet werden konnte (Jak 2015: 37).<sup>92</sup> Allerdings wurde dieser Mangel beseitigt. In der neusten metaSEM-Version (1.2.0) ist eine experimentelle Versionen des One-Stage-MASEM (OSMASEM) enthalten, der es ermöglicht, kontinuierliche Moderatoren direkt ins Untersuchungsmodell einzubeziehen und die gesamte Schätzung in einem einzigen Modellierungsschritt vorzunehmen (Jak/Cheung 2018a). Dieser Ansatz stellt eine Verknüpfung und Erweiterung des ML-MASEM-Ansatzes und der ersten Phase des Random-Effects-TSSEM-Modells dar. Konzeptionell bedienen sich die Autoren an einem Ansatz aus der Messinvarianzforschung zur Erklärung von Heterogenität von geschätzten Parameter aus unterschiedlichen Studien (Bauer 2017). Der ursprüngliche Ansatz des TSSEM-Verfahrens wird dabei um eine Mittelwertstruktur erweitert (Jak/Cheung 2018a: 9). Mit diesem Ansatz ist es möglich, Moderatoruntersuchungen auf die Parameter des Strukturgleichungsmodells durchzuführen, indem eine Regression der Modellparameter auf die Moderatorvariablen erfolgt. Dabei kann eine Moderation aller direkten Effekte (Regressionskoeffizienten) und Varianzen sowie Kovarianzen überprüft werden. Dies ermöglicht die Untersuchung von Teilen der Heterogenität zwischen den Studien. Dieser Ansatz hat gegenüber dem FIMASEM-Ansatz (Yu et al. 2016) und dem parameter-based-Ansatz (Cheung/Cheung 2016) Vorteile, denn der FIMASEM-Ansatz kann die Heterogenität mittels Credibility-Intervallen und Bootstrapping quantifizieren und der parameter-based Ansatz benötigt vollständige Daten.

---

<sup>92</sup> Die einzige Möglichkeit im ML-MASEM-Ansatz mit Heterogenität umzugehen, besteht darin, Subgruppenanalysen mit homogenen Gruppen auf Basis von a priori Variablen (Studiencharakteristiken) im Sinne von Moderatoranalysen durchzuführen oder einen Clusteranalyseansatz zu wählen (Oort/Jak 2016: 165f.).

**Tabelle 19: Ansätze zum Umgang mit Heterogenität**

Ansatz	Moderatoren	Fehlende Werte
TSSEM	Multigruppenanalyse mit kategorialen Moderatoren	Ja
FIMASEM	Keine Moderatoren möglich	Ja
Parameter-based	Kategoriale und kontinuierliche Variablen	Nein
OSMASEM	Kategoriale und kontinuierliche Variablen	Ja

**Anmerkungen:** In Anlehnung an Jak/Cheung (2018a).

Durch Verwendung des FIML-Schätzers sind fehlende Werte unproblematisch. Jak/Cheung (2018a) empfehlen zur Verbesserung der numerischen Stabilität die Standardisierung (z-Transformation) der kontinuierlichen Moderatorvariablen (Urban/Mayerl 2018: 215). TSSEM mit Subgruppenanalyse und OSMASEM liefern nahezu die gleichen Parameterschätzungen ohne Berücksichtigung von Moderatoren. Wenn jedoch Moderatoren untersucht werden sollen, hat OSMASEM deutliche Vorteile, da kein Informationsverlust durch Kategorisierung der Variablen entsteht und somit die Teststärke zur Aufdeckung von Moderatoreffekten höher ist (Jak/Cheung 2018a: 16). Die Anzahl der Moderatoren sollte jedoch auf wenige zentrale Variablen beschränkt bleiben, da es auch hier zu Konvergenzproblemen kommen kann. In Tabelle 19 ist ein Vergleich der verschiedenen Ansätze zu sehen.

## (6) Studienartefakte

Auch in einer TSSEM-Analyse ist es möglich, den Einfluss von Studienartefakten, wie bei Schmidt/Hunter (2015), zu korrigieren (Cheung 2015: 243f.). Es ist jedoch unwahrscheinlich, dass Primärstudien die dafür notwendigen statistischen Informationen für alle Artefakte liefern.<sup>93</sup> Immerhin 75 % der von Sheng et al. (2016) in einem Literaturreview gefundenen MASEM-Studien haben solche Korrekturen vorgenommen (Sheng et al. 2016: 196). Bei der Reliabilitätskorrektur wird jeder Korrelationskoeffizient einzeln korrigiert und anschließend die korrigierte Korrelationsmatrix für die MASEM-Analyse genutzt (Cheung 2015: 244). Dies führt jedoch zu Problemen, da nicht alle Studien die Reliabilitätskoeffizienten berichten, so dass eine Korrektur nicht bei allen Korrelationen durchgeführt werden kann. Darüber hinaus kann die Prozedur zu nicht lösbaren Korrelationsmatrizen führen. Ebenso kann es zu Verzerrungen der konditionalen Stichprobenkovarianzmatrix kommen. Ferner konnte bereits nachgewiesen werden, dass Studienartefakte bzw. deren Korrektur generell keinen Einfluss auf die substanziellen Ergebnisse haben (Michel/Viswesvaran/Thomas 2011: 183ff.). Nach bisherigen

<sup>93</sup> Falls Rohdaten vorliegen, kann natürlich die SEM/CFA-Prozedur auch zur Messfehlerbereinigung verwendet und eine andere Form der Modellierung genutzt werden (Cheung 2014a: 38f.; Urban/Mayerl 2014: 16ff.).

Erkenntnisstand kann deshalb bei einer modellbasierten Meta-Analyse auf die Artefaktkorrektur verzichtet werden.<sup>94</sup>

### **(7) Abhängige Effektstärken**

Eines der größten ungelösten Probleme besteht derzeit im Umgang mit abhängigen Effektstärken, die die gleiche Variablenbeziehung untersuchen. Diese praxisrelevante Problematik wurde in TSSEM-Arbeiten bislang kaum berücksichtigt, obwohl mehrere Lösungsstrategien vorliegen (Cheung 2015: 180ff.; López-López et al. 2018 o. S.; Scammacca/Roberts/Stuebing 2014: 329ff.).<sup>95</sup> Die herkömmlichen Problemlösungen bestehen darin, dass entweder eine a priori festgelegte Entscheidungsregel verwendet oder der Mittelwert aus den entsprechenden Korrelationen gebildet wird. Darüber hinaus ist noch die zufällige Auswahl einer einzelnen Korrelation (López-López et al. 2018), der Shifting the Unit of Analysis Ansatz (Cooper 2010) oder die Verwendung einer robusten Varianzschätzung (Hedges/Tipton/Johnson 2010) möglich. Besonders relevant in der human- und sozialwissenschaftlichen Praxis ist die statistische Abhängigkeit, die aufgrund der Verwendung multipler Messinstrumente für dasselbe Konstrukt entsteht (Gleser/Olkin 2009). Rein praktisch betrachtet, liegen einer solchen Studie, mehrere für die kumulierte Korrelationsmatrix relevante Korrelationen für eine Beziehung vor, so dass diese in einer einzelnen Zelle der Korrelationsmatrix zusammengefasst werden müssen. Wilson/Polanin/Lipsey (2016) schlagen dazu eine Modifikation der 1. MASEM-Phase vor.<sup>96</sup> Die Autoren greifen dazu auf das Three-Level-Design von van den Noortgate et al. (2013) zurück, das verschiedene Formen der Abhängigkeit von Effektstärken (Cluster auf verschiedenen Ebenen) berücksichtigen kann. In der nachfolgenden Abbildung 13 ist die schematische Darstellung dieser Modellierung zu sehen.

Die kumulierte Korrelationsmatrix wird mittels eines gewichteten multivariaten Mehrebenen-Mixed-Effects-Meta-Regressionsmodells erzeugt (Wilson/Polanin/Lipsey 2016: 125). Die kumulierten Korrelationskoeffizienten werden mithilfe von Dummyvariablen, die die Zugehörigkeit zu einer spezifischen Zelle anzeigen, mittels eines No-Intercepts-Modells und mit einer Gewichtung durch die inverse der jeweiligen Stichprobengröße, geschätzt. Die durch diese

---

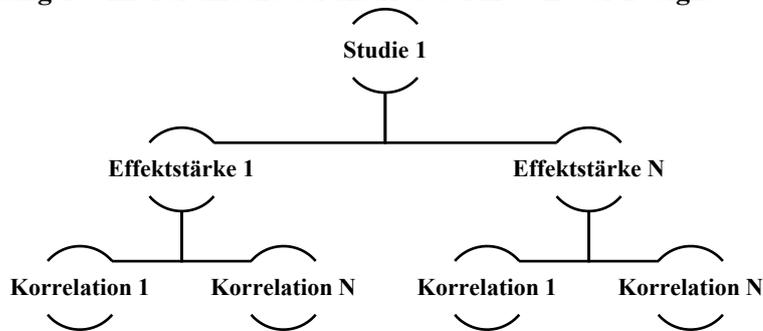
<sup>94</sup> Auch nach persönlicher Aussage von Suzanne Jak auf der „Research Synthesis 2018“ im Juli 2018 in Trier sollte man auf die Artefaktkorrektur verzichten, da bislang unklar ist, ob es beim TSSEM-Ansatz anschließend zu korrekten Ergebnissen führt.

<sup>95</sup> Auch bei der Betrachtung der meta-analytischen Praxis ist aufgefallen, dass viele Studien diese Problematik ignorieren (siehe Kapitel 4.1.5).

<sup>96</sup> Eine weitere Erweiterung die sie vorschlagen, besteht darin, dass sie Korrelationskoeffizienten bei der Zusammenfassung mittels Meta-Regression nach Studiencharakteristiken korrigieren, damit deren Einfluss auf die Effektstärken reduziert wird (Wilson/Polanin/Lipsey 2016: 126ff.). Damit entstehen allerdings künstliche Werte, die Schätzungen der Korrelationskoeffizienten darstellen. Dieser Ansatz stellt eine Alternative zur Moderatoranalyse dar, da die Effekte der Moderatoren auf die Effektstärken bereinigt anstatt analysiert werden.

Schätzung erzeugten Regressionskoeffizienten entsprechen den gewichteten kumulierten Korrelationskoeffizienten für einen bestimmten Zusammenhang. Die asymptotische Kovarianzmatrix, die über die Präzision der Schätzung informiert, wird bei dieser Schätzung ebenfalls berechnet und kann in der zweiten Phase der TSSEM-Schätzung genutzt werden.

**Abbildung 13: Hierarchische Struktur des Three-Level-Designs**



**Anmerkung:** Eigene Abbildung in Anlehnung an van den Noortgate et al. (2013).

Die Analyse in der zweiten Phase unterscheidet sich nicht vom herkömmlichen TSSEM-Ansatz und basiert auf der korrigierten kumulierten Korrelationsmatrix. Zusammenfassend erweitert der Ansatz das TSSEM-Verfahren um die Möglichkeit auch komplexe Datensätze mit mehreren Messinstrumenten für das gleiche Konstrukt zu analysieren.

Eine weitere Möglichkeit zur statistischen Modellierung der Abhängigkeit der Effektstärken, liegt in der robusten Varianzschätzung (RVE). Damit können verschiedene Arten von Abhängigkeiten der Effektstärken bei der Schätzung gleichzeitig berücksichtigt werden (Hedges/Tipton/Johnson 2010). Die Standardfehler werden dabei basierend auf den Abhängigkeiten korrigiert, so dass die Inferenzschlüsse korrekt sind. Die Anwendungsvoraussetzung liegt darin, dass mindestens 20 Studien vorliegen. Inzwischen liegt auch ein angepasster RVE-Schätzer vor, der auch in kleineren bis mittleren Stichproben robustere Ergebnisse liefert (Tipton 2015).

In Tabelle 20 ist eine Übersicht über die Möglichkeiten zum Umgang mit statistischer Abhängigkeit und deren jeweiligen Probleme zu sehen. Zum einen sind dies die sogenannten ad-hoc Lösungen und zum anderen die modelltechnischen Lösungen. Die ad-hoc Lösungen sind bei komplexen Fragestellungen, wie etwa bei der Untersuchung von Moderatorhypothesen auf Basis unterschiedlicher Messinstrumente, ungeeignet. Wenn jedoch keine expliziten Hypothesen bezüglich der Messinstrumente vorliegen, können diese Ansätze, mit Ausnahme der „Ignorieren“-Strategie, aufgrund ihrer Einfachheit angewandt werden. Sinnvoll ist dies allerdings nur, wenn alle Messinstrumente gleichermaßen zur Erfassung der jeweiligen Variable geeignet sind (López-López et al. 2018).

**Tabelle 20: Umgang mit statistischer Abhängigkeit der Korrelationskoeffizienten**

	Strategie	Auswirkungen/Probleme
Ad-Hoc Lösungen	Ignorieren	Verzerrte Standardfehler; falsche Inferenz
	Entscheidungsregel	Informationsverlust und unklar welche Regel
	Durchschnittsbildung	Informationsverlust
	Zufällige Auswahl	Informationsverlust
Modellierung	Shifting the Unit of Analysis	Informationsverlust; Datenmanagement sehr komplex
	Mehrebenenansatz	Unklar, ob die Interpretation analog zum TSSEM erfolgen kann; muss im R-Paket metafor durchgeführt werden
	Robuste Varianzschätzung	Muss in metafor oder anderen R-Paketen durchgeführt werden
	Constraint-Lösung	Nur einsetzbar, wenn alle Studien die gleichen Messinstrumente verwenden

Eine weitere, bislang nicht betrachtete Möglichkeit besteht darin, Constraints für die Gleichsetzung der Effekte zwischen den abhängigen Effektstärken und den anderen Modellvariablen zu nutzen. Dies wäre jedoch nur einsetzbar, wenn in allen Studien die gleichen Messinstrumente verwendet werden. Die Entscheidung, welche dieser Strategien angewandt werden soll, muss zunächst anhand von analytischen Überlegungen getroffen werden (López-López et al. 2018). Welche dieser Strategien empirisch die besten Ergebnisse liefert ist bislang ungeklärt und muss noch überprüft werden.<sup>97</sup>

#### 5.4.2 Desiderata Two-Stage-Structural-Equation-Modelling (TSSEM)

In diesem Abschnitt werden die methodischen Lücken des TSSEM-Ansatzes systematisiert und Lösungsansätze für unterschiedliche Adressaten formuliert, so dass weitere Forschungsvorhaben direkt daran anknüpfen können.

Da der TSSEM-Ansatz vorwiegend auf Basis von Korrelationen durchgeführt wird, wurde bislang der kovarianzbasierte TSSEM-Ansatz eher vernachlässigt, obwohl er dazu in der Lage ist, sehr differenzierte Analysen durchzuführen (Cheung/Hong 2017). Dies muss verstärkt in den Fokus genommen und entsprechende Evaluierungen durchgeführt, sowie konkrete Anwendungsgebiete gefunden werden.

Da die Umsetzung des Random-Effects-Modells im TSSEM-Ansatz noch neu ist, sollte die Performanz der unterschiedlichen Ansätze (parameter-based vs. correlation-based) mittels Monte-Carlo-Simulationsstudien näher evaluiert werden (Cheung 2015: 242). Besonderes Augenmerk sollte dabei auf die verschiedenen Vor- und Nachteile beider Verfahren gelegt und herausgearbeitet werden, unter welchen Voraussetzungen welcher Ansatz die besten Ergebnisse liefert. Die Leistungsfähigkeit des parameter-based-Ansatzes erscheint konzeptionell sehr viel-

<sup>97</sup> Hinweise, eine ausführlichere Diskussion der unterschiedlichen Strategien und deren Geeignetheit für jeweils unterschiedliche Forschungsziele sowie einen hilfreichen Entscheidungsbaum finden sich bei López-López et al. (2018).

versprechend, allerdings sind seine Anwendungsvoraussetzungen für die empirische Praxis bislang ungeeignet. Zukünftige Simulationsstudien, die das Random-Effects-Modell überprüfen, sollten realistischere Simulationsbedingungen (fehlende Variablen, unterschiedliche Anteile an Missings, unterschiedliche Fallzahlen etc.) als in der Vergangenheit verwenden und deren Auswirkungen auf beide Analysephasen evaluieren (Alamolhoda/Ayatollahi/Bagheri 2017: 456).

Eine weitere Lücke zeigt sich bei den unterschiedlichen Möglichkeiten den Modellfit zu evaluieren. Es muss künftig weitere Forschung erfolgen, um einen einheitlichen Weg zur Modellevaluation zu finden (Cheung 2014a: 38). Es werden (weitere) Simulationsstudien benötigt, die die empirische Performanz der unterschiedlichen Fit-Indizes miteinander vergleichen (Cheung/Chan 2005b: 61). Es bleibt trotz vorliegender Simulationsergebnisse unklar, welche Fit-Indizes (CFI, TLI, RMSEA, SRMR) mit welchen Grenzwerten verwendet werden sollten (Cheung 2015: 233). Reichen etwa die üblichen Grenzwerte („Daumenregeln“) aus der SEM-Literatur aus (Urban/Mayerl 2014: 99) oder werden aufgrund des meta-analytischen Designs und der damit einhergehenden Unschärfe weichere Grenzen benötigt? Bislang wird überwiegend der RMSEA und das SRMR zur Evaluation der Modellpassung empfohlen, allerdings muss auch hier die empirische Performanz näher untersucht und auf Basis von Simulationsstudien Empfehlungen für Grenzwerte gefunden werden (Jak 2015: 35). Beispielsweise bleibt unklar, ab welchem Grenzwert des RMSEA die Homogenitätsannahme in der ersten Phase verworfen werden soll.

Der Vorschlag von Wilson/Polanin/Lipsey (2016) zur Kombination von Three-Level-Meta-Analyse und TSSEM sollte mittels Simulationen einerseits und praktischer Anwendung andererseits evaluiert werden. Auch die anderen aufgeführten Strategien zum Umgang mit Abhängigkeiten zwischen den Effektstärken (siehe Tabelle 20) müssen betrachtet werden. Ein erster Vergleich der unterschiedlichen Ansätze liegt bereits vor und zeigt einen leichten Vorteil des robusten Varianzschätzers gegenüber dem 3-Ebenen-Modell (Park/Beretvas 2018). Allerdings ist diese Studie nicht im Kontext modellbasierter Meta-Analyse durchgeführt worden.

Wie von Schmidt/Hunter (2015) vorgeschlagen, können auch im TSSEM-Ansatz Studienartefakte korrigiert werden. Dazu liegen jedoch bislang keine empirischen Untersuchungen vor, die überprüfen, wie es um die Performanz dieser Artefaktkorrektur in MASEM beschaffen ist (Tang/Cheung 2016: 487). Vor allem die Auswirkungen der Reliabilitätskorrektur sollte untersucht werden (Cheung 2014a: 39, 2015: 244). Diese Studien werden dringend gebraucht, da diese Verfahren, obwohl ihre Leistungsfähigkeit unbekannt ist, bereits eingesetzt werden

(Sheng et al. 2016: 196). Auch die Verwendung von z-Werten anstelle von Korrelationskoeffizienten muss weiter untersucht werden. Es ist unklar, wie sich die Performanz von herkömmlichen Korrelationsmatrizen zu Korrelationsmatrizen auf Basis von z-Werten verhält (Cheung 2015: 243) und inwiefern die technischen Schwierigkeiten bei der Verwendung von z-Werten bei der Analyse von Strukturgleichungsmodellen überwunden werden sollen (Cheung/Hong 2017: o. S.). Besonderes Augenmerk sollte dabei auf die Auswirkungen der z-transformierten Werte auf die Parameterschätzungen in der zweiten Phase gelegt werden (Cho 2015: 61f.).

In der Zukunft sollte überprüft werden, inwiefern der MASEM-Ansatz mittels Bayes-Statistik erweitert werden kann (Cheung 2014a: 38). Erste Vorschläge liegen dazu von Prevost et al. (2007) und Ke/Zhang/Tong (2018) vor. Gebraucht werden systematische Vergleiche der verschiedenen modellbasierten Ansätze. Dazu müssen die erwähnten Ansätze FIMASEM, OSMA-SEM und One-Stage-Bayesian-Approach mit den herkömmlichen MASEM-Verfahren systematisch mithilfe von Monte-Carlo-Simulationsstudien verglichen werden.

Ein weiterer Aspekt, der verstärkt untersucht werden sollte, sind die Auswirkungen von fehlenden Werten auf die Ergebnisse von MASEM-Studien und welche Verfahren zum Umgang mit dieser Problematik angewandt werden können. So stellt sich etwa die Frage, wie sich die verschiedenen Missing-Mechanismen auf die Schätzungen in der 2. MASEM-Phase auswirken (Zhang 2011: 58). Besonders relevant ist dabei, welche Auswirkungen es hat, wenn die fehlenden Korrelationskoeffizienten „missing not at random“ (MNAR) sind (Cheung/Chan 2005b: 60; Jak/Cheung 2017: 11). Es hat sich bereits gezeigt, dass der vorgeschlagene OC-Ansatz (Jak/Roorda/Koomen 2013: 139) besser abschneidet als der herkömmliche OV-Ansatz (Jak/Cheung 2017: 11f.). Allerdings sollten die Ergebnisse dieser Simulation mittels Variation der Simulationsbedingungen einer weiteren Überprüfung unterzogen werden. Ebenfalls steht die Performanzüberprüfung von multiplen Imputationsverfahren (MI) mit meta-analytischen Daten aus. Bislang sind konzeptuelle Zweifel angebracht, ob sie überhaupt zu korrekten Ergebnissen im meta-analytischen Kontext führen (Furlow/Beretvas 2005: 251). Lediglich einzelne Simulationsstudien wurden mit einer beschränkten Anzahl von Simulationsbedingungen (uni-variater Ansatz, GLS-Ansatz, Fallzahl und Anzahl der Studien, Fixed-Effects-Modell) durchgeführt (Furlow/Beretvas 2010: 139ff.). Die Ergebnisse zeigen, dass MI zur Ersetzung fehlender Werte in einer MASEM-Analyse unter dem MCAR-Mechanismus gut funktioniert. Diese Ergebnisse müssen mittels weiterer Monte-Carlo-Simulationsstudien differenzierter und realitätsnaher (MAR-Missing-Mechanismus, Random-Effects-Modell und mehr Variation bei den Simulationsbedingungen) überprüft werden.

Viele statistische Verfahren basieren auf der Annahme, dass die Daten annäherungsweise normalverteilt und statistisch unabhängig voneinander sind. Dies gilt auch für die modellbasierten meta-analytischen Verfahren. Aus der SEM-Literatur sind mittlerweile jedoch verschiedene robuste Schätzer, die auch bei Verletzung dieser Annahme verwendet werden können, bekannt (Urban/Mayerl 2014: 66ff.). Diese Verfahren benötigen jedoch häufig Rohdaten, die in einer Meta-Analyse in der Regel nicht vorliegen. Somit bleibt ungeklärt, ob und wie solche Verfahren mathematisch in den TSSEM-Ansatz eingebaut werden können (Cheung 2014a: 37). Ein erster Vorschlag für robuste Standardfehler bei Verletzungen der Normalverteilungsannahme im Fixed-Effects GLS-Ansatz wird von Yuan (2016) eingeführt. Diese Anregung muss nun noch auf das TSSEM-Modell übertragen und seine empirische Äquivalenz mit dem ML-Ansatz überprüft werden (Yuan 2016: 229).<sup>98</sup>

Einige Autoren haben bereits TSSEM-Analysen auf Basis von kumulierten Korrelationsmatrizen anderer Meta-Studien durchgeführt (Sheng et al. 2016: 195). Es muss überprüft werden, ob dieser Ansatz, etwa aufgrund von unterschiedlichen Selektionskriterien, in den einzelnen Meta-Studien zu korrekten Ergebnissen gelangt (Tang/Cheung 2016: 487).

Wie sich bei der Synthese der Forschungslücken zeigt, sind weitere Monte-Carlo-Simulationsstudien, die die relevanten Entscheidungen bei der Durchführung einer modellbasierten Meta-Analyse empirisch näher evaluieren, notwendig. Dazu zählen etwa die Performanz der Schätzer, die Verwendung von z-Werten und die Evaluation der Fit-Indizes sowie ein systematischer Vergleich zwischen den TSSEM-Erweiterungen bzw. dessen Alternativen. Darüber hinaus müssen die Ansätze für die Entwicklung robuster meta-analytischer Schätzer und deren Einbindung in die vorliegenden Softwarelösungen weiterverfolgt werden. Generell müssen die Simulationen unter realistischeren Bedingungen durchgeführt werden (Cheung 2017a: 29). Das bedeutet, dass Simulationskonfigurationen gewählt werden, die den praktischen Gegebenheiten der meta-analytischen Forschung entsprechen. Bei bisherigen Simulationen wurde oftmals die Fallzahl in den Primärstudien nicht variiert und lediglich ein Fixed-Effects-Modell angenommen. Diese Aspekte sollten in neuen Studien als weitere Konfigurationen in das Simulationsstudiendesign aufgenommen werden (Cho 2015: 60ff.; Furlow/Beretvas 2005: 252). Daneben muss auch eine größere Variabilität in Bezug auf die Anzahl der Studien und/oder Variablen, die Komplexität der Modelle und das Muster fehlender Werte in Simulationen berücksichtigt werden (Oort/Jak 2016: 166; Zhang 2011: 57). Vor der Durchführung weiterer Simulationen

---

<sup>98</sup> Darüber hinaus gibt es bei Verletzungen der Unabhängigkeitsannahme, wie bereits angesprochen, auch die Möglichkeit die Standardfehler mit dem Verfahren von Hedges/Tipton/Johnson (2010) zu korrigieren. Eine feste Einbindung in das metaSEM-Paket ist wünschenswert.

sollte deshalb mittels eines systematischen Literaturreviews, die typischen Anwendungen der modellbasierte Meta-Analyse erhoben werden, um im nächsten Schritt auf Basis dieser Informationen die relevanten Simulationskonfigurationen abzuleiten. Neben den genannten Simulationsarbeiten werden jedoch auch weitere empirische Analysen von Realdaten benötigt, die die unterschiedlichen Methoden anwenden und deren Ergebnisse miteinander vergleichen.

## **5.5 Publikationsbias und Sensitivitätsanalyse**

Nun folgen Ausführungen zu zwei spezifischen Aspekten einer Meta-Analyse, die für modellbasierte Verfahren bislang noch kaum diskutiert wurden. Zunächst wird der sogenannte Publikationsbias und die entsprechenden Diagnosemöglichkeiten eingeführt. Daran anknüpfend werden verschiedene Vorschläge für Sensitivitätsanalysen systematisiert.

### **5.5.1 Publikationsbias**

Eine wichtige Annahme für die Gültigkeit meta-analytischer Ergebnisse ist, neben der Unabhängigkeit der Beobachtungseinheiten, eine unverzerrte Datenbasis. Das potenzielle Vorliegen des so genannten „publication bias“, also die systematische Form der Selektivität von bestimmten Primärstudien, gefährdet allerdings die Ergebnisvalidität (Urban/Fiebig 2015: 161).<sup>99</sup> Jede Meta-Analyse sollte testen, ob ein Publikationsbias vorliegt (Sutton 2005: 175ff.). Dies ist notwendig, da aufgrund des exponentiellen Zuwachses an (wissenschaftlicher) Information kaum die Möglichkeit besteht, alle relevanten Studien in einer Meta-Studie zu berücksichtigen. Dabei sollten stets mehrere Verfahren angewandt werden. Auch die modellbasierte Meta-Analyse ist gegen solche Verzerrungen nicht geschützt.<sup>100</sup>

Dieser Bias kann verschiedene Ursachen haben (Thornton/Lee 2000: 207ff.). Zum einen liegt dies am so genannten „file-drawer-problem“, also dem Phänomen, dass Studien mit signifikanten Ergebnissen häufiger publiziert werden (Hox 2010: 218).<sup>101</sup> Zum anderen kann er aus der Tatsache resultieren, dass bestimmte Veröffentlichungsformen (graue Literatur) über die systematische Literatursuche schwerer zugänglich sind (bspw. Konferenzpapiere, Hochschulveröffentlichungen, unveröffentlichte Forschungsberichte etc.). Wird dies ignoriert, kann es zu

---

<sup>99</sup> Das Vorliegen eines Publikationsbias ist im Little-Rubin-Framework zur Klassifizierung fehlender Werte ein NI/NMAR (non-ignorable/Not-Missing-at-Random), weil die Wahrscheinlichkeit, dass Studienergebnisse berichtet werden, von dem Ergebnis selbst abhängen (bspw. von der statistischen Signifikanz).

<sup>100</sup> Allerdings kann argumentiert werden, dass das Problem dort geringer ist, da ein ganzes Netz von Variablen in die Analyse eingeschlossen wird, wobei teilweise sogar Variablen eingeschlossen werden, nach denen nicht explizit gesucht wird.

<sup>101</sup> Ausführlichere Erläuterungen, über die Entstehung und die Probleme, die daraus resultieren (Sutton 2009: 435ff.; Urban/Fiebig 2015: 161) sowie ausführliche Beschreibungen zu den Überprüfungsverfahren (Card 2012: 260ff.; Urban/Fiebig 2015: 160ff.) finden sich an anderer Stelle.

Verzerrungen der meta-analytischen Ergebnisse führen (Hedges 2009: 44; Rothstein/Sutton/Borenstein 2005: 2). Die beste Strategie den Publikationsbias zu vermeiden, liegt deshalb in der sorgfältigen Planung der systematischen Literatursuche basierend auf unterschiedlichen Quellen, damit insbesondere auch graue Literatur gefunden wird (Hopewell/Clarke/Mallett 2005: 49ff.; Sutton 2009: 436).

Neben der Möglichkeit den Publikationsbias bereits bei der Planung der systematischen Literatursuche zu vermeiden, gibt es darüber hinaus unterschiedliche Analysemöglichkeiten die Ergebnisse von Meta-Analysen auf das Vorliegen von Verzerrungen zu untersuchen (Fiebig/Rüdebusch/Urban 2012). Eine Überprüfung, ob ein Publikationsbias besteht, ist obligatorisch und sollte möglichst mit mehreren der nachfolgenden Verfahren evaluiert werden (Card 2012: 260). Diese Verfahren können in graphische und statistische Analysemethoden unterschieden werden. Das Problem dieser Verfahren ist, dass sie nicht für multivariate Modelle vorliegen, so dass es für MASEM-Analysen üblich ist, die entsprechenden Verfahren jeweils auf die bivariaten Beziehungen anzuwenden (Scherer/Siddiq/Tondeur 2019: 20).

Eine der einfachsten Varianten, die in der empirischen Praxis häufig angewendet wird, ist die Überprüfung mittels Moderatoranalysen. Mit einem dichotomen Moderator „Veröffentlichungsstatus“ kann überprüft werden, ob die Effektstärke in Abhängigkeit davon variiert. Allerdings braucht dieses Vorgehen eine entsprechende Anzahl nicht publizierter Studien, um zu validen Aussagen zu gelangen.

Zur graphischen Analyse können Funnel-Plots explorativ ausgewertet werden (Sterne/Becker/Egger 2005: 75ff.). Hierzu werden die standardisierten Effektstärken (z-transformierte Werte) in einem Koordinatensystem gegenüber der Stichprobengröße, der inversen Stichprobenvarianz oder dem Standardfehler mithilfe von Scatterplots in Beziehung gesetzt (Hox 2010: 219). Die Form sollte symmetrisch sein und etwa wie ein Trichter aussehen, falls kein Publikationsbias vorliegt. Die Idee dahinter ist, dass die Effektstärken gleich verteilt, um den dahinterliegenden „wahren“ Effekt streuen und das bei kleineren Studien (niedrige Fallzahl) eine größere Variabilität vorliegt als bei großen Studien. Allerdings muss beachtet werden, dass „funnel-plots“ als subjektive Hilfsmittel mit Vorsicht zu interpretieren und keinesfalls allein dazu geeignet sind, statistisch valide Aussagen über das Vorliegen eines Publikationsbias zu treffen (Sterne/Becker/Egger 2005: 75).<sup>102</sup>

---

<sup>102</sup> Auch andere Ursachen können für asymmetrische Verläufe verantwortlich sein und besonders bei einer kleinen Anzahl von Primärstudien können keine validen Schlüsse gezogen werden (Card 2012: 264ff.).

Darüber hinaus sind formale statistische Tests möglich (Sutton 2009). Diese Tests wenden die Logik der Funnel-Plots an, indem sie Effektstärke und Fallzahl miteinander in Verbindung bringen. Sie überprüfen, ob etwa kleine Studien einen systematischen Effekt auf das Gesamtergebnis der Meta-Analyse haben (Sterne/Egger 2005: 107). Bei einer positiven mittleren Effektstärke liegt ein Publikationsbias vor, wenn Studien mit kleiner Fallzahl größere Effektstärkeschätzungen als Studien mit größerer Fallzahl liefern (Card 2012: 266). Es kann der Rangkorrelationskoeffizient (auch „Begg and Mazumdar’s rank correlation method“ genannt) und verschiedene Varianten (bspw. mit robusten Standardfehlern) des sogenannten Egger-Tests berechnet werden (Sterne/Egger 2005: 100ff.; Sutton 2009: 440ff.). Die Rangkorrelationsmethode überprüft mittels Paarvergleich, ob es einen Zusammenhang von Effektstärke und Standardfehler gibt. Liegt ein Publikationsbias vor, zeigt sich dieser, indem hohe Standardfehler (kleine Studien mit kleiner Stichprobe) mit größeren Effektstärken zusammenhängen. Der Egger-Test hingegen nutzt eine lineare Regression, um den Effekt des Standardfehlers (der Präzision) auf die Effektstärke zu bestimmen. Beide Methoden sollten Simulationsergebnissen zufolge nicht angewandt werden, wenn weniger als 17 Studien (Egger-Test) bzw. weniger als 40 Studien (Rangkorrelation) vorliegen, da ansonsten die Teststärke zu gering ist (Card 2012: 268; Sterne/Egger 2005: 105). Darüber hinaus hat Egger-Test eine höhere Teststärke als die Rangkorrelationsmethode. Die Rangkorrelationsmethode ist nur sensitiv gegenüber einem ausgeprägten Publikationsbias.

Eine illustrative Methode zur Überprüfung, eines Publikationsbias, stellt die fail-safe-N-Methode dar (Rosenthal 2010). Dieser Methode liegt die Frage zugrunde, wie viele nicht signifikante Publikationen zusätzlich in eine Meta-Studie eingeschlossen werden müssten, um den Gesamteffekt der Meta-Analyse nicht signifikant werden zu lassen (Hox 2010: 219). Dabei handelt es sich um eine Methode, die die Robustheit der meta-analytischen Ergebnisse gegenüber der Existenz nicht berücksichtigter Studien überprüft (Card 2012: 268ff.; Urban/Fiebig 2015: 168f.). Zu dieser Methode gibt es verschiedene Abwandlungen, wie etwa die Rosenberg- oder die Orwin-Variante. Die Besonderheit der Rosenberg-Variante liegt in der Möglichkeit der Gewichtung bei der Berechnung der fail-safe-Ns. Das Grundprinzip zur Evaluation ist bei allen Varianten einfach. Bei einem großen berechneten Wert ist die Wahrscheinlichkeit gering, dass ein Bias vorliegt.

Eine weitere Möglichkeit zur Überprüfung, ob ein Publikationsbias vorliegt, stellt die nichtparametrische „trim and fill“-Methode dar (Duval 2005: 127ff.; Sutton 2009: 443f.). Bei diesem Verfahren geht es, ähnlich wie bei den Funnel-Plots, um die Symmetrie der Werteverteilung. Liegt keine symmetrische Verteilung vor, wird mit einer iterativen zweiphasigen Korrektur die

Anzahl der fehlenden Studien ermittelt, die notwendig ist, um Symmetrie herzustellen (Urban/Fiebig 2015: 169f.). Das Ergebnis einer solchen Berechnung ist eine aktualisierte präzisere Effektstärke, die auch Studien beinhaltet, die nicht vorliegen. Dies sollte jedoch nicht als eine Art von Imputation fehlender Werte (hier: Studien) verstanden werden. Viel mehr wird die Robustheit der Ergebnisse im Sinne einer Sensitivitätsanalyse überprüft.

Für modellbasierte Meta-Analysen liegen keine spezifischen Vorschläge vor. Bislang können lediglich die herkömmlichen Verfahren für bivariate Beziehungen angewandt werden. Die empirische Praxis im MASEM-Bereich klammert das Problem des Publikationsbias bislang fast vollständig aus.<sup>103</sup> Deshalb ist es notwendig, neue Vorgehensweisen zu entwickeln, um sicherstellen zu können, dass auch in einem multivariaten meta-analytischen Modell kein Publikationsbias vorliegt. Ebenfalls muss ein Vorgehen für den Fall, dass nur in Teilen des Gesamtmodells eine verzerrte Studienbasis vorliegt, entwickelt werden. Für die vorliegende Arbeit werden, neben der sorgfältigen Planung der Literatursuche, eine Kombination aus fünf Verfahren angewandt. Zunächst erfolgt eine Kombination aus graphischer Analyse der Funnel-Plots und dem Trim-and-Fill-Verfahren auf Basis der z-transformierten Korrelationskoeffizienten in Relation zum Standardfehler. Sowohl der Rangkorrelationskoeffizient (Kendalls Tau), als auch der Egger-Test werden für die einzelnen Beziehungen berechnet. Der Egger-Test wird in Form einer gewichteten Regression der Effektstärke auf den Standardfehler zur statistischen Überprüfung potenzieller Funnel-Plot-Asymmetrie durchgeführt. Der eher illustrierte Rosenbergs fail-safe-N-Ansatz bildet den Abschluss dieser Überprüfung.

### 5.5.2 Sensitivitätsanalyse

Sensitivitätsanalysen sind notwendig, um die Robustheit meta-analytischer Ergebnisse im Hinblick auf die statistischen Annahmen und getroffenen Entscheidungen sicherzustellen (Borenstein et al. 2011: 368ff.). Meta-analytisch generierte Schlüsse sind nur zulässig, wenn sie unter der Voraussetzung einer unverzerrten Datenbasis entstanden sind. Vor allem im Hinblick auf das hohe Vertrauen, das meta-analytischen Ergebnissen aus der Praxis entgegen gebracht wird, sind sie somit eigentlich obligatorisch (Bosnjak/Viechtbauer 2009: 329; Greenhouse/Iyengar 2009: 418). Auch die Analyse von Ausreißerstudien bzw. von besonders einflussreichen Studien (bspw. sehr kleine oder sehr große Effektstärken) kann zu einem genaueren Bild führen (Viechtbauer/Cheung 2010: 124). Alle Phasen der Meta-Analyse, von der Formulierung der

---

<sup>103</sup> Dies ergibt zumindest eine Durchsicht einiger aktueller MASEM-Studien (Bamberg/Möser 2007; Bergh et al. 2014; Boer et al. 2016; Curtiss et al. 2017; Eller 2015; Hildebrand 2015; Hong/Cheung 2015; Kraft/Bausch 2016; Luhmann et al. 2012; Montazemi/Qahri-Saremi 2015; Niemand 2014; Scalco et al. 2017; Schlaegel/Koenig 2014; Schudey/Jensen/Kraicz 2013; Steinmetz/Isidor/Baeuerle 2012).

Forschungsfrage bis hin zur Präsentation der Ergebnisse, und somit alle Entscheidungen, die in einer Meta-Analyse getroffen werden müssen, sind für Sensitivitätsanalysen geeignet.

Häufig angewendet werden Variationen bei der Datenanalyse, wie etwa der Vergleich der meta-analytischen Ergebnisse, wenn unterschiedliche Selektionskriterien verwendet werden (Borenstein et al. 2011: 368f.). Dies ist jedoch mit hohem Aufwand verbunden, der bereits beim Screening der Studien bedacht werden muss. Ebenso ist es möglich, die Ergebnisse im Hinblick auf die Verwendung von unterschiedlichen Effektstärkekoeffizienten miteinander zu vergleichen. Eine weitere Möglichkeit der Sensitivitätsanalyse besteht darin, dass Studienergebnisse, die keine oder nur wenige fehlende Werte aufweisen, mit den Studien verglichen werden, die viele fehlende Werte aufweisen (Matt/Cook 2009: 544). Auch ist es bezüglich fehlender Werte denkbar, die Robustheit der Ergebnisse im Hinblick auf unterschiedliche Missing-Data-Ersetzungsverfahren zu überprüfen. Auch der Vergleich des Fixed- mit dem Random-Effects-Modell kann einen wichtigen Beitrag zur Sensitivitätsanalyse leisten (Greenhouse/Iyengar 2009: 425). Außerdem können die im vorangegangenen Abschnitt erläuterten Ansätze zum Aufspüren eines potenziellen Publikationsbias, wie das Trim-and-Fill-Verfahren, ebenfalls als Sensitivitätsanalyse für die Robustheit der Ergebnisse verstanden werden (Duval 2005: 131).<sup>104</sup>

Bei all diesen Ansätzen geht es im Grunde darum, die Grundgesamtheit der Studien oder die Analysemethode zu variieren (Greenhouse/Iyengar 2009: 418). Damit sollen die Auswirkungen, die bei der Durchführung einer Meta-Analyse zu treffenden Entscheidungen, auf die meta-analytischen Ergebnisse deutlich gemacht werden. Es ist aufschlussreich zu beobachten, wie sich die meta-analytischen Ergebnisse verändern, wenn einzelne Studien oder Subgruppen von Studien aus der Meta-Analyse ausgeschlossen werden.

Wie beim Publikationsbias gibt es hinsichtlich der Sensitivitätsanalyse bislang keine Vorschläge, wie sie in einer modellbasierten Meta-Analyse umgesetzt werden können. Allerdings lassen sich die angesprochenen Vorschläge leicht auf eine modellbasierte Meta-Analyse übertragen. Variationen der Analyseverfahren lassen sich ebenfalls umsetzen. Darüber hinaus ist es möglich, ein Robustheitstest unter Verwendung der unterschiedlichen Schätzverfahren durchzuführen. In der vorliegenden Studie werden verschiedene Sensitivitätsanalysen durchgeführt. Zum einen wird das Trim-and-Fill-Verfahren angewandt, um die bivariaten Beziehungen auf

---

<sup>104</sup> Die wohl aufwendigste Möglichkeit eine Sensitivitätsanalyse durchzuführen besteht darin, eine kumulative Meta-Analyse vorzunehmen, um zu zeigen, wie sich der kumulierte Effekt und die zugehörige Varianz mit jeder neu hinzugefügten Studie verändert (Borenstein et al. 2011: 371ff.). Dazu wird jeweils eine Studie zur Meta-Analyse hinzugefügt und die Serie der einzelnen Analysen in einer Tabelle oder einem Plot, meist chronologisch, dargestellt. Somit kann schnell ein Überblick darüber geschaffen werden, wie sich die wissenschaftliche Evidenz über die Zeit hinweg entwickelt.

einen möglichen Publikationsbias hin zu überprüfen. Zum anderen werden die Ergebnisse von unterschiedlichen Methoden zur Zusammenfassung von abhängigen Effektstärken innerhalb einer Studie miteinander verglichen. Darüber hinaus werden zur Untersuchung der Moderatorhypothesen zwei unterschiedliche Verfahren angewandt.



## 6 Literatursuche und Vor-Screening

Nun werden zunächst die Ergebnisse der beiden Bausteine der Suchstrategie (siehe Tabelle 12) der Reihe nach vorgestellt. Danach werden die Ergebnisse des Screenings beschrieben. Anschließend folgt eine Sammlung von Beobachtungen, die während der Literaturerhebung und dem Vor-Screening gemacht wurden. Abschließen wird das Kapitel mit einer Zusammenfassung der Ergebnisse dieser Arbeitsschritte.

### 6.1 Erhebung der Studien

Die Studierhebung erfolgt in zwei Schritten. Im ersten Schritt werden die Primärstudien aus den bereits bekannten Meta-Analysen, Literaturreviews und der Studiendatenbank des „Handbook of Religion and Health“ extrahiert. Diese beinhalten insgesamt 975 Literaturstellen (siehe Tabelle 21).

**Tabelle 21: Meta-Studien und Literaturreviews**

Meta-Studie und Literaturreviews	Anzahl
Hackney/Sanders 2003	34
Smith/McCullough Poll 2003	147
Sawatzky/Ratner/Chiu 2005	51
Visser/Garssen/Vingerhoets 2010	40
Koenig/King/Carson 2012	221
Schreiber/Brockopp 2012	18
Yonker/Schnabelrauch/DeHaan 2012	75
Eger/Maridal 2015	164
Salsman et al. 2015	148
Rizvi et al. 2016	77

Allerdings muss eine Reihe von Dubletten und Studien, die vor dem Jahr 2000 erschienen sind, ausgeschlossen werden. Nach der Datenbereinigung stehen noch 518 Studien für das Vor-Screening in Citavi<sup>105</sup> zur Verfügung. Der zweite Baustein der Literatursuche besteht aus der Suche in den Zitationsdatendanken. Dieser Teil der Erhebung wurde zwischen dem 20. und 27. April 2017 von einer wissenschaftlichen Hilfskraft und mir durchgeführt (siehe Tabelle 22).

**Tabelle 22: Datenbanksuche**

Datenbank	Suchdatum	Suchperson
PubMed	20.04.2017	AG
PubPsych	20.04.2017	AG
Web of Science	25.04.2017	MW
SOWIPORT	26.04.2017	MW
BASE	27.04.2017	MW

**Anmerkungen:** AG=Alexander Geisler, MW=Marius Wuketich.

<sup>105</sup> Umfangreiche Erläuterungen über die Funktionsweise und die Möglichkeiten von Citavi finden sich auf der Homepage des Herstellers: <https://www.citavi.com>. Das Programm bietet die Möglichkeit, die einzelnen Literaturstellen mit dem zugehörigen PDFs zu verknüpfen und somit ergibt sich die Möglichkeit papierlos zu arbeiten. Ebenso gibt es nützliche Funktionen wie „Dubletten finden“ oder „Volltext finden“, die die Erhebung oftmals vereinfacht haben.

Dabei wurden die zuvor beschriebenen Suchstrings für die Suche genutzt (siehe Tabelle 15). Die Suchstrings bezogen sich bei der Suche auf das „allgemeinste Feld“ der jeweiligen Datenbank.<sup>106</sup> Dabei wird Titel, Abstract und Keywords sowie oftmals (falls vorhanden) das gesamte Dokument durchsucht. Die beschriebene Suche wurde in allen Datenbanken eingeschränkt, um den Inklusionskriterien zu entsprechen. Zum einen wurde die Zeitspanne der Suche auf die Jahre 2000 bis zum Erhebungszeitpunkt (20. – 27 April 2017) und zum anderen wurden die Publikationssprachen auf Deutsch und Englisch eingeschränkt.<sup>107</sup> Somit konnten insgesamt 30.077 Literaturstellen durch die elektronische Datenbanksuche gefunden werden.

Im nächsten Schritt werden diese Literaturstellen mit der Citavi-Projektdatei, bestehend aus den bereits bekannten Meta-Studien und Literaturreviews, zusammengeführt. Diese Projektdatei enthält insgesamt 31.052 Literaturstellen. Alle doppelten Datenbankeinträge werden anschließend gelöscht. Dadurch verringert sich die Anzahl der Publikationen um 9.404 Datensätze. Nach der Datenpflege liegen noch 21.648 Datensätze für das Vor-Screening vor.

## 6.2 Vor-Screening

Nun folgt zunächst die Erläuterung des groben Vor-Screenings der Literaturstellen. Insgesamt werden 21.648 Literaturstellen einzeln auf ihre Eignung hin überprüft. Dafür werden die in Kapitel 4.2 formulierten Studienselektionskriterien angewandt, um reliable und nachvollziehbare Entscheidungen über den Ein- bzw. Ausschluss sicherstellen zu können.

Dabei werden Titel und Abstract zunächst auf die Passung hin überprüft. Erst wenn auf Grundlage dieser Informationen keine Entscheidung für „potenziell relevant“ oder „nicht relevant“ getroffen werden kann, wird der Volltext der Studie, mit besonderem Fokus auf den Analyseteil, gesichtet. Nach der Bereinigung der Daten („Datenpflege“) und dem Vor-Screening bleiben noch insgesamt 720 Literaturstellen übrig, die vorläufig als „vielleicht relevant“ gekennzeichnet sind. Dies entspricht etwa 3,3 % der bereinigten Studiendatenbank.<sup>108</sup> In Tabelle 23 ist eine Übersicht dieser Arbeitsschritte zu sehen.

Zu allen Studien werden anschließend, falls nicht bereits durch die Citavi-Funktion „Volltexte finden“ vorhanden, die Volltexte in PDF-Form online gesucht. Viele nicht direkt zugängliche

---

<sup>106</sup> Diese Suchstrings mussten teilweise einzeln verknüpft werden, da zwei Datenbanken (PubMed und PubPsych) eine Verknüpfung von AND und OR Verknüpfungen zum Zeitpunkt der Erhebung nicht erlaubten.

<sup>107</sup> Bei der Bielefeld Academic Search Engine (BASE) konnte das Sprachkriterium erst bei der Datenpflege kontrolliert werden, da die Spracheingrenzung nicht richtig funktionierte.

<sup>108</sup> Dies ist ein Hinweis dafür, dass die Literatursuche „zu breit“ bzw. die Suchstrings zu unpräzise waren. Deshalb wurden viele Studien gefunden, die für die empirische Untersuchung nicht relevant sind.

Literaturstellen konnten durch kontaktieren der Autoren (über [www.researchgate.com](http://www.researchgate.com))<sup>109</sup> beschafft werden. Der allergrößte Teil der Literatur ist somit zugänglich, nur einige ältere Monographien bzw. Sammelbände mussten per Fernleihe bestellt werden. Lediglich 49 Publikationen konnten nicht beschafft werden.

**Tabelle 23: Literatursuche und Vor-Screening**

<b>Herkunft</b>	<b>Anzahl</b>
Meta-Analysen/Literaturreviews	975
Web of Science	8.033
SOWIPORT	12.196
PubMed	8.181
PubPsych	1.159
BASE	508
<b>Gesamt</b>	<b>31.051</b>
Datenbereinigung	9.436
Vor-Screening Datensatz	21.615
<b>Vielleicht relevant</b>	<b>720</b>

Dabei handelt es sich vorwiegend um Hochschulschriften (Dissertationen und Masterarbeiten), die nicht direkt zugänglich sind (N=17) oder um Abstracts von Konferenzbeiträgen (N=12). Darüber hinaus sind 17 Artikel aus wissenschaftlichen Journals und drei Monographien nicht (mehr) zugänglich. Ausgehend von 31.051 Literaturstellen zu Beginn der Erhebung, wurden nach der Datenpflege noch 21.615 Literaturstellen in der Vor-Screeningphase durchgesehen. Dies führt zu 720 potenziell relevanten Studien.

---

<sup>109</sup> Bei dem Portal „Researchgate“ handelt es sich um ein internationales soziales Netzwerk im Wissenschaftsbereich. Dort sind sowohl Abstracts zu vielen Publikationen, aber auch Volltexte sowie unveröffentlichte Arbeiten und umfangreiche Foren- und Austauschmöglichkeiten zu finden.



## 7 Kodierung

Im Anschluss an die Dokumentation der Datenerhebung und des Vor-Screenings folgt nun die Dokumentation des Fein-Screenings und der Kodierung, sowie damit zusammenhängend, eine Zusammenfassung und ein Überblick über die, für die modellbasierte Meta-Analyse relevanten Publikationen.<sup>110</sup>

Zunächst wird das Vorgehen bei der Kodierung beschrieben.<sup>111</sup> Zunächst werden allgemeine Auffälligkeiten, die während der Kodierung zu Tage getreten sind, diskutiert. Anschließend wird das Ergebnis der Kodierung ausführlich anhand verschiedener zentraler Aspekte beschrieben. Dort folgt eine Übersicht über die berücksichtigten Messinstrumente der Variablen.

### 7.1 Vorgehen bei der Kodierung

Zur besseren Darstellung ist die Abfolge dieser Arbeitsschritte in Abbildung 14 veranschaulicht. Vor Beginn der eigentlichen Kodierung werden die im Screening bereits überprüften Publikationen (N=720) abermals auf ihre Eignung hin überprüft (Feinscreening). Besonderes Augenmerk liegt hierbei auf drei spezifischen Aspekten. Es wird geprüft, ob es sich bei den berichteten Korrelationen im Einzelfall auch um bivariate Korrelationen (Produkt-Moment-Korrelation) und nicht um partielle Korrelationen handelt.<sup>112</sup> Ebenfalls wird die Datenherkunft der jeweiligen Publikationen kontrolliert, um diejenigen Datensätze herausfiltern zu können, die entweder eine länderübergreifende Stichprobe verwenden oder die mehrmals in verschiedenen Publikationen genutzt wurden. Trifft einer der beiden geschilderten Fälle zu, werden die entsprechenden Publikationen bzw. Stichproben ausgeschlossen. Im Falle von doppelter Nutzung von Daten ist die Entscheidungsregel, dass die ältere Studie ausgeschlossen wird.<sup>113</sup> Der

---

<sup>110</sup> Die Unterscheidung von Publikationen und Stichproben ist von zentraler Bedeutung. Eine Publikation besteht bspw. aus einem Journalartikel. Eine einzelne Publikation kann mehrere Stichproben beinhalten. Eine Stichprobe hingegen wird definiert als eine mit den anderen Stichproben innerhalb einer Publikation überschneidungsfreie Stichprobe, deren Korrelationen separat in der Studie berichtet wird. Ein Beispiel hierfür sind, Korrelationen, die für Männer und Frauen oder für Studierende und Schüler getrennt voneinander berichtet werden. Diese können als unabhängige Stichproben in einer modellbasierten Meta-Analyse eingeschlossen werden, wenn es keine Überschneidungen zwischen den Stichproben gibt (Cheung/Hafdahl 2016: 117).

<sup>111</sup> Während der Vorarbeiten zur Kodierung wurden die Erhebungsmaterialien stetig weiterentwickelt. Das Kodierschema wurde während der Kodierung mehrfach verändert (von Version 1.0 bis 2.0) und die Kodierregeln wurden stetig erweitert, so dass sichergestellt werden konnte, dass alle Publikationen gleichbehandelt werden und getroffene Entscheidungen für nachfolgende Publikationen dokumentiert sind. Besonders im Hinblick auf tautologische Messinstrumente bei der unabhängigen Variable Religiosität/Spiritualität war dies sinnvoll, da die Informationen über die tautologischen Messinstrumente für die nachfolgenden Kodierungen dann zur Verfügung standen.

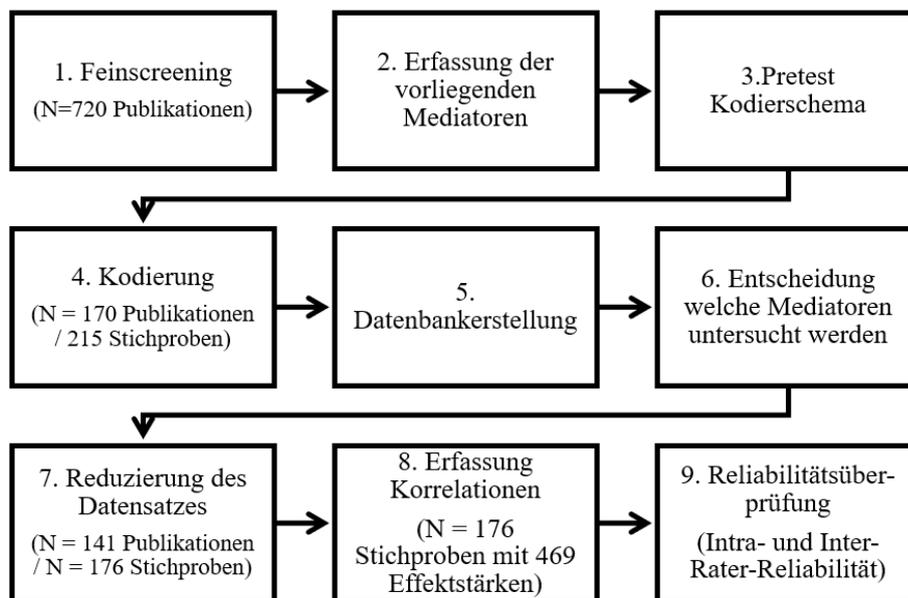
<sup>112</sup> Partielle Korrelationen sind nach bestimmten Variablen (oftmals demographische Variablen wie Alter) bereinigt (Urban/Mayerl 2018: 167).

<sup>113</sup> Der Grund hierfür liegt in der Erfahrung, dass oftmals ältere Publikationen in der Dokumentation ihres statistischen Vorgehens nachlässiger sind.

dritte Aspekt des Feinscreenings besteht darin, dass alle Messinstrumente von Religiosität/Spiritualität und Lebenszufriedenheit abermals betrachtet werden.

Alle Literaturstellen werden anschließend auf die untersuchten Mediatoren hin überprüft. In diesem Schritt wird nicht nur die Art des Mediators, sondern auch das verwendete Messinstrument erhoben, um das Kodierschema für eine standardisierte Erhebung der Mediatoren weiterzuentwickeln. Neben den vier berücksichtigten Erklärungsmechanismusbereichen (Kohäsion, kognitive Orientierung, positive Gefühle, alternative Werte) wurden weitere in der Empirie vorgefundene Mediatorvariablen in einen fünften Bereich „empirisch“ zusammengefasst. Das Kodierschema wurde dann an einer zufälligen Stichprobenauswahl einem Pretest unterzogen. Die Erkenntnisse aus diesen Tests werden dann bei der weiteren Fortentwicklung des Kodierschemas berücksichtigt.

**Abbildung 14: Ablauf der Stichprobenkodierung**



**Anmerkung:** Eigene Abbildung.

Die somit erhobenen Daten werden anschließend, mit Ausnahme der erhobenen Korrelationen, in eine Datenbank übertragen. Auf Grundlage dieser Datenbank entscheidet sich im nächsten Schritt, welche der empirisch vorliegenden Mediatorvariablen in der nachfolgenden Analyse untersucht werden.<sup>114</sup> Die Auswahl der zu berücksichtigten Mediatoren erfolgt dabei sowohl empirisch, als auch auf Basis von konzeptionellen Überlegungen. Lediglich die Mediatoren, die am häufigsten in der Stichprobengrundgesamtheit untersucht werden und aus unterschiedlichen

<sup>114</sup> Eine Beschränkung der Mediatoren ist aufgrund der Komplexität der statistischen Modellierung notwendig, da ansonsten eine ordnungsgemäße Durchführung der modellbasierten Meta-Analyse mit dem TSSEM-Verfahren, aufgrund der Vielzahl von zu schätzender Parameter nicht mehr gewährleistet werden kann. Bei einem übermäßig komplexen Modell und vielen fehlenden Korrelationen sind Konvergenzprobleme unvermeidlich.

Theoriebereichen stammen, werden für die Analyse verwendet. Somit werden drei Mediatorvariablen für die weitere Analyse ausgewählt. Daran anschließend werden zwei Makroindikatoren ausgewählt, die als Kulturproxyvariablen in der modellbasierten Meta-Analyse dazu genutzt werden, das Modell auf seine kulturelle Bedingtheit hin zu überprüfen (siehe Ausführungen in Kapitel 3.4). Die relevanten Korrelationen aus den verbleibenden 176 Stichproben werden anschließend ebenfalls in eine analysefähige Datenbank übertragen. Insgesamt handelt es sich dabei um 456 Effektstärken. Zum einen wird das hierarchische Design von Wilson/Polanin/Lipsey (2016) zum Umgang mit komplexen meta-analytischen Daten für die Korrelationsdatenbank angewandt und zum anderen wird eine herkömmliche modellbasierte meta-analytische Datenbank erstellt. Dies ermöglicht die Anwendung verschiedener Verfahren zum Umgang mit abhängigen Effektstärken.

Im letzten Schritt werden die erhobenen Daten auf ihre Reliabilität hin überprüft. Auch wenn es sich bei der vorliegenden Arbeit lediglich um eine Exemplifizierung der modellbasierten meta-analytischen Methodik handelt, ist eine Überprüfung der Kodierungen obligatorisch. Dazu werden die beiden für die Analyse zentralen Variablen überprüft. Zum einen sind dies die Korrelationen der relevanten Modellbeziehungen und zum anderen die zur jeweiligen Studie zugehörige Fallzahl. Zunächst werden alle Korrelationen und die Fallzahlen erneut überprüft (Intra-Rater-Reliabilität), bevor im nächsten Schritt 20 % der Stichproben ( $n = 35$ ) zufällig ausgewählt und von einer weiteren Wissenschaftlerin wiederholt überprüft werden (Inter-Rater-Reliabilität). Auffälligkeiten werden anschließend kontrolliert und angepasst.

## **7.2 Kodierungsergebnisse**

Zunächst werden nun allgemeine Auffälligkeiten berichtet, die während der Kodierung deutlich geworden sind, bevor im Anschluss daran, die Ergebnisse deskriptiv dargestellt werden.

Was die fachliche Herkunft der Literaturstellen betrifft, stammen die meisten aus den Humanwissenschaften. Mit Ausnahme weniger Arbeiten sind fast alle in Fachzeitschriften aus den Bereichen Psychologie oder Medizin veröffentlicht worden. Einige lassen sich jedoch auch den Sozialwissenschaften zuordnen. Die geographische Herkunft ist ebenso ungleich verteilt. Nahezu die Hälfte der Analysen wurde in den Vereinigten Staaten durchgeführt. Die Effektstärken aus den USA und dem arabischen Raum scheinen dabei höher zu sein, als in stärker säkularen Staaten. Die Fallzahl der inkludierten Stichproben ist heterogen. Wie in den Humanwissenschaften üblich, sind diese dort eher gering. Jedoch haben einige Stichproben auch große Samples mit tausenden Probanden untersucht. Besonders auffällig ist, dass vor allem die unabhängige Variable Religiosität/Spiritualität oftmals mit tautologischen Messinstrumenten erfasst wird.

Viele dieser Messinstrumente beinhalten Wohlbefindensindikatoren oder Indikatoren aus dem Bereich der Mediatorvariablen. Eine Aufzählung der tautologischen Messinstrumente ist in Tabelle 24 zu sehen. Stichproben, die lediglich ein tautologisches Messinstrument zur Erfassung von R/S verwenden, werden entsprechend der Selektionskriterien ausgeschlossen, da die unverfälschte Messung des Zusammenhangs ein zentrales Ziel der vorliegenden Studie ist. Sind mehrere Messinstrumente zur Erfassung der Variable vorhanden, werden nur die überschneidungsfreien Messinstrumente mit den zugehörigen Korrelationen erfasst.

Mit zunehmenden Ausschluss von tautologischen Messinstrumenten bzw. Stichproben war eine immer moderatere Anzahl von Messinstrumenten relevant (siehe dazu Tabelle 27). Die Korrelationen der Stichproben, die tautologische Messungen verwenden, sind subjektiv betrachtet etwas höher, als diejenigen Korrelationen, die auf tautologiefreien Messinstrumenten beruhen. Es muss davon ausgegangen werden, dass verunreinigt gemessene Korrelationen zu einer moderaten Überschätzung des tatsächlichen Zusammenhangs führen.

**Tabelle 24: Tautologische Messinstrumente**

<b>Konstrukte</b>	<b>Tautologische Messinstrumente</b>
Religiosität/Spiritualität	Spirituality Rating Scale (Pinto/Pais-Ribeiro 2007) Religious Meaning System Questionnaire The Spirituality Scale (Delaney 2005) Daily Spiritual Experience Scale (DSES) (Underwood/Teresi 2002) Functional Assessment of Chronic Illness Therapy –Spiritual Well-Being Scale(FACIT-sp) und FACIT-SP12 Jarel Spiritual Well-Being Scale (Hungelmann et al.1996) Religious Coping Activities Scale – subscale Discontent (Pargament et al. 1990) Rush Religious/Spiritual Struggle Screening Protocol (Fitchett/Risk 2009) Spiritual Index of Well-Being (SIWB) (Daaleman/Frey 2004) Spiritual Well-Being Scale (SWBS) (Ellison 1983; Ellison/Smith 1991) Existential Well-Being Scale Religious Well-Being Scale Spiritual Well-Being Questionnaire (SWBQ) (Gomez/Fisher 2003) WHOQOL Spirituality, Religion and Personal Beliefs (WHOQOL SRPB) (O’Connell et al. 2006) Ironson Woods S/R-Index Faith Maturity Scale (FMS) Spiritual Involvement and Beliefs Scale Revised (SIBS R) Brief Multidimensional Measure of Religiousness/Spirituality (mit Ausnahmen siehe Tabelle 27) Spiritual Assessment Scale (Howden 1992) Spiritual Meaning Scale (SMS) (Mescaro 2004/2006) Expressions of Spirituality Inventory ESI (MacDonald, 2000a, 2000b)
Lebenszufriedenheit	Psychological well-being (Ryff/Keyes)

**Anmerkung:** Die Skalen werden nicht im Literaturverzeichnis zitiert.

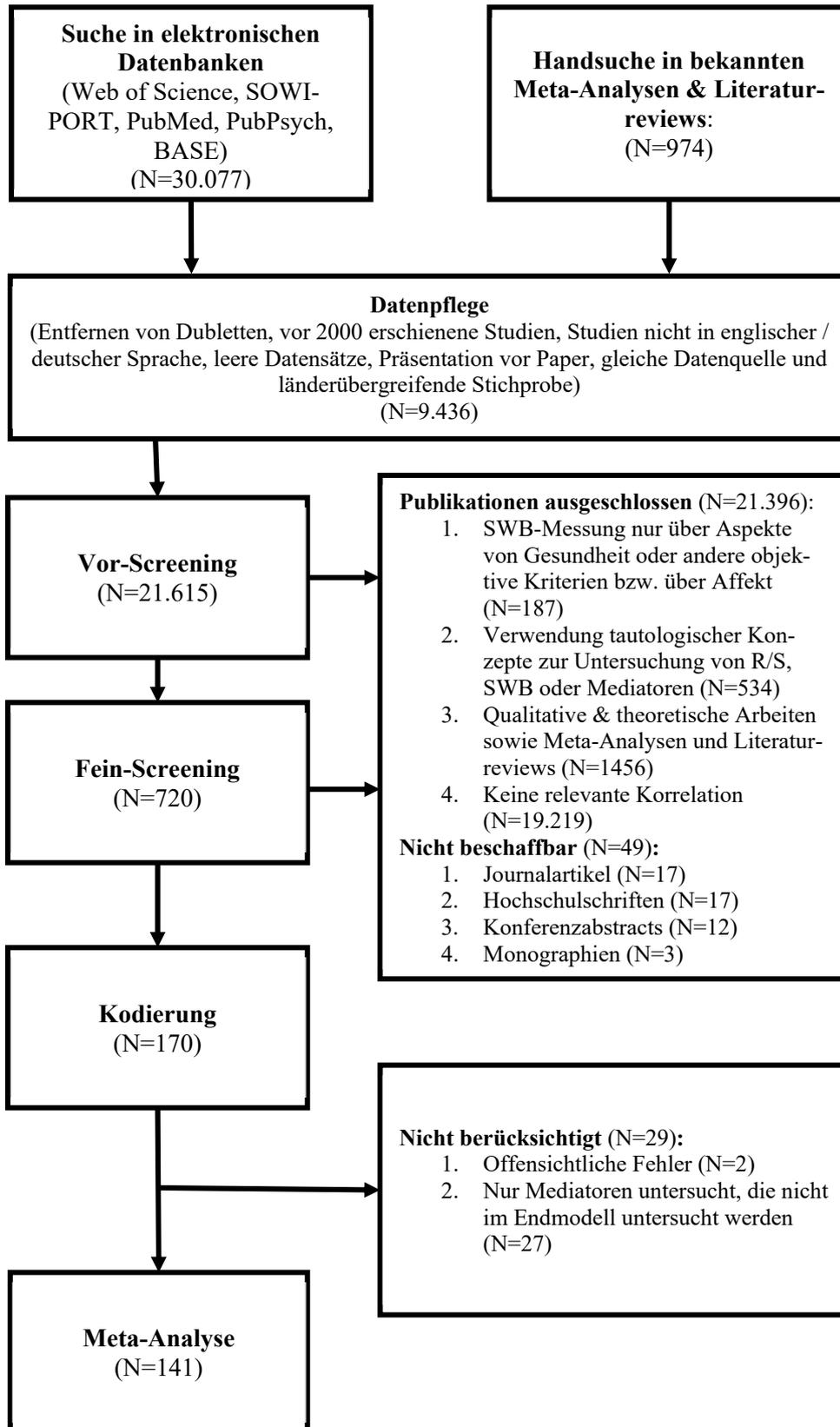
Bereits im vorangegangenen Kapitel wurde berichtet, dass viele Stichproben mehrere Messinstrumente zur Erfassung von R/S, Lebenszufriedenheit und der Mediatoren verwendet haben.

Diese mussten entsprechend getrennt voneinander erfasst werden, um anschließend die verschiedenen Methoden zur Berücksichtigung abhängiger Effektstärken miteinander vergleichen zu können. Häufig kommen fehlende Werte, im Sinne von nicht berichteten, aber untersuchten Korrelationskoeffizienten, vor.<sup>115</sup> Nach Beendigung der Kodierung werden alle relevanten Stichproben abermals in Hinblick auf die Korrelationskoeffizienten und die Fallzahlen kontrolliert, um die Intra-Rater-Reliabilität dieser zentralen Variablen zu überprüfen. Dabei können eine Reihe von Fehlern, im wesentlichen Tippfehler, aufgefunden und beseitigt werden (n = 34). Im nächsten Schritt werden 20% der Stichproben (n = 35) zufällig ausgewählt und von einer weiteren Wissenschaftlerin, nach einer Einführung in Kodierschema und Kodierprozess, abermals überprüft. Dabei wurden sieben Auffälligkeiten festgestellt, von denen sich bei der nachfolgenden Kontrolle vier als tatsächliche Fehler herausgestellt haben, die verbessert wurden. In Abbildung 15 sind zur Veranschaulichung die beiden Arbeitsschritte Screening und Kodierung in Form eines PRISMA-Flow-Diagramms dargestellt (Liberati et al. 2009).

---

<sup>115</sup> Zwei Stichproben mussten aus der weiteren Untersuchung ausgeschlossen werden, da offensichtliche Fehler bei der Berichterung der Korrelationen vorliegen und nicht klar ist, ob es sich um eine positive oder negative Korrelation handelt.

**Abbildung 15: Zusammenfassung von Screening und Kodierung**



**Anmerkung:** Eigene Abbildung.

In Tabelle 25 ist eine Übersicht über die Mediatorvariablen und deren Untersuchungshäufigkeit zu finden. Die erst bei der Durchsicht entdeckten Variablen Self-Efficacy und Sense of Control

lassen sich den zuvor herausgearbeiteten Theoriebereichen zuordnen. Sense of Control (verschiedene Kontrollüberzeugungen) fällt in den Bereich „Theorie der Verhaltensregulierung“. Der Mediator „Self-Efficacy“ (Selbstwirksamkeitsüberzeugung) kann dem Bereich der „positiven Gefühle“ zugeordnet werden.

**Tabelle 25: Empirisch gefundene Mediatoren**

Mediator	Häufigkeit
Soziale Unterstützung	41
Self-Esteem / Selbstwert	39
Optimismus	29
Meaning in Life / Lebenssinn	20
Purpose in Life / Lebenszweck	17
Sense of Control	15
Hoffnung	14
Self-Efficacy	13
Forgiveness	8
Kohärenzgefühl	6
Gratitude	3

**Anmerkung:** Basierend auf 170 Publikationen.

Wie bereits angedeutet, werden aufgrund der Durchführbarkeit der nachfolgenden statistischen Analysen drei Mediatorvariablen auf Grundlage der Häufigkeit und der Abdeckung der verschiedenen Bereiche ausgewählt. Die drei Variablen „soziale Unterstützung“ (Bereich: Kohäsion), „Optimismus“ (Bereich: kognitive Orientierung) und „Selbstwert“ (Bereich: positive Gefühle) werden ausgewählt, da diese Variablen am häufigsten über alle Stichproben hinweg untersucht wurden.

Alle Stichproben bzw. Publikationen, die lediglich die Beziehungen zwischen abhängiger Variable oder unabhängiger Variable und eine der nun nicht mehr relevanten Mediatorvariablen beinhalten, werden nach diesem Arbeitsschritt entfernt, so dass die finale Untersuchungsdatenbank 141 Publikationen mit insgesamt 176 Stichproben umfasst. In Anhang B ist eine vollständige Übersicht zu sehen. Es ist deutlich erkennbar, dass keine der Stichproben alle drei der ausgewählten Mediatoren enthält. Jedoch untersuchen einige Stichproben zwei der ausgewählten Erklärungsmechanismen (N=16). Immerhin 74 Stichproben haben zumindest einen der drei relevanten Mediatoren in ihrem Untersuchungsdesign berücksichtigt.

Betrachtet man die Anzahl der Effektstärken (Korrelationen), die die Stichproben jeweils zur Meta-Analyse beitragen, wird eine große Spannweite deutlich. Viele Stichproben liefern nur eine einzige Korrelation für die Untersuchung, während die Studie mit der höchsten Anzahl 15 Effektstärken beisteuert. Je Studie liegt der Mittelwert bei 2,7 Effektstärken und der Median bei 2. In Abbildung 16 ist ein Überblick über die Anzahl der Effektstärken je Studie zu sehen.

67 Stichproben liefern nur eine Effektstärke, die restlichen 109 Stichproben liefern mehr als eine Effektstärke für die Untersuchung.<sup>116</sup>

**Abbildung 16: Anzahl der Effektstärken**



**Anmerkung:** Eigene Abbildung.

Immerhin sind Stichproben aus 30 unterschiedliche Staaten in der Datenbank vertreten und fast alle geographischen Bereiche der Erde abgedeckt (siehe Tabelle 26).

**Tabelle 26: Stichprobenübersicht nach Ländern**

Land	Anzahl
USA	91
Kuwait	21
Australien	5
Iran	5
Kanada	4
Libanon	5
Indien	4
Katar	4
Ägypten	3
Algerien	3
Deutschland	3
Irland	3
Polen	3
Belgien	2
Palästinensische Autonomiegebiete	2
Sambia	2
Vereinigtes Königreich	2
Je 1: China, Griechenland, Hongkong, Malaysia, Malta, Neuseeland, Pakistan, Philippinen, Portugal, Slowakei, Spanien, Südkorea, Ungarn	1

**Anmerkung:** Die Palästinensischen Autonomiegebiete sind völkerrechtlich zwar kein anerkannter Staat, aber kulturell sicherlich von Israel zu unterscheiden.

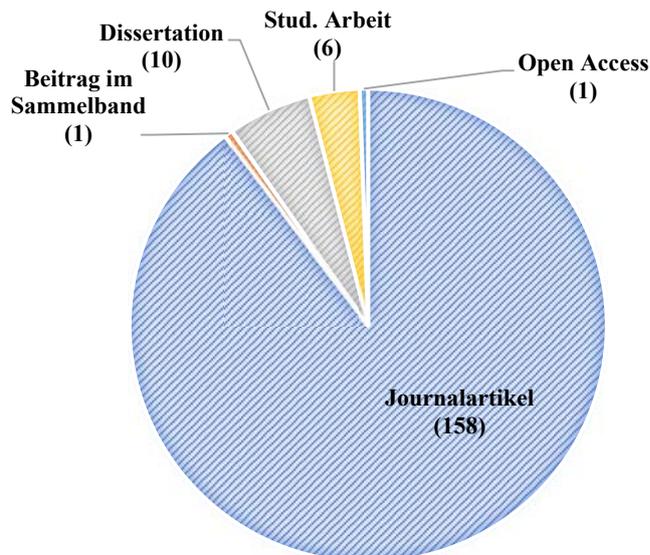
Die gesamte Stichprobengröße beträgt 205.485 Befragte. Dabei gibt es eine große Varianz zwischen den Stichproben. Die kleinste Stichprobengröße beträgt gerade mal 17 Personen, während die größte 93.679 Personen beinhaltet. Die durchschnittliche Stichprobengröße beträgt

<sup>116</sup> Hierbei wird die Notwendigkeit von multivariaten meta-analytischen Verfahren und die Verwendung von Methoden zur Berücksichtigung abhängiger Effektstärken deutlich.

1.167 Personen, wird allerdings durch die großen Stichproben verzerrt. Der Median beträgt 251 Personen, so dass die eingeschlossenen Stichproben eher auf kleinen Fallzahlen beruhen.

Der Ursprung der Stichproben ist eindeutig (siehe Abbildung 17). 90,3% der Stichproben (N=159) stammt aus herkömmlichen Quellen wie Journalartikel und Sammelbandbeiträge. Die restlichen 9,6 % können dem Bereich der grauen Literatur (N=17) zugeordnet werden.

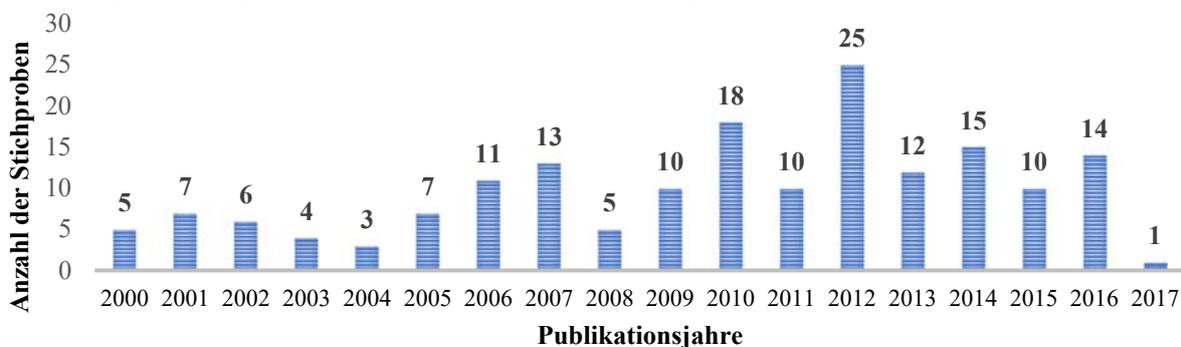
**Abbildung 17: Stichprobenübersicht nach Art der Studie**



**Anmerkung:** Eigene Abbildung.

Die herkömmliche Literatur besteht aus 158 Stichproben aus wissenschaftlichen Journalen und nur einem Beitrag in einem Sammelband. Lediglich ein Journal („Mental Health, Religion & Culture“) ist mit 26 Stichproben häufig vertreten. Die graue Literatur setzt sich aus 10 Dissertationen, 6 studentischen Arbeiten (Bachelor und Master) sowie einem Open-Access-Beitrag aus einer Schriftenreihe zusammen. Betrachtet man die Anzahl der Stichproben nach Publikationsjahr, so lassen sich kaum Auffälligkeiten bestimmen (siehe Abbildung 18).

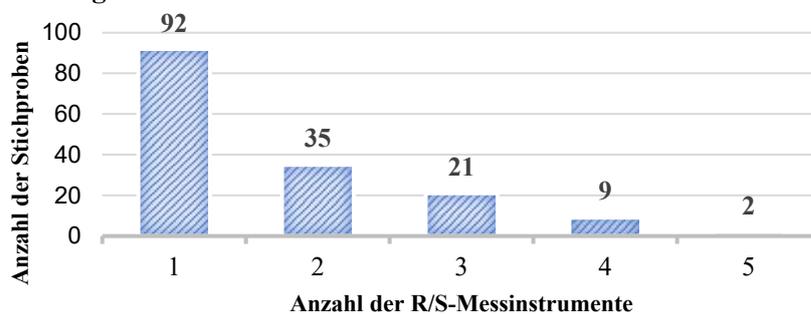
**Abbildung 18: Stichprobenübersicht nach Publikationsjahr**



**Anmerkung:** Eigene Abbildung.

Über die gesamten 17 Erhebungsjahre (2000 - 2017) hat die Anzahl der jährlich publizierten Stichproben leicht zugenommen. Dies entspricht dem Eindruck, dass das Forschungsfeld seit etwa den 70er Jahren des vergangenen Jahrhunderts zunehmend an Beliebtheit gewinnt.<sup>117</sup> Die Zahl für das Jahr 2017 ist aufgrund des frühen Erhebungszeitraums (20.04.2017 – 27.04.2017) niedrig. Bezüglich der Anzahl von Messinstrumenten für die unabhängige Variable Religiosität (siehe Abbildung 19) wird deutlich, dass fast die Hälfte aller Stichproben (42,5 %), die eine R/S-Messung beinhalten, mehr als ein Messinstrument zur Erfassung von R/S verwendet haben (n = 67). Dies entspricht einem Mittelwert von 1,7 Messinstrumenten pro Studie und einem Median von 1.

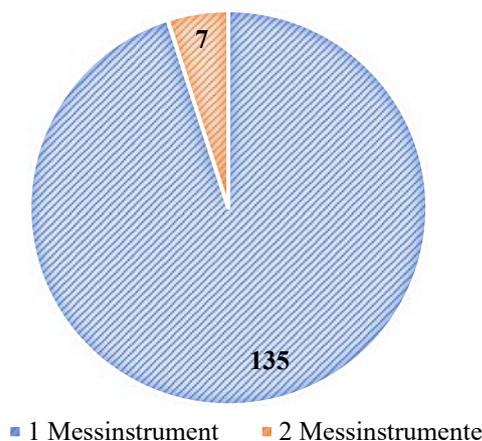
**Abbildung 19: Anzahl von R/S Messinstrumenten**



**Anmerkung:** Eigene Abbildung; 159 Stichproben haben R/S gemessen.

Ganz anders hingegen sieht es bei der Messung der abhängigen Variablen Lebenszufriedenheit aus (siehe Abbildung 20). Insgesamt beinhalten 141 der 176 Stichproben ein Messinstrument zur Erfassung der Variable. 135 Stichproben dieser Stichproben enthalten ein Messinstrument und nur 7 Stichproben enthalten zwei Messinstrumente (5 %).

**Abbildung 20: Anzahl von Lebenszufriedenheitsmessinstrumenten**



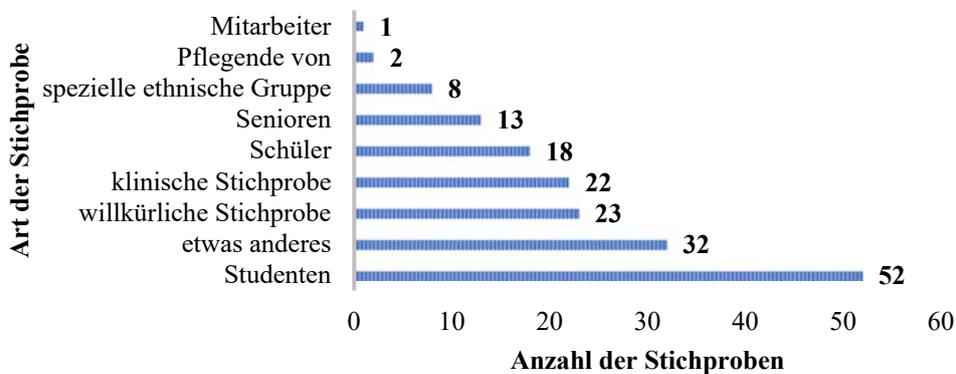
**Anmerkung:** Eigene Abbildung; 141 Stichproben haben Lebenszufriedenheit gemessen.

<sup>117</sup> Jedoch kann die Expansion von wissenschaftlichen Veröffentlichungen ebenfalls dafür verantwortlich gemacht werden.

Nach der Reduzierung der Mediatorvariablen auf die drei oben genannten Variablen, gibt es lediglich bei einer Studie mehreren Messinstrumenten zur Erfassung desselben Mediatorkonstrukts (S-ID 160\_1).

Betrachtet man die Repräsentativität der zugrundeliegenden Stichproben kommt man zu einem wenig überraschenden Ergebnis. Lediglich 5 der 176 (2,8 %) Stichproben sind repräsentativ für die jeweilige Bevölkerung (USA, Deutschland, Polen, Neuseeland und Algerien). Die Fallzahl dieser repräsentativen Stichproben beträgt insgesamt 28.448. Dies entspricht 13,8 % der gesamten, der Meta-Analyse zugrundeliegenden Stichprobe. Im Umkehrschluss sind 171 Stichproben (97,2 % aller Stichproben) Spezialpopulationen. In der nachfolgenden Abbildung 21 ist eine Übersicht über die verwendeten unterschiedlichen Populationen zu finden.

**Abbildung 21: Nicht-repräsentative Stichproben nach Stichprobenart**



**Anmerkung:** Eigene Abbildung; 171 von 176 Stichproben basieren auf der Untersuchung einer Spezialpopulation. Besonders viele Stichproben basieren auf der Untersuchung von Studierenden (ca. 30 % aller Stichproben). Aber auch andere Personengruppen, wie spezifische ethnische Gruppen (bspw. Ureinwohner oder Afroamerikaner) oder Patienten (Krebs, Schlaganfall etc.), sind dort zu finden. Das durchschnittliche Alter der verwendeten Stichproben wurde in 141 von 176 Stichproben berichtet (79,6 %). Die jüngste Stichprobe ist dabei lediglich 14,1 Jahre alt und die älteste 80,5 Jahre. Über alle Stichproben hinweg beträgt das durchschnittliche Stichprobenalter 37,2 Jahre bei einem Median von 32. Betrachtet man die Geschlechterverteilung aller Stichproben machen 11 der 176 Stichproben keinerlei Angaben zur Geschlechterverteilung. In den Studien, die entsprechende Angaben machen, sind 39 % aller Personen Männer und 61 % Frauen. Wobei 32 Stichproben nur aus Frauen und 23 Stichproben nur aus Männern bestehen.

In der nachfolgenden Tabelle 27 ist eine Übersicht über die in den Stichproben verwendeten Messinstrumente für alle Konstrukte zu sehen. Es ist zu erkennen, dass mit Ausnahme der Mediatoren M4 (Optimismus) und M7 (Selbstwert) eine große Anzahl unterschiedlicher Messinstrumente vorliegt.

**Tabelle 27: Berücksichtigte Messinstrumente**

Konstrukt	Messinstrumente
Religiosität/Spiritualität	<p>Duke University Religion Index            Attitudes toward Islam Scale            Religious Orientation Scale-Revised (Gorsuch/McPherson 1989)            Spiritual Transcendence Index (STI; Seidlitz et al. 2002)            Spirituality Perspective Scale (SPS; Reed 1987)            Spiritual Involvement Skala – Unterskala Spiritual Life Integration            The Spiritual Maturity Inventory (Ellison 1983)            Religious Commitment Inventory (RCI-10; Worthington et al. 2003)            The Intrinsic Religious Motivation scale (Hoge 1972)            Religiousness involvement scale            Religiosity Scale (Chatters et al. 1992)            Religious Importance Scale (RIS; Putney/Middleton 1961)            4-item religiosity scale (Cohen/Rankin 2004)            Centrality of Religious Meaning System Scale (Dezutter et al.)            Brief Multidimensional Measure of Religiousness and Spirituality (BMMRS; Fetzer Institute/National Institute on Aging Working Group 1999) – Unterskalen: Values/beliefs; private religious practice; organizational religiousness            Francis Scale of Attitude towards Christianity (FSAC)            Santa Clara Strength of Religious Faith Questionnaire (SCSRF; Plante/Boccaccini 1997)            Feagin's (1964) Skala (9 Items)            revised Quest Scale (Batson/Schoenrade 1991)            Cultural Connectedness Scale-Short Version (CSS-S) – Spiritualitätsskala            GES Questionnaire (Benito et al. 2014) - intrapersonal            10-item Intrinsic Religiosity Scale (IR)            Assessment of Spirituality and Religious Sentiments (ASPIRES; Piedmont 1999)            Religious Belief Salience Scale (Blaine/Crocker 1995)            Religiosity Measure (Rohrbaugh/Jessor 1975)            Religious Background and Behavior (RBB)            DUREL Scale            25 Item ASP1            Index of Religiousness            eigene Messung (1 Item)            eigene Messung (mehrere Items)</p>
Soziale Unterstützung (M1)	<p>Anzahl enger Freunde            Duke Social Support Index            Multidimensional Scale of Perceived Social Support (MSPSS; Zimet et al. 1988)            Inventory of Socially Supportive Behaviors (ISSB)            Network Analysis Profile (NAP; Cohen/Sokolovsky 1979; Sokolovsky/Cohen, 1981)            Lubben Social Network Scale (LSNS-R; Lubben et al., 2001)            Sense of Support Scale (SSS; Dolbier and Steinhart 2000)            Social Provisions Scale (SPS; Cutrona &amp; Russell, 1987)            Social Support Questionnaire + SSQ6 (Sarason/Sarason/Shearin 1986)            The Social Support Appraisals Scale (Vaux et al., 1986)            Adult Social Support Questionnaire            Personal Resource Questionnaire (PRQ2000; Weinert 2003)            2-Way Social Support Survey (Shakespeare-Finch/Obst, 2011)            Perceived social support (PSS-Fr) - Freund-Skala (Procidano/Heller 1983)            Skala von Caplan et al. 1980            eigene Messung mehrere Items</p>
Optimismus (M4)	<p>Arabic Scale of Optimism and Pessimism (ASOP)            The Life Orientation Test-Revised (LOT-R)            eigene Messung mehrere Items</p>
Selbstwert (M7)	<p>Self-esteem scale (S-ES; Rosenberg 1965)            eigene Messung mehrere Items</p>
Lebenszufriedenheit	<p>3-item Satisfaction with Life Scale (SWLS)            5-item Satisfaction with Life Scale (SWLS)            Satisfaction with Life Scale (SWLS) – ohne nähere Angabe            Life Satisfaction Index-Z (LSI-Z 12)            Cantril Ladder/Ladder Scale of General Well-Being (1 Item -10er Skala)            Messinstrument von Campbell, Converse, and Rodgers (1976)            Brief Multidimensional Life Satisfaction Scale (BMLSS)            Messinstrument von Poloma/Pendleton 1989            the Student Life Satisfaction Scale (SLSS)            The Satisfaction with Life Scale for Children            Life Satisfaction Index-A (LSI-A)            eigene Messung (1 Item)            eigene Messung (mehrere Items)</p>

**Anmerkungen:** Es sind nur diejenigen Skalen erfasst, die aufgrund der Angaben in den Publikationen präzise erfasst werden konnten. Die Skalen werden nicht im Literaturverzeichnis aufgeführt.

Lediglich zwei Makro-Indikatoren werden für die kulturelle Moderatoranalyse zusätzlich, zu den in der Kodierung erhobenen Daten, in den Datensatz aufgenommen. Zum einen der „Human-Development-Index (HDI)“ und zum anderen die aggregierte „Wichtigkeit von Religion“ basierend aus Umfragedaten des World-Value-Survey (WVS). Die Daten des HDI liegen vollständig für alle Länder vor.<sup>118</sup> Der jeweilige Human-Development-Wert wird bezüglich des Erhebungsjahres der Daten, oder wenn dieses nicht vorliegt, bezüglich des Publikationsdatums einer Studie zugeordnet. Der HDI schwankt über alle Stichproben hinweg von .487 (Pakistan) bis zu .939 (Australien). Der Mittelwert beträgt .850 und der Median .889.

Lediglich für neun der 176 Stichproben (5,1 %) liegen die Werte der zweiten Makrovariable „Wichtigkeit von Religion“ nicht vor.<sup>119</sup> Dies hat zur Folge, dass in der nachfolgenden Analyse ein vollständiger HDI-Datensatz (HDI-DS) und ein reduzierter WR-Datensatz (WR-DS) verwendet wird. Bei den Ausprägungen dieser Makrovariable liegt eine große Spannbreite vor. Die kleinste Zustimmungsrates bei der WR-Variable liegt bei 2,6 % in China, während der höchste Wert in Katar 98,9 % Prozent vorliegt. Somit gibt es eine sehr große Heterogenität in Bezug auf die durchschnittliche Religiosität der Gesellschaften, in denen die Stichproben angefertigt worden sind.

Insgesamt werden mit dem geschilderten Vorgehen 141 relevante Publikationen ausgewählt, die insgesamt 176 unabhängige Stichproben beinhalten. Nach Reduzierung der zu untersuchenden Mediatorvariablen auf drei Konstrukte enthalten diese 456 relevante Effektstärken (Korrelationen) auf Basis von 205.485 Befragten.

---

<sup>118</sup> Siehe: <http://hdr.undp.org/en/data>, aufgerufen am 25.06.2018.

<sup>119</sup> Dabei handelt es sich um die folgenden Stichproben: 058\_1; 059\_1; 073\_1; 079\_2; 109\_1; 114\_1; 117\_1; 123\_1; 162\_1.



## 8 Erwartungen

Vor der Durchführung der modellbasierten Meta-Analyse werden auf Basis der bisherigen Erkenntnisse Erwartungen zum Untersuchungsmodell und zu den methodischen Fragen ausformuliert. Auf eine vollständige Verschriftlichung, aller in einem solch komplexen Untersuchungsmodell eingeschlossenen Hypothesen, wird an dieser Stelle, in Anlehnung an die modellbasierte meta-analytische Praxis, verzichtet. Das konstruierte Erklärungsmodell enthält ohnehin die entsprechenden inhaltlichen Hypothesen. Darüber hinaus handelt es sich um eine explorative Arbeit, bei der zum meta-analytischen Untersuchungsmodell bislang keine empirischen Erkenntnisse vorliegen, so dass vor allem bezüglich der Auswirkungen der Moderatoren auf die einzelnen Regressionskoeffizienten kaum konkrete Erwartungen formuliert werden können. Anstelle von konkreten Hypothesen werden jedoch ausgewählte Erwartungen im Folgenden formuliert. Eine Übersicht über die formulierten Erwartungen und in welchem Kapitel diese getestet werden, ist in Tabelle 28 zu sehen. Ebenfalls ist zu sehen, wo die Befunde auf Sensitivität überprüft werden.

Die erste Erwartung bezieht sich auf die Konvergenz des Modells (E1). Es wird vermutet, dass das analytisch hergeleitete Modell auch empirisch reproduziert werden kann. Bereits bei der Sichtung der Studien zeigt sich eine starke Befundheterogenität. Deshalb wird vermutet, dass auch empirisch auf Basis der  $Q$ -Statistik und der  $I^2$ -Werte ein Random-Effects-Modell in Einklang mit dem analytischen Fokus dieser Arbeit (Generalisierung) geschätzt werden sollte (E2).

Inhaltlich wird der Literatur folgend ein positiver Haupteffekt von Religiosität auf Lebenszufriedenheit erwartet (E3.1). Das gleiche gilt für die drei ins Untersuchungsmodell eingeschlossenen Mediatorvariablen. Es wird sowohl ein positiver Zusammenhang zwischen Religiosität und den Mediatoren soziale Unterstützung (E3.2), Optimismus (E3.3) und Selbstwert (E3.4), als auch zwischen den Mediatoren und Lebenszufriedenheit (E3.5, E3.6 und E3.7) erwartet. Für alle drei Erklärungsmechanismen wird eine partielle Mediation vermutet, die sich darin zeigt, dass sowohl der direkte, als auch der indirekte Effekt signifikant ist (E4.1, E4.2 und E4.3). Zwischen den Mediatorvariablen werden positive Kovariationen (E5.1, E5.2 und E5.3) angenommen. Über die Stärke der Effekte werden an dieser Stelle keine expliziten Vermutungen aufgeführt, sie dürften aber im schwachen bis moderaten Bereich liegen.

Auch bezüglich der Wirkungen der Moderatorvariablen liegen bestimmte Vermutungen vor. Für den HDI wird erwartet, dass in höher entwickelten Ländern die Effekte generell schwächer

ausgeprägt sind. Vor allem beim Haupteffekt, also zwischen Religiosität und Lebenszufriedenheit, wird ein moderierender Effekt vermutet (E6.1). Bei der Variable Wichtigkeit von Religiosität wird hingegen angenommen, dass die Effekte in Ländern, in denen die aggregierten Werte höher sind, auch stärker ausfallen (E6.2). Es wird jedoch nicht ausgeschlossen, dass die beiden Kultur-Makrovariablen auch noch auf andere Regressionseffekte des Untersuchungsmodells wirken.

**Tabelle 28: Erwartungen**

	Beschreibung	Prüfung	Sensitivität
E1	Untersuchungsmodell konvergiert	9.3	9.4 und 9.5
E2	Random-Effects-Modell	9.3	9.4 und 9.5
E3.1	b34: positiver Effekt (Religiosität → Lebenszufriedenheit)	9.3	9.4 und 9.5
E3.2	b31: positiver Effekt (Religiosität → soziale Unterstützung)	9.3	9.4 und 9.5
E3.3	b32: positiver Effekt (Religiosität → Optimismus)	9.3	9.4 und 9.5
E3.4	b33: positiver Effekt (Religiosität → Selbstwert)	9.3	9.4 und 9.5
E3.5	b35: positiver Effekt (soziale Unterstützung → Lebenszufriedenheit)	9.3	9.4 und 9.5
E3.6	b36: positiver Effekt (Optimismus → Lebenszufriedenheit)	9.3	9.4 und 9.5
E3.7	b37: positiver Effekt (Selbstwert → Lebenszufriedenheit)	9.3	9.4 und 9.5
E4.1	ind1: partieller positiver Mediator (soziale Unterstützung)	9.3	9.4 und 9.5
E4.2	ind2: partieller positiver Mediator (Optimismus)	9.3	9.4 und 9.5
E4.3	ind3: partieller positiver Mediator (Selbstwert)	9.3	9.4 und 9.5
E5.1	k1: positive Kovariation (soziale Unterstützung und Optimismus)	9.3	9.4 und 9.5
E5.2	k2: positive Kovariation (soziale Unterstützung und Selbstwert)	9.3	9.4 und 9.5
E5.3	k3: positive Kovariation (Optimismus und Selbstwert)	9.3	9.4 und 9.5
E6.1	Negativer Moderator HDI → b34; weitere Interaktionsterme möglich	9.4.1	9.4.2
E6.2	Positiver Moderator WR → b34; weitere Interaktionsterme möglich	9.4.1	9.4.2
E6.3	Kleinere Effektstärken bei Studien mit großer Fallzahl	9.4.1	-
E6.4	Kein Publikationsbias (herkömmliche Literatur vs. graue Literatur)	9.4.1	-
E6.5	Kleinere Effektstärken bei repräsentativen Stichproben	9.4.1	-
E6.6	Kleinere Effektstärken bei schlechter bewerteten Studien	9.4.1	-
E6.7	Kein Einfluss des Publikationsjahrs	9.4.1	-
E7.1	Konvergenzprobleme beim Subgruppendesign	9.4.2	-
E7.2	Präzisere Schätzung (schmalere LBCIs) beim Wilson- und robusten Ansatz	9.5.2	-

Darüber hinaus werden fünf weitere methodische Moderatoruntersuchungen vorgenommen. Die fünf untersuchten Variablen sind Fallzahl, Publikationsstatus, Repräsentativität, Studienqualität und Publikationsjahr. Studien mit großer Fallzahl haben vermutlich kleinere Effektstärken, als kleinere Studien, da in kleineren Studien häufig extremere Werte zu finden sind (E6.3). Der Moderator Publikationsstatus wird als Dummyvariable überprüft, wobei herkömmliche Literatur, grauer Literatur gegenübergestellt wird, um zu überprüfen, ob ein Publikationsbias vorliegt (E6.4). Ebenfalls eine Dummyvariable stellt der Moderator Repräsentativität dar. Dabei wird überprüft, ob Studien auf Basis repräsentativer Stichproben niedrigere Effekte liefern, als Studien aus Spezialpopulationen (E6.5). Die Variable Studienqualität überprüft, ob subjektiv

schlechter bewertete Studien extremere Werte liefern (E6.6). Das Publikationsjahr wird ebenfalls überprüft, um herauszufinden, ob aufgrund des langen Erhebungszeitraums ein Periodeneffekt vorliegt (E6.7).

Bezüglich der methodischen Fragestellungen lassen sich ebenfalls Vermutungen formulieren. Beim Vergleich der beiden Methoden zur Untersuchung von Moderatorvariablen kann angenommen werden, dass die Teststärke im OSMASEM-Ansatz höher ist, als beim Subgruppen-TSSEM-Ansatz, da die metrischen Variablen nicht kategorisiert werden und somit kein Informationsverlust entsteht. Wahrscheinlich wird es im Subgruppen-Design zu Konvergenzproblemen kommen (E7.1), da viele fehlende Werte im Datensatz vorliegen sowie ein Random-Effects-Modell vermutet wird und dies zu Problemen führt (Jak/Cheung 2018a, 2018b). Vor allem diejenigen Beziehungen, die auf nur wenigen Effektstärken basieren, sollten dabei zu Problemen führen. Eine Reduzierung des Modells auf weniger Beziehungen erscheint dabei wahrscheinlich.

Die unterschiedlichen Methoden zur Berücksichtigung abhängiger Effektstärken sind schwer zu bewerten, da bislang nur spärliche Erkenntnisse über deren Auswirkungen vorliegen. Es ist anzunehmen, dass die komplexeren Verfahren (Wilson-Ansatz und robuster Ansatz) eine höhere Präzision bei der Schätzung (E7.2) haben (López-López et al. 2018; Park/Beretvas 2018). Der Vorteil des robusten Ansatzes, dass gleichzeitig mehrere Formen von Abhängigkeiten berücksichtigt werden können, spielt in der vorliegenden Arbeit keine Rolle, so dass nur ein geringer Unterschied zwischen dem Wilson- und dem robusten Ansatz erwartet wird.



## 9 Durchführung der Meta-Analyse und Ergebnisdarstellung

Dieses Kapitel dokumentiert die Ergebnisse der modellbasierten Meta-Analyse zur Frage des Zusammenhangs von Religiosität und Lebenszufriedenheit und den für die Beziehung verantwortlichen Erklärungsmechanismen. Neben der engeren Analyse der inhaltlichen Fragen wird die kulturelle Bedingtheit des Erklärungsmodells mittels entsprechenden Makrovariablen überprüft. Weitere methodische Moderatoren werden ebenfalls getestet. Der methodische Fokus der Analyse liegt auf dem Vergleich dreier unterschiedlicher Methoden zum Umgang mit abhängigen Effektstärken und dem Vergleich zweier unterschiedlicher Methoden zur Untersuchung von Moderatorvariablen.

Zunächst wird nun die programmtechnische Umsetzung der statistischen Analyse in R beschrieben. Daran anknüpfend folgt ein Gesamtüberblick über die einzelnen Modellierungsschritte. Im Anschluss daran folgen die Analyseergebnisse.

### 9.1 Programmtechnische Umsetzung

Die vollständige Umsetzung der statistischen Analyse wird mithilfe der freien Statistikumgebung R vorgenommen (R Core Team 2018). Bei R handelt es sich ursprünglich um eine Programmiersprache, die auf der Sprache S basiert und eine Vielzahl von statistischen Analysen ermöglicht. Ergänzt wird die Programmiersprache um eine graphische Oberfläche beziehungsweise eine Entwicklungsumgebung mit der Software RStudio (RStudio Team 2015).

Die Grundfunktionalität von R kann mit Paketen erweitert werden. Zur Umsetzung einer TSSEM-Analyse sind die Pakete metaSEM (Cheung 2014b) und OpenMx notwendig (Neale et al. 2016). Das metaSEM Paket ermöglicht die Durchführung SEM-basierter Meta-Analysen und MASEM-Analysen. Die meta-analytischen Strukturgleichungsmodelle werden von metaSEM in OpenMx umgesetzt. Die Modellspezifizierung des metaSEM-Pakets basiert dabei auf der sogenannten RAM-Notation (Cheung 2015: 15ff.). Das Reticular Action Model (RAM) unterscheidet sich von den sonst üblichen LISREL-, EQS- und *Mplus*-Notationen (Little/Kline 2016; McArdle/McDonald 1984). Im RAM-Ansatz werden Strukturgleichungsmodelle mithilfe von drei Matrizen ( $A$ ,  $S$ ,  $F$ ) und einem Vektor  $M$  ausgedrückt. In der asymmetrischen  $A$ -Matrix werden Regressionskoeffizienten und Faktorladungen spezifiziert. In der symmetrischen  $S$ -Matrix werden hingegen die Varianzen und Kovarianzen des Modells festgelegt. Bei abhängigen Variablen entsprechen sie den Residuen und bei unabhängigen Variablen den Varianzen. Die  $F$ -Matrix gibt Auskunft darüber, ob es sich um manifeste oder latente Variablen handelt

und der M-Vektor repräsentiert die Mittelwerte (Intercepts) der Variablen. In der neusten Version von metaSEM können auch Modelle mittels der lavaan-Syntax spezifiziert und anschließend in die RAM-Formulierung umgewandelt werden.<sup>120</sup>

Zur Umsetzung des von Wilson/Polanin/Lipsey (2016) vorgeschlagenen Ansatzes zur Berücksichtigung abhängiger Effektstärken mittels Mehrebenenmodellierung und der robusten Varianzschätzung (Hedges/Tipton/Johnson 2010) wird das Paket metafor benötigt (Viechtbauer 2010). Bei metafor handelt es sich um ein R-Paket zur Durchführung und Visualisierung von herkömmlichen Meta-Analysen. Bei den beiden genannten Methoden zur Berücksichtigung abhängiger Effektstärken wird jeweils die erste Phase der TSSEM-Analyse, also die Kumulierung der Effektstärken, in metafor durchgeführt. Anschließend werden die Ergebnisse dieses Analyseschritts an das metaSEM Paket übergeben und dort die zweite Phase der Analyse ausgeführt. Für die Darstellung der Meta-Analyse werden eine Reihe weiterer R-Pakete genutzt. Zur automatisierten Ausspeicherung der Korrelationstabellen der ersten TSSEM-Phase wird das R-Paket sjPlot verwendet (Daniel Lüdtke 2018). Die Ergebnistabellen der zweiten Phase werden mit dem Paket texreg ausgespeichert (Leifeld 2013).<sup>121</sup> Die Darstellung der meta-analytischen Strukturgleichungsmodelle in der RAM-Notation werden mithilfe des Pakets semPlot vorgenommen (Epskamp 2015). Die vollständige Session-Info der Analysen mit allen Programm- und Paketversionen findet sich in Anhang D. Im Online-Anhang (<https://osf.io/29rqv>) befinden sich darüber hinaus alle Datensätze und Syntaxdateien zur Reproduktion der nachfolgenden Analysen.

---

<sup>120</sup> Dies erleichtert die Zugänglichkeit des metaSEM-Pakets für viele Anwender und wird in dieser Arbeit zur Umsetzung des OSMASEM-Ansatzes zur Untersuchung von Moderatoren angewandt.

<sup>121</sup> Für die Ausspeicherung musste dazu jedoch eigens eine neue Funktion geschrieben werden, die es ermöglicht, die Ergebnisobjekte von metaSEM zu übernehmen.

## 9.2 Übersicht über die statistischen Modellierungsschritte

Das Ausgangsmodell besteht aus der unabhängigen Variable „Religiosität“ und der abhängigen Variable „Lebenszufriedenheit“. Dazwischen befinden sich die drei Mediatorvariablen „soziale Unterstützung“, „Optimismus“ und „Selbstwert“.<sup>122</sup> Neben diesem Grundmodell werden eine Reihe von Moderatoranalysen mittels zweier unterschiedlicher Verfahren durchgeführt sowie drei unterschiedliche Methoden zur Berücksichtigung abhängiger Effektstärken angewandt.

In der nachfolgenden Tabelle 29 ist die Korrelationsmatrix, wie sie dem Modell zugrunde liegt, dargestellt. Die abgetragene Anzahl der Korrelationen bezieht sich dabei auf die Gesamtheit der aus den Studien extrahierten Effektstärken und entspricht somit dem Wilson et al. Ansatz. Ein deutliches Ungleichgewicht zwischen den Zellen ist zu erkennen. Während die Korrelationen der Hauptuntersuchungsvariablen R/S und Lebenszufriedenheit sehr häufig untersucht wird ( $n = 219$ ), sind die anderen Zellen deutlich schwächer besetzt.

**Tabelle 29: Korrelationsmatrix**

	Religiosität	Soz. Unterstützung (M1)	Optimismus (M4)	Selbstwert (M7)	Lebenszufriedenheit
Religiosität	1	Zelle 1	Zelle 2	Zelle 3	Zelle 4
Soz. Unterstützung (M1)	53	1	Zelle 5	Zelle 6	Zelle 7
Optimismus (M4)	41	6	1	Zelle 8	Zelle 9
Selbstwert (M7)	60	2	8	1	Zelle 10
Lebenszufriedenheit	219	31	15	21	1

**Anmerkungen:** Über der Diagonale ist die Zellenzuordnung und unterhalb der Diagonale ist die Anzahl der berichteten Effektstärken für die jeweilige Beziehung zu sehen.

Besonders die Korrelationen zwischen den Mediatoren untereinander (Kovarianzen) liegen nur selten vor. Die Korrelationen zwischen den Mediatorvariablen und (un-)abhängiger Variable sind im Vergleich häufiger verfügbar.

Vor allem bei den Zellen mit vielen fehlenden Werten, also den Zellen 5, 6 und 8 (Kovarianzen zwischen den Mediatoren), aber auch 2, 9 und 10 können Probleme bei der Schätzung auftreten. Besonders bei der Subgruppen-TSSEM-Moderatoranalyse ist dies vermutlich problematisch. Dies liegt daran, dass durch den Median-Split vermehrt leere Zellen entstehen können und Modelle dann nicht mehr konvergieren. Deshalb wird schrittweise das Modell reduziert, um Datensätze mit weniger fehlenden Werten zu schaffen und die Gefahr leerer Zellen zu reduzieren. Dazu wird zunächst Optimismus, dann Selbstwert und zum Schluss beide Variablen aus dem

<sup>122</sup> Aus Gestaltungsgründen wird im Folgenden zumeist auf die Setzung von Anführungszeichen bei den Variablen verzichtet.

Datensatz entfernt. Dementsprechend werden diejenigen Studien ausgeschlossen, die nur Beziehungen zwischen der jeweiligen Variablen und der abhängigen oder der unabhängigen Variable untersucht haben, so dass die Modelle jeweils auch auf einer anderen Studienanzahl basieren. Da die Makrovariable „Wichtigkeit von Religion“ nicht für alle Studien vorliegt, basiert diese Analyse auf einer anderen Studienzusammensetzung als die Untersuchung des HDI-Moderators, der mit dem vollständigen Datensatz durchgeführt wird.

In den genannten Modellen wird sowohl für den HDI-Datensatz als auch für den WR-Datensatz getestet, ob eine Moderatoranalyse möglich ist. Es zeigt sich in beiden Datensätzen, dass jeweils nur das Modell mit nur einer Mediatorvariable (soziale Unterstützung) konvergiert. Alle anderen Modelle haben in der ersten oder zweiten Analysephase Konvergenzprobleme.<sup>123</sup> Der Wilson-Ansatz sowie der robuste Ansatz zur Berücksichtigung abhängiger Effektstärken wird nur am Untersuchungsmodell ohne Moderatoren getestet. Dies hängt damit zusammen, dass die R-Funktion zur Subgruppenanalyse in metaSEM nur zur Überprüfung von in metaSEM kumulierten Effektstärken eingesetzt werden können. In Tabelle 30 ist die Übersicht der jeweils ausgeschlossenen Studien anhand der K-IDs zu sehen. Im WR-Datensatz werden bereits vor Beginn der Modellierung neun Studien ausgeschlossen, da für diese Studien die Makrovariable „Wichtigkeit von Religion“ nicht vorliegt.

**Tabelle 30: Ausgeschlossene Studien (K-IDs)**

	<b>Komplett</b>	<b>Ohne Optimismus</b>	<b>Ohne Selbstwert</b>	<b>Ohne Optimismus und Selbstwert</b>
<b>HDI-DS</b>	-	29, 30, 44, 46, 110, 132, 172	53, 60, 90, 91, 104, 118, 120, 137, 138, 139, 140, 164	29, 30, 44, 46, 49, 53, 59, 60, 67, 104, 109, 110, 118, 120, 123, 124, 125, 132, 137, 138, 139, 140, 164, 172
<b>WR-DS</b>	71, 72, 84, 89, 117, 120, 125, 130, 169	29, 30, 44, 46, 110, 132, 172	53, 60, 90, 91, 104, 118, 137, 138, 139, 140, 164	29, 30, 44, 46, 49, 53, 59, 60, 67, 104, 109, 110, 118, 120, 123, 124, 125, 132, 137, 138, 139, 140, 164, 172

In Tabelle 31 ist eine Übersicht über die jeweilige Fallzahl und Stichprobenanzahl über die Modellierungsschritte hinweg zu sehen. Im vollständigen HDI-Datensatz (HDI-DS) liegen 176 Stichproben mit 205.483 Befragten vor.

<sup>123</sup> Konvergenz bedeutet hierbei, dass das Verfahren eine Lösung berechnen kann und OpenMx zusätzlich den Status 1 anzeigt (Cheung 2015: 165). Bei allen Modellen, wo dies nicht zutrifft, wird im nächsten Schritt mittels „rerun“-Befehl versucht, doch noch zu einer gültigen Lösung zu kommen.

**Tabelle 31: Stichproben- und Fallzahl**

		Komplettes Modell	Ohne Optimismus	Ohne Selbstwert	Ohne Optimismus/Selbstwert	Mit Moderator
Fallzahl (N)	HDI-DS	205.483	203.517	180.333	177.088	Low: 34.819 High: 142.269
	WR-DS	203.317	200.689	180.450	172.746	Low: 49.374 High: 123.372
Stichprobenanzahl (k)	HDI-DS	176	170	164	152	Low: 72 High: 80
	WR-DS	167	159	158	142	Low: 56 High: 86

Der WR-Datensatz ist mit 167 Stichproben und 203.317 Befragten etwas kleiner. Durch Ausschluss weiterer Stichproben, aufgrund der Reduzierung des Untersuchungsmodells, verringert sich die Gesamtfallzahl und die Studienanzahl. Da die Subgruppen-Moderatoranalyse aufgrund von Konvergenzproblemen erst im reduzierten Modell, also ohne die Variablen Optimismus und Selbstwert, möglich ist, sind Fallzahl und Stichprobenanzahl nur für das verringerte Untersuchungsmodell in der rechten Spalte ausgewiesen. Der Median-Split der beiden Moderationsvariablen ergibt in Bezug auf die Stichprobenanzahl etwa gleich gewichtete Gruppen. Im Hinblick auf die zugrundeliegende Fallzahl gibt es jedoch ein deutliches Übergewicht bei den High-Gruppen<sup>124</sup>. In der nachfolgenden Tabelle 32 ist abschließend die Übersicht über die Anzahl der Effektstärken, die in den jeweiligen Analysen eingeschlossen sind, zu sehen.

**Tabelle 32: Anzahl der Effektstärken**

		Gesamtmodell	Ohne Optimismus	Ohne Selbstwert	Ohne Optimismus und Selbstwert	Mit Moderator
Mittelwert-TSSEM	HDI-DS	293	243	230	187	Low: 79 High: 108
	WR-DS	281	231	222	180	Low: 64 High: 116
Wilson-Ansatz/ robuster Ansatz	HDI-DS	456	386	365	303	-
	WR-DS	431	361	348	286	-

<sup>124</sup> Bei den beiden Kulturproxymvariablen auf Makroebene wird jeweils die Gruppe unterhalb des Medians Low-Gruppe und Gruppe überhalb des Medians High-Gruppe genannt. Siehe dazu auch Tabelle 40.

Diese werden nach (Mittelwert-)TSSEM und Wilson- bzw. robuster Ansatz getrennt ausgewiesen. Der vollständige HDI-Datensatz basiert auf 456 und der WR-Datensatz auf 431 Effektstärken. Im Mittelwert-TSSEM-Ansatz sind diese Effektstärken bereits zusammengefasst, so dass dort im vollständigen Untersuchungsmodell 293 bzw. 281 Effektstärken vorliegen. Im Endmodell (ohne Optimismus und Selbstwert) sind beim HDI-Datensatz noch 187 und beim WR-Datensatz 180 Effektstärken übrig. Beim Wilson-Ansatz hingegen befinden sich am Ende der Modellreduktion 303 bzw. 286 Effektstärken in der Untersuchung. Die Verteilung der Effektstärken in Bezug auf den Mediansplit ist ebenfalls verzerrt, so dass sich in den High-Gruppen mehr Effektstärken befinden als in den Low-Gruppen. Dies gilt für beide Moderatoren gleichermaßen.

In Tabelle 33 ist eine deskriptive Übersicht über die zehn im Untersuchungsmodell befindlichen Effektstärken zu sehen. Alle Effektstärken zeichnen sich durch eine ausgedehnte Spannweite aus. Teils reicht diese von stark negativen bis zu sehr stark positiven Zusammenhängen. Dabei wird zum einen die Notwendigkeit eines Random-Effects-Modells aufgrund der Spannweite der Effektstärken deutlich und zum anderen die große Anzahl von fehlenden Werten. Diese hohe Anzahl an Missings ist verantwortlich für die beschriebenen Konvergenzprobleme.

**Tabelle 33: Deskriptive Übersicht über die Effektstärken**

	Mittelwert	Median	Min.	Max.	NAs
Zelle 1	.135	.119	-.211	.580	142
Zelle 2	.153	.170	-.090	.360	155
Zelle 3	.152	.145	-.120	.495	144
Zelle 4	.207	.200	-.235	.481	51
Zelle 5	.398	.385	.300	.480	170
Zelle 6	.195	.195	.050	.340	174
Zelle 7	.335	.340	.060	.610	147
Zelle 8	.284	.281	.100	.570	168
Zelle 9	.377	.390	.059	.650	161
Zelle 10	.386	.480	-.440	.618	155

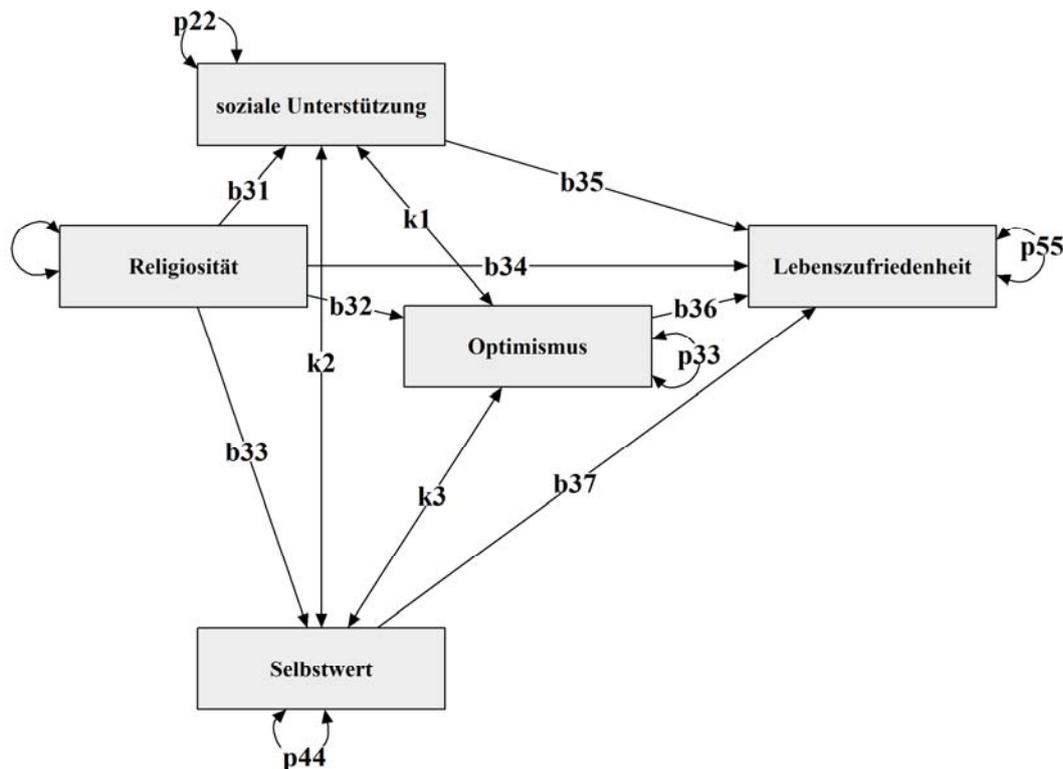
**Anmerkung:** Bezieht sich auf den vollständigen Mittelwert-TSSEM-Datensatz.

Zur Illustration der Verteilung der z-transformierten Effektstärken, in Bezug auf die jeweilige Beziehung, sind die zugehörigen Forest-Plots in Anhang E zu sehen. Diese Forest-Plots basieren auf bivariaten Random-Effects-Schätzungen, die im Anschluss an die Durchführung des TSSEM-Verfahrens auch für die Analyse des Publikationsbias genutzt werden. Die Forest-Plots sind nach den z-transformierten Effektstärkekoeffizienten für die jeweiligen Studien sortiert (Caterpillar-Plot), um einen Überblick über die Verteilung des Effekts zu erhalten. Betrachtet man diese Abbildungen, wird die Variabilität der Effektstärken deutlich. Im Einklang mit den analytischen Überlegungen sollte deshalb ein Random-Effects-Modell spezifiziert werden.

### 9.3 Analyseergebnisse Mittelwert-TSSEM-Ansatz

In der nachfolgenden Abbildung 22 ist das abgeleitete Gesamtmodell dargestellt. Es sind alle zu schätzenden Koeffizienten zu sehen, wie sie in der RAM-Notation in OpenMx festgelegt sind. Das Modell wird entsprechend den Vorgaben von Cheung (2015) und Jak (2015) konstruiert. Nur die A- und S-Matrizen müssen für das vorliegende Untersuchungsmodell spezifiziert werden, da es sich dabei um ein Pfadmodell mit ausschließlich manifesten Variablen handelt und keine Mittelwertstruktur notwendig ist. Im vorliegenden Fall werden zur Schätzung aller Modelle die Startwerte für Regressionskoeffizienten (b31 – b37) auf 0.1 und für Varianzen bzw. Residuen (p22, p33, p44, p55) auf 1.0 gesetzt.<sup>125</sup> Die Varianz der unabhängigen Variable Religiosität muss für die Schätzung des Modells auf 1.0 gesetzt werden (Jak/Cheung 2018a: 10). Der Modellfit wird in Anlehnung an die methodologische Praxis nur bei überidentifizierten Modellen beurteilt. Auf eine Fixierung spezifischer Effekte zur Erzeugung eines überidentifizierten Modells wird verzichtet (Jak/Cheung 2018b).

Abbildung 22: Gesamtmodell



Anmerkung: Eigene Abbildung.

<sup>125</sup> Eine Variation der Startwerte wird zum Sensitivitätstest durchgeführt. Eine Veränderung der Startwerte der Pfadkoeffizienten auf .2 und Varianzen auf 2 zeigt dabei keinerlei Auswirkungen auf die Schätzergebnisse. Dies spricht für eine robuste Schätzung in Bezug auf die Startwerte.

Die Kovarianzen zwischen den Mediatorvariablen werden im Grundmodell, in Anlehnung an die analytischen Überlegungen in Kapitel 3.4, ebenfalls spezifiziert ( $k_1, k_2, k_3$ ). Da das vorliegende Modell Mediatoren enthält, können mit TSSEM in metaSEM keine Standardfehler und damit keine auf der Wald-Statistik beruhenden Konfidenzintervalle berechnet werden. Als Ersatz dienen jedoch die sogenannten Likelihood-based-Konfidenzintervalle (siehe Ausführungen in Kapitel 5.4). Da deskriptiv bereits die Heterogenität der Effektstärken deutlich wird (siehe Tabelle 33 und Anhang E) und analytisch ebenfalls ein Random-Effects-Modell plausibel ist, wird mit der Schätzung eines Random-Effects-Modells begonnen. Dies hat den Vorteil, dass fehlende Werte unproblematisch sind.

In der nachfolgenden Tabelle 34 ist das Ergebnis der ersten Phase der Schätzung zu sehen. Das Gesamtmodell basiert auf 176 unabhängigen Stichproben, 293 Effektstärken und 205.483 Befragten. Über der Diagonale befinden sich die kumulierten Korrelationskoeffizienten. Der kumulierte Korrelationskoeffizient der Hauptbeziehung zwischen Religiosität und Lebenszufriedenheit (.204) ist im Vergleich zu den bisherigen Meta-Studien etwas stärker (siehe Kapitel 3.3). Unterhalb der Diagonale ist zum einen die Anzahl der Stichproben, auf denen der jeweilige Korrelationskoeffizient basiert, zu sehen und zum anderen in Klammern der  $I^2$ -Wert zu sehen.

**Tabelle 34: (Random-Effects-) Ergebnisse erste Phase Mittelwert-TSSEM-Ansatz**

	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		0.128	0.151	0.149	0.204
Soz. Unterstützung	34 (0.894)		0.384	0.180	0.332
Optimismus	21 (0.856)	6 (0.000)		0.288	0.377
Selbstwert	32 (0.927)	2 (0.940)	8 (0.949)		0.391
Lebenszufriedenheit	125 (0.9216)	29 (0.945)	15 (0.960)	21 (0.982)	

**Anmerkung:**  $n = 205.483$ ;  $k = 176$  Stichprobe; 293 Effektstärken. Oberhalb der Diagonale sind die kumulierten Korrelationskoeffizienten und unterhalb die Anzahl der Effektstärken für die jeweilige Beziehung sowie die  $I^2$ -Werte zu sehen.

Die kumulierten Korrelationskoeffizienten zwischen allen Variablen können für alle Beziehungen als moderat bis stark bezeichnet werden (Gignac/Szodorai 2016). Entsprechend der Daumenregel ( $I^2$ -Werte  $> 75\%$ ) ist die Variabilität der Effektstärken zwischen den Studien sehr hoch, da sich die  $I^2$ -Werte zwischen .856 und .982 bewegen. Lediglich die Beziehung zwischen den Variablen soziale Unterstützung und Optimismus (Zelle 5) ist homogen, was

damit zusammenhängt, dass diese Beziehung auf nur wenigen Studien basiert. Der formale Homogenitätstest zeigt, dass keine Homogenität der Korrelationsmatrix vorliegt, da die Nullhypothese mit  $Q(df = 283) = 2082$  mit  $.0001, p < 0.00$  verworfen wird. Es besteht somit ein erhebliches Ausmaß an Heterogenität in der kumulierten Korrelationsmatrix. In Tabelle 35 sind für das vollständige Modell die Fallzahlen nach Beziehung geordnet zu sehen.

**Tabelle 35: Fallzahlen nach Beziehungen**

	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		16.974	5.314	31.860	71.863
Soz. Unterstützung			1.534	1.173	105.689
Optimismus				3.086	4.420
Selbstwert					9.343
Lebenszufriedenheit					

Vor allem die Kovarianzen zwischen den Mediatoren beruhen auf nur wenigen Studien. Die Parameterschätzungen für Beziehungen mit nur wenigen Effektstärken sollten deshalb zurückhaltend interpretiert werden.

Die Schätzung eines Random-Effects-Modells in der zweiten Phase des TSSEM-Verfahrens ist somit aufgrund analytischer Überlegungen und der deskriptiven Einschätzung sowie der empirischen Resultate notwendig. In Tabelle 36 ist das Schätzergebnis der zweiten Phase auf Basis der kumulierten Random-Effects-Korrelationsmatrix des vollständigen TSSEM-Modells sowohl mit Kovarianz  $k_2$ , also der Beziehung zwischen den Variablen soziale Unterstützung und Selbstwert, als auch ohne zu sehen. Ebenfalls sind dort die geschätzten Pfadkoeffizienten ( $r$ ), Kovarianzen ( $k$ ), Residuen ( $p$ ) und die indirekten Effekte ( $ind$ ) zu sehen. Alle Schätzungen sind mit ihrem zugehörigen Likelihood-Konfidenzintervall (95%igen LBCI) versehen.

Im unteren Teil der Tabelle finden sich die Fit-Indizes und die weiteren relevanten Modellangaben. Im vollständigen Modell können aufgrund der fehlenden Freiheitsgrade, da es sich um ein genau identifiziertes Modell handelt, keine Fit-Indizes berechnet werden. Die Konfidenzintervalle (95%ige LBCIs) der geschätzten Parameter sind vor allem bei den Pfadkoeffizienten und den Kovarianzen sehr breit, was die Heterogenität der zugrundeliegenden Studien, etwa durch die Unterschiedlichkeit der verwendeten Messinstrumente, und grundlegende Unterschiede bei den jeweiligen Grundgesamtheiten unterstreicht. Deutlich wird auch, dass für die Effekte zwischen Religiosität und den Mediatoren ( $b_{31}$ ,  $b_{32}$  und  $b_{33}$ ) die Konfidenzintervalle der Schätzung deutlich schmaler sind als zwischen den Mediatoren und Lebenszufriedenheit

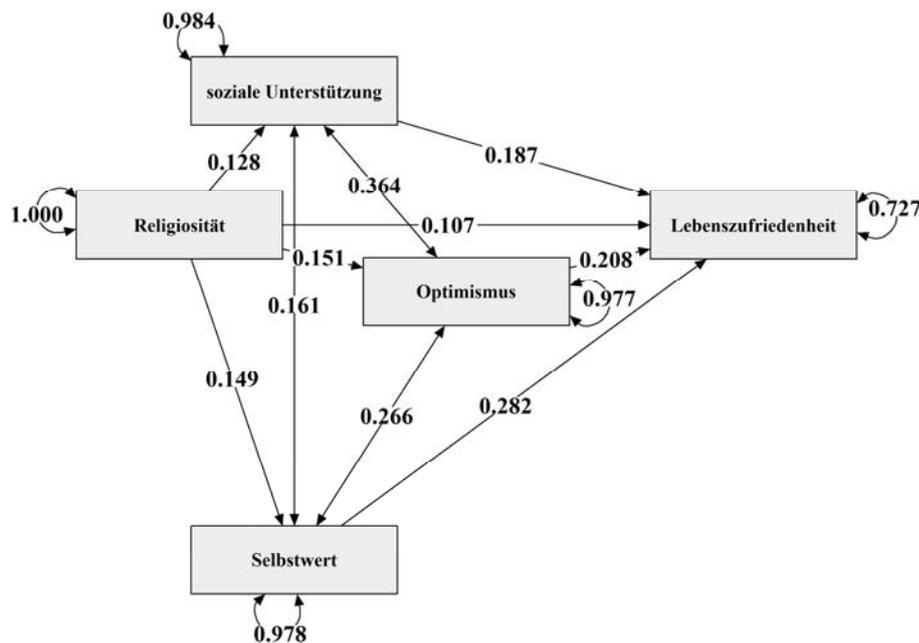
(b35, b36 und b37). Die wenig präzise Schätzung hängt sicherlich auch mit der teils sehr niedrigen Anzahl von Effektstärken für die entsprechenden Beziehungen zusammen. Das LBCI des Haupteffekts (b34) ist hingegen vergleichsweise schmal (.0704 - .1407).

**Tabelle 36: Zweite Phase Mittelwert-TSSEM-Ansatz (Gesamtmodell mit k2 und ohne k2)**

	Mit k2	Ohne k2
b31	0.1277 [0.0882; 0.1672]	0.1287 [0.0892; 0.1681]
b32	0.1513 [0.1065; 0.1961]	0.1513 [0.1065; 0.1961]
b33	0.1485 [0.1032; 0.1938]	0.1496 [0.1043; 0.1949]
b34	0.1071 [0.0704; 0.1407]	0.0982 [0.0614; 0.1324]
b35	0.1874 [0.0822; 0.2948]	0.2454 [0.1621; 0.3289]
b36	0.2077 [0.0721; 0.3322]	0.1755 [0.0367; 0.3046]
b37	0.2817 [0.1504; 0.4113]	0.3210 [0.1934; 0.4468]
p22	0.9837 [0.9721; 0.9922]	0.9834 [0.9717; 0.9920]
k1	0.3643 [0.3094; 0.4192]	0.3642 [0.3093; 0.4191]
k2	0.1614 [-0.0371; 0.3598]	-
p33	0.9771 [0.9616; 0.9887]	0.9771 [0.9616; 0.9887]
k3	0.2659 [0.1572; 0.3746]	0.2658 [0.1571; 0.3745]
p44	0.9779 [0.9624; 0.9893]	0.9776 [0.9620; 0.9891]
p55	0.7274 [0.6376; 0.7967]	0.7069 [0.6144; 0.7804]
Ind1 (b31*b35)	0.0239 [0.0102; 0.0409]	0.0316 [0.0193; 0.0473]
Ind2 (b32*b36)	0.0314 [0.0108; 0.0546]	0.0263 [0.0051; 0.0495]
Ind3 (b33*b37)	0.0418 [0.0213; 0.0674]	0.0469 [0.0260; 0.0728]
Chi-Quadrat-Veränderung	-	2.5397
DF	0	1.0000
p	-	0.1110
RMSEA	-	0.0027
RMSEA unt. 95%iges KI	-	0.0000
RMSEA ober. 95%iges KI	-	0.0071
CFI	-	0.9983
SRMR	-	0.0509
Fallzahl/Effektstärken	205.483/293	205.483/293

In Abbildung 23 sind zur besseren Übersichtlichkeit die geschätzten Parameter des Pfadmodells mit k2 abgetragen. Der direkte Pfad von Religiosität auf Lebenszufriedenheit ist eher schwach (b34 = .107). Auch die Effekte der unabhängigen Variable auf die Mediatorvariablen soziale Unterstützung (b31 = .128), Optimismus (b32 = .151) und Selbstwert (b33 = .149) sind relativ schwach. Deutlich stärker sind die Effekte der Mediatorvariablen soziale Unterstützung (b35 = .187), Optimismus (b36 = .208) und Selbstwert (b37 = .282) auf die unabhängige Variable. Die Kovarianzen zwischen den Mediatorvariablen betragen k1 = .364, k2 = .161 und k3 = .266.

Abbildung 23: Schätzergebnis des Gesamtmodells mit k2

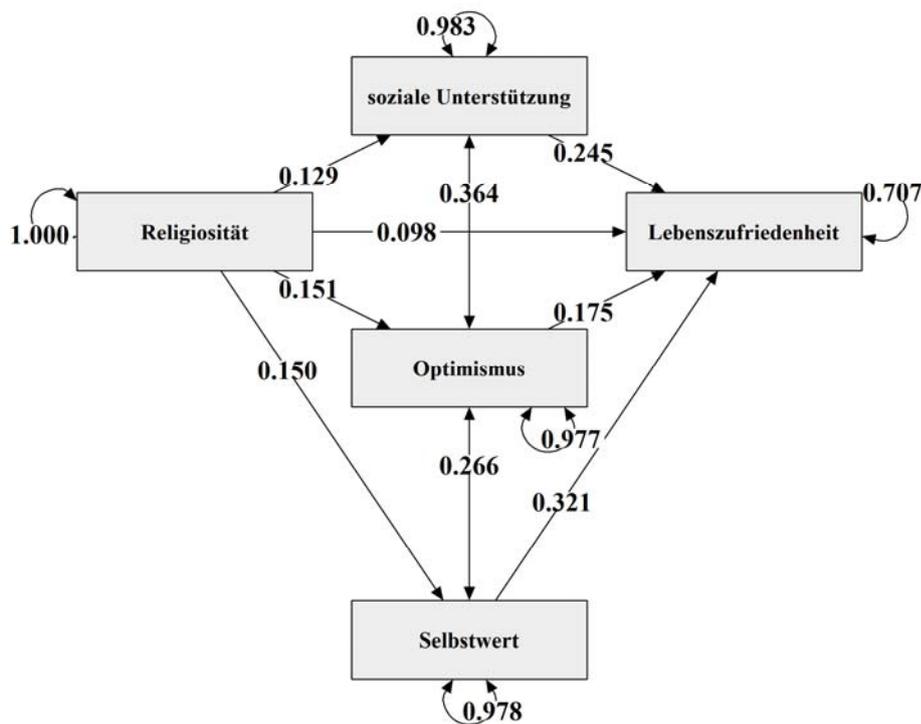


**Anmerkung:** Eigene Abbildung; n = 205.483.

Auch die drei indirekten Effekte von Religiosität und den Mediatoren werden bei der Modellschätzung berücksichtigt. Der indirekte Effekt und sein 95%iges LBCI von Religiosität über soziale Unterstützung auf Lebenszufriedenheit beträgt .024 (.010 - .040), der indirekte Effekt von Religiosität über Optimismus auf Lebenszufriedenheit beträgt .031 (.011 - .055), und der indirekte Effekt von Religiosität über Selbstwert beträgt .042 (.021 - .064). Da in keinem der 95%igen LBCIs der Wert „0“ eingeschlossen ist, können alle drei indirekten Effekte als signifikant betrachtet werden. Dies bedeutet, dass es sich wie vermutet um partielle Mediatoren handelt. Das Modell kann 27,4 % der Varianz von Lebenszufriedenheit statistisch erklären. Dies ist durchaus beachtenswert, wenn man die vielfältigen Faktoren, die einen Einfluss auf Lebenszufriedenheit haben, in Betracht zieht.

Lediglich das LBCI der Kovarianz k2 schließt den Wert „0“ mit ein, so dass das Modell abermals ohne die Kovarianz geschätzt wird (siehe Abbildung 24). Dies hat den Vorteil, dass ein Freiheitsgrad gewonnen wird und somit Fit-Indizes berechnet werden können. Der Likelihood-Ratio-Test (2.5397) zeigt dabei keine signifikante Verschlechterung des Modells an.

Abbildung 24: Schätzergebnis des Gesamtmodells ohne k2



**Anmerkung:** Eigene Abbildung; n = 205.483.

Die Fit-Indizes zeigen eine sehr gute Passung des Modells an, da CFI (.9983), RMSEA (.0027) sowie dessen Konfidenzintervall (.0000 - .0071) und SRMR (.0509) sehr gute Werte aufweisen. Im Vergleich zum Modell mit k2 gibt es leichte Veränderungen der geschätzten Pfadkoeffizienten. Der direkte Pfad zwischen Religiosität und Lebenszufriedenheit (b34) ist kleiner, während die Pfade zwischen den beiden Mediatoren soziale Unterstützung (b35) sowie Selbstwert (b37) und Lebenszufriedenheit gestiegen sind. Lediglich der Pfad von Optimismus zu Lebenszufriedenheit (b36) ist geringer geworden.

Auch in dem Modell ohne k2 werden die indirekten Effekte getestet. Hierbei lassen sich leichte Veränderungen feststellen. Der indirekte Effekt und sein 95%iges LBCI von Religiosität über soziale Unterstützung auf Lebenszufriedenheit beträgt .032 (.019 - .047), der indirekte Effekt von Religiosität über Optimismus auf Lebenszufriedenheit beträgt .026 (.006 - .050) und der indirekte Effekt von Religiosität über Selbstwert beträgt .048 (.027 - .074). Die indirekten Effekte über soziale Unterstützung und Selbstwert nehmen leicht zu und der indirekte Effekt über Optimismus wird geringer. Alle LCBI schließen den Wert „0“ nicht ein, so dass auch hier partielle Mediatoren vorliegen. Die Erklärungsleistung dieses Modells ist im Vergleich zum Modell mit k2 leicht angestiegen. 29,3 % der Varianz von Lebenszufriedenheit lassen sich mit dem zugrunde gelegten Modell statistisch erklären.

Die schwachen Pfade zwischen den Variablen sind kaum überraschend, wenn man die Natur der abhängigen Variable Lebenszufriedenheit und die Vielzahl von Prädiktoren für dieses soziale Konstrukt bedenkt. In Kapitel 3.4 wird die Vielzahl der gefundenen Erklärungsmechanismen (Mediatoren) angesprochen. Bei den drei hier ins Untersuchungsmodell eingeschlossenen Variablen handelt es sich nur um eine Auswahl, der in der Literatur diskutierten Mediatoren. Was jedoch bereits in diesem Modell deutlich wird ist, dass bei Berücksichtigung von Mediatoren der direkte Pfad von Religiosität auf Lebenszufriedenheit kleiner wird. Dies spricht für die dieser Arbeit zugrundeliegenden These, dass die vielfach empirisch gezeigten positiven Effekte von Religiosität auf das Leben von Individuen vorwiegend auf spezifische Erklärungsmechanismen zurückzuführen sind. Die individuelle Wirksamkeit von Religiosität liegt in psychosozialen Wirkketten. Es kann vermutet werden, dass sich der direkte Effekt zwischen Religiosität und Lebenszufriedenheit bei Berücksichtigung weiterer relevanter Mediatoren (siehe Kapitel 3.4) weiter verringert.

Bezüglich der Methodik des TSSEM-Verfahrens bleibt festzuhalten, dass die explorative Anwendung gut funktioniert. Das konstruierte Untersuchungsmodell kann als Basis weiterer Untersuchungen des Zusammenhangs in der Zukunft dienen. Trotz sehr unterschiedlicher eingeschlossener Studien lassen sich inhaltlich gehaltvolle Schlüsse aus der modellbasierten Meta-Analyse ableiten, die für eine Vielzahl von Forschungsbereichen relevant sein können. Somit bleibt zu vermuten, dass das Verfahren in einer Vielzahl von wissenschaftlichen Bereichen zur meta-analytischen Zusammenfassung von komplexen Modellen genutzt werden kann, auch wenn bislang kein etabliertes Erklärungsmodell vorliegt. Die Sensitivität der Analyse wird in den folgenden Kapiteln überprüft.

## 9.4 Untersuchung der Moderatorvariablen

Zur Untersuchung der Kulturabhängigkeit des Erklärungsmodells und zur Exemplifizierung der Möglichkeiten zur Durchführung von Moderatoranalysen (Jak/Cheung 2018a, 2018b) wird nun untersucht, ob sich die Regressionskoeffizienten in Studien aus Ländern mit hohem bzw. niedrigem Human Development Index (HDI) unterscheiden. Die zweite Kulturmoderatorvariable, die untersucht wird, ist die aggregierte Makrovariable „Wichtigkeit von Religion“ aus dem World Value Survey. Darüber hinaus werden ausgewählte methodische Moderatoren auf ihre Wirksamkeit hin überprüft. Dies wird nun mittels zweier unterschiedlicher Verfahren der Reihe nach überprüft. Zuerst folgen die OSMASEM-Analysen (siehe 9.4.1), bevor anschließend die Subgruppenanalyse (9.4.2) durchgeführt wird.

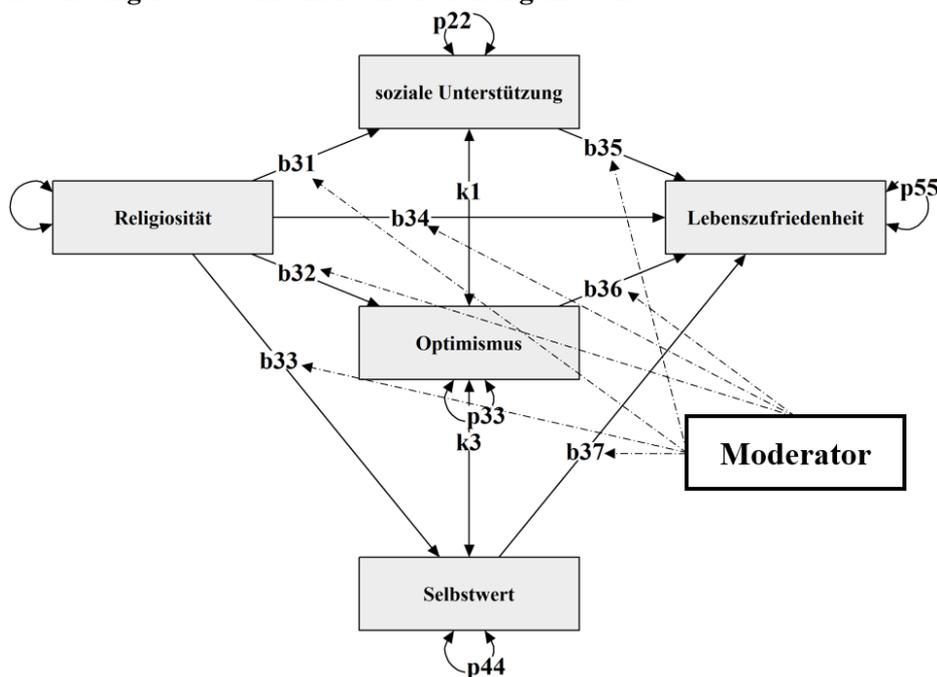
Dazu wird zunächst der OSMASEM-Ansatz auf das vollständige Untersuchungsmodell mit drei Mediatoren und ohne die Kovarianz  $k_2$  angewandt. Dabei wird getestet, welchen Einfluss die Moderatorvariable auf die Regressionskoeffizienten ( $b_{31} - b_{37}$ ) des Untersuchungsmodells haben. Zur besseren numerischen Stabilität werden alle kontinuierlichen Moderatorvariablen z-standardisiert und die restlichen Variablen als Dummy (0/1) in das Modell aufgenommen (Jak/Cheung 2018a). Anschließend wird als Sensitivitätstest an einem reduzierten Modell, mit nur einem Mediator (soziale Unterstützung), ein Vergleich des OSMASEM-Ansatzes mit der Subgruppen-Untersuchung des TSSEM-Ansatzes vorgenommen.

Im Subgruppen-Design wird zunächst das reduzierte Modell über beide Gruppen hinweg modelliert. Danach werden auf Basis des Variablenmedians die Gruppen gebildet. In beiden Gruppen wird dann jeweils die erste Phase des TSSEM-Verfahrens mittels Random-Effects-Modellierung durchgeführt. Anschließend wird das reduzierte Untersuchungsmodell in beiden Gruppen auf Basis der jeweiligen kumulierten Korrelationsmodelle getestet. Dann wird ein Multigruppenvergleich durchgeführt (Urban/Mayerl 2014: 40ff.). Dazu wird für eine der Gruppen, im vorliegenden Fall die High-Gruppe, eine eigene A-Matrix konstruiert. Die S-Matrix bleibt für beide Gruppen gleich. Anschließend wird die zweite Phase mit der in einer Gruppe modifizierten A-Matrix durchgeführt. Im nächsten Schritt wird das Multigruppenmodell mit beiden Gruppen konstruiert, das die drei Regressionskoeffizienten auf Gleichheit testen soll. Anschließend wird mittels Chi-Quadrat-Differenzentest die Modellveränderung statistisch überprüft. Eine Überprüfung der einzelnen Regressionskoeffizienten (einzelne Elemente der A-Matrix) wird aufgrund der niedrigeren Teststärke nicht durchgeführt.

### 9.4.1 One-Stage-MASEM (OSMASEM)

Nun folgt zunächst das Ergebnis der Moderatoruntersuchung der HDI-Variable. Das Untersuchungsmodell entspricht dabei Abbildung 24, also ohne die Kovarianz  $k_2$ , da diese sowohl im TSSEM- als auch OSMASEM-Ansatz nicht signifikant ist. Zusätzlich wird bei diesem Modell jeweils die Moderatorvariable mit in das Untersuchungsmodell aufgenommen, die auf die einzelnen Regressionskoeffizienten (Elemente der A-Matrix) wirkt. Dies ist in Abbildung 25 zu sehen.

Abbildung 25: OSMASEM Untersuchungsmodell



**Anmerkung:** Eigene Abbildung.

In OSMASEM stehen Standardfehler (SE) zur Verfügung, so dass Signifikanztests durchgeführt werden können. Ebenfalls liegen Fit-Indizes zur Evaluation des Modellfits vor (siehe Tabelle 37). In der Tabelle sind von oben nach unten zunächst alle geschätzten Regressionskoeffizienten ( $b$ ), Residuen ( $p$ ) und Kovarianzen ( $k$ ) zu sehen. Danach folgen die drei indirekten Effekte ( $ind$ ) und die Interaktionsterme der Moderatorvariable ( $HDI \rightarrow b$ ). Im Grundmodell bestehen keine Unterschiede zwischen OSMASEM und TSSEM. Erst in der zweiten oder dritten Nachkommastelle der geschätzten Parameter gibt es Unterschiede, die vor allem auf den unterschiedlichen Umgang mit fehlenden Korrelationen zurückzuführen sind. Alle geschätzten Parameter sind signifikant und die Ergebnisse und Interpretationen sind analog zur TSSEM-Analyse. Ebenfalls entspricht der Modellfit des OSMASEM-Ausgangsmodells dem des TSSEM. Sowohl der RMSEA (.0021) als auch das zugehörige Konfidenzintervall (0 - .0073) zeigen eine sehr gute Passung ( $<.05$ ) des Modells an. Auch das SRMR ist mit dem Wert .0487 in einem sehr guten Bereich ( $<.08$ ).

Auch die drei indirekten Effekte der Mediatoren werden in OSMASEM überprüft. Hierbei gibt es die Möglichkeit, die Standardfehler der indirekten Effekte oder wie beim TSSEM die LBCIs der indirekten Effekte zu berechnen. Der indirekte Effekt 1 wird mit .0316 (.0193 - .0473), der indirekte Effekt 2 mit .0263 (.0051 - .0495) und der indirekte Effekt 3 mit .0469 (.0260 - .0728) geschätzt. Die LBCIs schließen jeweils den Wert „0“ nicht ein. Auch im OSMASEM-Ansatz wirken alle drei Mediatoren partiell. Dies kann als Sensitivitätstest der bisherigen Erkenntnisse aufgefasst werden, da die Ergebnisse auch unter Anwendung eines abweichenden Schätzverfahrens bestehen bleiben.

**Tabelle 37: Vergleich TSSEM und OSMASEM (HDI-Datensatz)**

	TSSEM	OSMASEM	OSMASEM mit HDI
	Schätzung (95%igen LBCI)	Schätzung (SE)	Schätzung (SE)
b31	0.1287 [0.0892; 0.1681]	0.1292 (0.0199)	0.1146 (0.0202)
b32	0.1513 [0.1065; 0.1961]	0.1495 (0.0218)	0.1479 (0.0223)
b33	0.1496 [0.1043; 0.1949]	0.1474 (0.0220)	0.1432 (0.0225)
b34	0.0982 [0.0614; 0.1324]	0.0986 (0.0175)	0.0918 (0.0185)
b35	0.2454 [0.1621; 0.3289]	0.2445 (0.0412)	0.2313 (0.0438)
b36	0.1755 [0.0367; 0.3046]	0.1759 (0.0667)	0.2256 (0.0711)
b37	0.3210 [0.1934; 0.4468]	0.3182 (0.0642)	0.3791 (0.0721)
p22	0.9834 [0.9717; 0.9920]	0.9833 (0.0052)	0.9921 (NA)
k1	0.3642 [0.3093; 0.4191]	0.3615 (0.0232)	0.3657 (0.0222)
p33	0.9771 [0.9616; 0.9887]	0.9776 (0.0065)	0.9681 (NA)
k3	0.2658 [0.1571; 0.3745]	0.2711 (0.0553)	0.2536 (0.0516)
p44	0.9776 [0.9620; 0.9891]	0.9783 (0.0065)	0.9729 (NA)
p55	0.7069 [0.6144; 0.7804]	0.7089 (0.0416)	0.6904 (NA)
Ind1 (b31*b35)	0.0315 [0.0192; 0.0471]	0.0316 (0.0070)	0.0265 (0.0066)
Ind2 (b32*b36)	0.0265 [0.0055; 0.0496]	0.0263 (0.0106)	0.0334 (0.0114)
Ind3 (b33*b37)	0.0480 [0.0271; 0.0742]	0.0469 (0.0116)	0.0348 (0.0129)
HDI → b31	-	-	0.0354 (0.0206) n. s.
HDI → b32	-	-	-0.0425 (0.0239) n. s.
HDI → b33	-	-	-0.0297 (0.0206) n. s.
HDI → b34	-	-	-0.0600 (0.0170)
HDI → b35	-	-	-0.0424 (0.0431) n. s.
HDI → b36	-	-	0.1038 (0.0615) n. s.
HDI → b37	-	-	0.0646 (0.0557) n. s.
Chi-Quadrat-Veränderung	2.5397	1.8743	-
DF	1.0000	1.0000	-
p	0.1110	0.1710	-
RMSEA	0.0027	0.0021	-
RMSEA unt. 95%iges KI	0.0000	0.0000	-
RMSEA ober. 95%iges KI	0.0071	0.0073	-
CFI	0.9983	-	-
SRMR	0.0509	0.0487	-
Fallzahl/Effektstärken	205.483/293	205.483/293	205.483/293

**Anmerkung:** Der CFI ist nach persönlicher Aussage von Mike Cheung bislang (Stand: Dezember 2018) im OSMASEM noch nicht korrekt definiert.

Im nächsten Schritt wird der kontinuierliche Wert der HDI-Variable als Moderator aufgenommen. Dabei wird überprüft, ob die sieben Regressionskoeffizienten des Untersuchungsmodells

durch den Human-Development-Index moderiert werden. Auch in diesem Fall gibt es in Bezug auf die geschätzten Koeffizienten kaum einen Unterschied zwischen TSSEM und OSMASEM. Alle Koeffizienten und auch die indirekten Effekte<sup>126</sup> sind signifikant. Der Omnibustest des Modells mit dem HDI-Moderator ist signifikant ( $\chi^2(7) = 27,192, p < .05$ ). Es zeigt sich, dass nur der Haupteffekt (b34) statistisch signifikant durch den Moderator HDI beeinflusst wird. Erhöht sich der Human Development Index um eine Standardabweichung, sinkt der geschätzte b34 Effekt um  $-0.06$ . Im Vergleich lässt sich zwischen dem Modell mit Moderator und ohne Moderator 19,0 % der Varianz der Korrelationen von Religiosität und Lebenszufriedenheit durch HDI statistisch erklären.

Daran anknüpfend wird nun in einem gesonderten Modell der zweite Makro-Moderator „Wichtigkeit von Religion“ aus dem World Value Survey überprüft (siehe Tabelle 38). Der Omnibustest des Modells mit dem WR-Moderator ist signifikant ( $\chi^2(7) = 52,886, p < .05$ ). In diesem Fall gibt es leichte Unterschiede zwischen den geschätzten Effekten zu den Modellen ohne WR-Moderatorvariable. Allerdings hat sich an den Richtungen nichts verändert. Der relative Fit des Modells ist sehr gut. Dies zeigen RMSEA (.0020) und das zugehörige Konfidenzintervall (0 - .0074). Das SRMR ist mit dem Wert .0484 ebenfalls in einem sehr guten Bereich ( $< .08$ ).

Auch die indirekten Effekte sind signifikant und zeigen somit eine partielle Mediation der drei berücksichtigten Mediatorvariablen. Der indirekte Effekt 1 beträgt  $.0315$  ( $.0192 - .0472$ ), der indirekte Effekt 2  $.0267$  ( $.0055 - .0500$ ) und der indirekte Effekt 3  $.0488$  ( $.0260 - .0765$ ). Auch nach dem Hinzufügen der WR-Moderatorvariable bleiben alle geschätzten Parameter signifikant. Alle Koeffizienten, die indirekten Effekte<sup>127</sup> und der Omnibustest des Modells mit dem WR-Moderator sind signifikant ( $\chi^2(7) = 52,886, p < .05$ ). Dabei zeigt sich, dass die Effekte b32, b33 und b37 durch den Moderator „Wichtigkeit von Religion“ beeinflusst werden. Erhöht sich WR um eine Standardabweichung, dann steigt der geschätzte Effekt von b32 um  $.0517$ , von b33 um  $.0779$  und von b37 um  $.1512$ .

Durch die WR-Variable können 33,3 % der Varianz der Korrelation zwischen Religiosität und Optimismus (Zelle 2 bzw. b32) statistisch erklärt werden. 24,4 % der Varianz der Korrelation zwischen Religiosität und Selbstwert (Zelle 3 bzw. b33) und 11,6 % der Varianz der Korrelation

---

<sup>126</sup> Der indirekte Effekt 1 wird im OSMASEM mit HDI-Moderator mit  $.0265$  ( $.0150 - .0412$ ), der indirekte Effekt 2 mit  $.0334$  ( $.0118 - .0597$ ) und der indirekte Effekt 3 mit  $.0543$  ( $.0312 - .0835$ ) geschätzt. Die Berechnung der LBCIs ist in dem Moderationsmodell rechenintensiv, so dass je nach Hardware darauf verzichtet werden sollte.

<sup>127</sup> Der indirekte Effekt 1 beträgt  $.0358$  ( $.0211 - .0560$ ), der indirekte Effekt 2  $.0275$  ( $.0052 - .0512$ ) und der indirekte Effekt 3  $.0370$  ( $.0136 - .0626$ ).

zwischen Selbstwert und Lebenszufriedenheit (Zelle 10 bzw. b37) können statistisch erklärt werden.

**Tabelle 38: Vergleich TSSEM und OSMASEM (WR-Datensatz)**

	TSSEM	OSMASEM	OSMASEM mit WR
	Schätzung (95%iges LBCI)	Schätzung (SE)	Schätzung (SE)
b31	0.1288 [0.0893; 0.1682]	0.1293 (0.0199)	0.1380 (0.0251)
b32	0.1513 [0.1065; 0.1961]	0.1498 (0.0217)	0.1563 (0.0196)
b33	0.1592 [0.1135; 0.2050]	0.1565 (0.0223)	0.1555 (0.0200)
b34	0.0977 [0.0596; 0.1332]	0.0981 (0.0181)	0.0837 (0.0181)
b35	0.2443 [0.1609; 0.3278]	0.2435 (0.0413)	0.2592 (0.0418)
b36	0.1779 [0.0389; 0.3071]	0.1783 (0.0667)	0.1765 (0.0675)
b37	0.3144 [0.1809; 0.4460]	0.3116 (0.0671)	0.2382 (0.0702)
p22	0.9834 [0.9717; 0.9920]	0.9832 (0.0052)	0.9718 (NA)
k1	0.3642 [0.3093; 0.4190]	0.3615 (0.0232)	0.3627 (0.0218)
p33	0.9771 [0.9616; 0.9887]	0.9775 (0.0065)	0.9467 (NA)
k3	0.2643 [0.1556; 0.3730]	0.2700 (0.0552)	0.2709 (0.0503)
p44	0.9746 [0.9580; 0.9871]	0.9755 (0.0070)	0.9284 (NA)
p55	0.7103 [0.6152; 0.7847]	0.7125 (0.0424)	0.5105 (NA)
Ind1 (b31*b35)	0.0315 [0.0191; 0.0469]	0.0315 (0.0070)	0.0358 (0.0086)
Ind2 (b32*b36)	0.0269 [0.0059; 0.0501]	0.0267 (0.0106)	0.0275 (0.0110)
Ind3 (b33*b37)	0.0501 [0.0274; 0.0780]	0.0566 (0.0124)	0.0370 (0.0118)
WR → b31	-	-	0.0207 (0.0310) n. s.
WR → b32	-	-	0.0517 (0.0219)
WR → b33	-	-	0.0779 (0.0195)
WR → b34	-	-	0.0118 (0.0191) n. s.
WR → b35	-	-	0.0930 (0.0650) n. s.
WR → b36	-	-	-0.0476 (0.0746) n. s.
WR → b37	-	-	0.1512 (0.0631)
Chi-Quadrat-Veränderung	2.4996	1.8576	-
DF	1.0000	1.0000	-
p	0.1139	0.1729	-
RMSEA	0.0027	0.0020	-
RMSEA unt. 95%iges KI	0.0000	0.0000	-
RMSEA ober. 95%iges KI	0.0072	0.0074	-
CFI	0.9983	-	-
SRMR	0.0505	0.0484	-
Fallzahl/Effektstärken	203.317/281	203.317/281	203.317/281

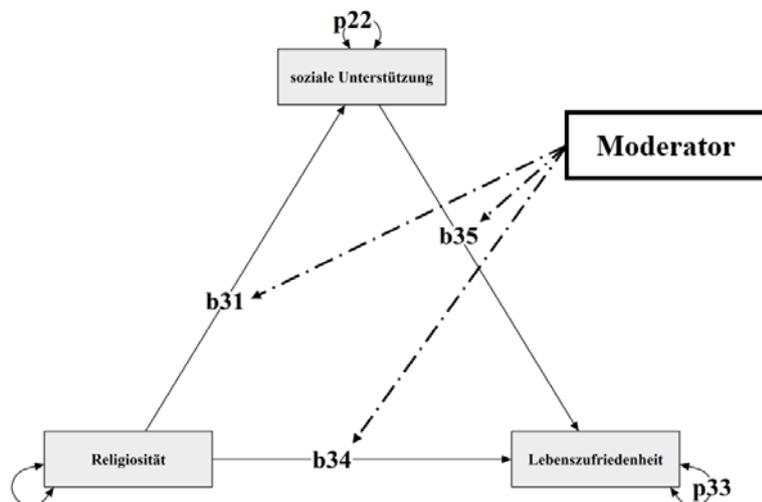
Auch methodische Moderatoren werden in Anlehnung an den Review der meta-analytischen Praxis (siehe Kapitel 4.1.5) überprüft. Die Grundmodelle der einzelnen Modellierungen unterscheiden sich in ihrer grundsätzlichen Aussagekraft (Richtung, Stärke der Effekte und Modellfit) nicht von den bisherigen Modellen und werden hier nicht ausgewiesen.

Eine Überprüfung der Fallzahl als Moderator bleibt ergebnislos ( $\chi^2(7) = 3,11$ ,  $p = .211$ ). Keiner der Moderatoreffekte ist signifikant und die subjektive Annahme, dass vor allem kleine Studien extremere Effektstärken aufweisen, kann somit verworfen werden. Ebenso verhält es sich mit dem Dummymoderator Publikationsstatus (graue Literatur = 1; Rest = 0). Der Omnibustest ist zwar mit  $\chi^2(7) = 19,12$ ,  $p < .05$  signifikant, allerdings keiner der Interaktionsterme. Lediglich

die Moderation auf b31 ist beinahe signifikant. Dies kann bereits als erste Falsifizierung eines potenziellen Publikationsbias verstanden werden (siehe Kapitel 5.5.1). Insgesamt bleiben die Ergebnisse jedoch auch unter Berücksichtigung des Publikationsstatus robust. Auch die Überprüfung des Repräsentativitätsdummys (repräsentative Stichprobe = 1; Rest = 0) bleibt vermutlich aufgrund der nur geringen Anzahl von repräsentativen Studien und der damit zusammenhängenden zu geringen Teststärke ergebnislos, da es zu keiner gültigen konvergierenden Lösung kommt. Die subjektiv eingeschätzte Studienqualität ist auch keine erklärungskräftige Moderatorvariable ( $\chi^2(7) = 3,11, p = .855$ ). Lediglich das Publikationsjahr der Studien hat einen moderierenden Einfluss auf die Regressionskoeffizienten des Untersuchungsmodells ( $\chi^2(7) = 21,95, p < .05$ ). Der Einfluss der Variable Publikationsjahr macht sich dabei in Form eines Periodeneffekts auf die beiden Effekte b33 und b36 bemerkbar. Erhöht sich die Variable Publikationsjahr um eine Standardabweichung, dann steigt der geschätzte Effekt von b33 um .0529 beziehungsweise sinkt der geschätzte Wert von b36 um -.1801. Durch die Variable Publikationsjahr können 32,2 % der Varianz der Korrelation zwischen Religiosität und Selbstwert (Zelle 3 bzw. b33) und 31,0 % der Varianz der Korrelation zwischen Optimismus und Lebenszufriedenheit (Zelle 9 bzw. b36) statistisch erklärt werden. Hinter diesen Moderatoreffekten des Publikationsjahres verbergen sich vermutlich gesamtgesellschaftliche Veränderungen und wissenschaftliche Moden, die sich durch die Veränderung der Variablen bzw. der dahinterstehenden Konstrukte bemerkbar machen.

Um die beiden Methoden zur Analyse von Moderatoreffekten anschließend vergleichen zu können (OSMASEM und TSSEM mit Subgruppenanalyse), wird im nächsten Schritt ein reduziertes Untersuchungsmodell mit nur einem Mediator getestet (siehe Abbildung 26 und Tabelle 39). Das Modell ist genau identifiziert und somit können keine Fit-Indizes berechnet werden.

**Abbildung 26: Reduziertes OSMASEM Untersuchungsmodell**



**Anmerkung:** Eigene Abbildung.

Zunächst wird das reduzierte Modell mit den HDI-Daten betrachtet. Die im Modell eingeschlossenen Effekte sind ähnlich wie im vollständigen Untersuchungsmodell. Auch der indirekte Effekt 0.0400 mit seinem zugehörigen LBCI (.0270 - 0.0555) ist in dem verringerten Modell signifikant. Soziale Unterstützung wirkt somit als partieller Mediator.

**Tabelle 39: Reduziertes OSMASEM Untersuchungsmodell**

	OSMASEM mit HDI	OSMASEM mit WR
	Schätzung (SE)	Schätzung (SE)
b31	0.1286 (0.0199)	0.1287 (0.0200)
b34	0.1633 (0.0138)	0.1659 (0.0141)
b35	0.3114 (0.0293)	0.3111 (0.0293)
p22	0.9835 (0.0051)	0.9834 (0.0051)
p55	0.8633 (0.0184)	0.8624 (0.0184)
Ind1 (b31*b35)	0.0400 (0.0071)	0.0400 (0.0070)
Mod. → b31	0.0293 (0.0216) n. s.	0.0272 (0.0338) n. s.
Mod. → b34	- 0.0480 (0.0123)	0.0337 (0.0213) n. s.
Mod. → b35	0.0059 (0.0333) n. s.	0.0912 (0.0537) n. s.
Fallzahl/Effektstärken	177.088/179	172.746/180

Der Omnibustest des Modells mit dem HDI-Moderator ist signifikant ( $\chi^2(3) = 14.71$ ,  $p < .05$ ). Der Effekt b34 wird signifikant durch den HDI beeinflusst. Erhöht sich die HDI-Variable um eine Standardabweichung, dann sinkt der geschätzte Effekt von b34 um -.0480. Durch die HDI-Variable können 10,8 % der Varianz der Korrelation zwischen Religiosität und Lebenszufriedenheit (Zelle 4 bzw. b34) statistisch erklärt werden. Ähnlich verhält sich das Modell im WR-Datensatz. Die berücksichtigten Effekte sind ähnlich wie im vollständigen Untersuchungsmodell. Der indirekte Effekt (0.0400) mit seinem zugehörigen LBCI (.0270 - 0.0554) ist auch in dem verringerten Modell signifikant und somit liegt ein partieller Mediator vor. Der Omnibustest des Modells mit dem WR-Moderator ist signifikant ( $\chi^2(3) = 29.709$ ,  $p < .05$ ). Jedoch ist keiner der drei Interaktionsterme signifikant.

## 9.4.2 TSSEM mit Subgruppenanalyse

Für die Moderatoranalyse mittels Subgruppendesign werden die Daten anhand der beiden (quasi-)metrischen Makrovariablen mittels Median geteilt. In Tabelle 40 ist eine Übersicht der Trennwerte zu sehen.

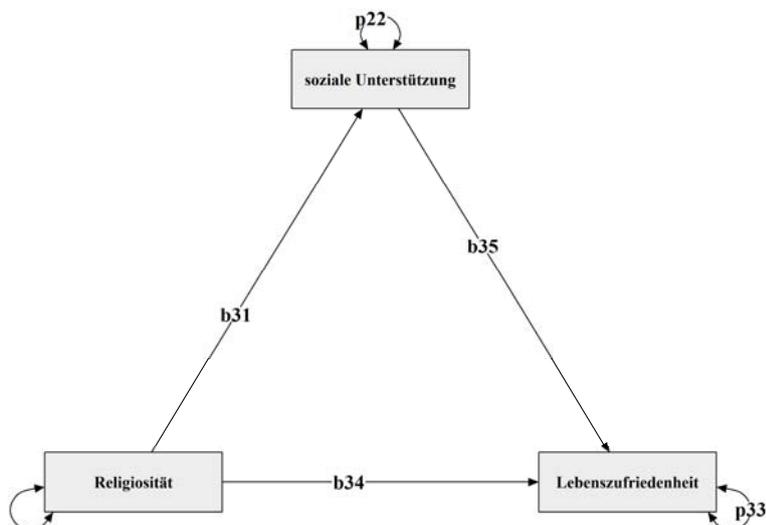
**Tabelle 40: Mediansplit der Makrovariablen für die Moderatoranalyse**

Variable	Eigenschaft	Low-Gruppe	High-Gruppe
Human Development Index (HDI)	Median	< .889	>= .889
	Stichprobenanzahl	72	80
	Anzahl der Effektstärken	79	108
	Fallzahl	34.819	142.269
Wichtigkeit von Religion	Median	<47	>=47
	Stichprobenanzahl	56	86
	Anzahl der Effektstärken	64	116
	Fallzahl	49.374	123.372

**Anmerkungen:** n (HDI-DS) = 177.088 Befragte; n (WR-DS) = 172.746.

Das reduzierte Untersuchungsmodell ist in Abbildung 27 veranschaulicht. Zunächst folgen nun die Ergebnisse der Untersuchung der Moderatorvariable HDI auf Basis von 152 Stichproben mit 187 Effektstärken und 177.088 Befragten.

**Abbildung 27: Ausgangsmodell für die Moderatoranalyse**



**Anmerkung:** Eigene Abbildung.

In Tabelle 41 ist das Ergebnis der ersten Phase der TSSEM-Analyse zu sehen. Über der Diagonale sind die kumulierten Korrelationskoeffizienten und in Klammern die Fallzahlen für die jeweiligen Beziehungen aufgeführt. Die kumulierten Korrelationskoeffizienten entsprechen den Werten, die auch im vollständigen Modell geschätzt werden. Unterhalb der Diagonale sind die Anzahl der Stichproben auf der die jeweilige kumulierte Effektstärke beruht, und in Klammer der zugehörige  $I^2$ -Wert abgetragen. Dieser schwankt zwischen .884 und .940.

**Tabelle 41: Erste Phase reduziertes Untersuchungsmodell (HDI-Datensatz)**

	Religiosität	Soz. Unterstützung	Lebenszufriedenheit
Religiosität		0.128 (16.974)	0.204 (71.636)
Soz. Unterstützung	34 (0.884)		0.332 (105.689)
Lebenszufriedenheit	124 (0.915)	29 (0.940)	

**Anmerkung:** Über der Diagonale befinden sich die kumulierten Korrelationskoeffizienten und in Klammern die Fallzahlen für die jeweiligen Beziehungen. Unterhalb der Diagonale sind die Anzahl der Stichproben auf der die jeweilige kumulierte Effektstärke beruht, und in Klammer der zugehörige  $I^2$ -Wert zu sehen.

Der formale Homogenitätstest zeigt, dass keine Homogenität in der Korrelationsmatrix vorliegt, da die Nullhypothese mit  $Q$  ( $df = 184$ ) = 1148 mit .0001,  $p < 0.00$  verworfen wird. Aufgrund des Ausmaßes an Heterogenität ist ein Random-Effects-Modell notwendig. Das Ergebnis der zweiten Phase der TSSEM-Analyse ist in Tabelle 42 zu sehen. Die Beziehung zwischen Religiosität und sozialer Unterstützung (b31) ist mit .1278 ähnlich wie im Gesamtmodell. Der Haupteffekt b34 hingegen hat sich deutlich nach oben verschoben (.1638). Noch deutlicher ist die Veränderung bei b35, also der Beziehung zwischen sozialer Unterstützung und Lebenszufriedenheit (.3107). Im Vergleich zum Gesamtmodell sind die LBCIs bei b34 und b35 schmaler.

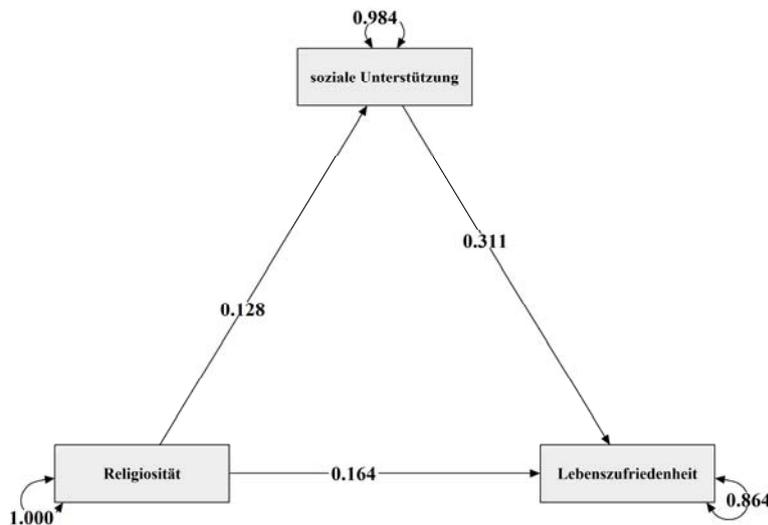
**Tabelle 42: Zweite Phase reduziertes Untersuchungsmodell (HDI-Datensatz)**

	Ohne Optimismus und Selbstwert
b31	0.1278 [0.0885; 0.1672]
b34	0.1638 [0.1363; 0.1907]
b35	0.3107 [0.2533; 0.3679]
p22	0.9837 [0.9720; 0.9922]
p33	0.8636 [0.8247; 0.8965]
Chi-Quadrat Nullmodell	467.0503
Fallzahl	177.088

In Abbildung 28 ist das zugehörige Pfadmodell mit den Ergebnissen der Schätzung visualisiert. Der indirekte Effekt und sein 95%iges LBCI beträgt .040 (.027 - .054).<sup>128</sup> Der indirekte Effekt ist signifikant und die Variable soziale Unterstützung wirkt als partieller Mediator. Im Vergleich zum Gesamtmodell hat sich der indirekte Effekt leicht erhöht, ist jedoch ebenfalls gering. Die Erklärungsleistung dieses Modells ist mit 13,6 % statistischer Varianzaufklärung der abhängigen Variable Lebenszufriedenheit schwächer als im Gesamtmodell.

<sup>128</sup> Aus Platzgründen wird der indirekte Effekt in den Ergebnistabellen der Subgruppen-Untersuchung nicht ausgewiesen.

**Abbildung 28: Pfadmodell reduziertes Untersuchungsmodell (HDI-Datensatz)**



**Anmerkung:** Eigene Abbildung; n = 177.088.

Im nächsten Schritt wird die erste Phase getrennt für beide Gruppen (Low- und High-HDI) durchgeführt. In Tabelle 43 sind die  $I^2$ -Werte in Klammern und die Anzahl von Effektstärken je Beziehung für beide Gruppen zu sehen. Über der Diagonale sind die Werte für die Low-HDI-Gruppe und unterhalb für die High-HDI-Gruppe abgetragen. Deutlich erkennbar ist, dass die Beziehung von Religiosität und sozialer Unterstützung sowie zwischen sozialer Unterstützung und Lebenszufriedenheit in der Low-HDI-Gruppe auf nur wenigen Effektstärken beruht. Es zeigen sich deutliche Unterschiede im Vergleich zur Berechnung über beide Gruppen hinweg. Die Heterogenität zwischen den Studien ist durch die Aufteilung der Stichproben in zwei Gruppen gesunken.

**Tabelle 43: Erste Phase für beide Gruppen der HDI-Moderatoranalyse ( $I^2$ -Werte)**

	Religiosität	Soz. Unterstützung	Lebenszufriedenheit
Religiosität		6 (.543)	66 (.851)
Soz. Unterstützung	28 (.914)		7 (.814)
Lebenszufriedenheit	58 (.911)	22 (.957)	

**Anmerkung:** Low-Gruppe oberhalb der Diagonale und High-Gruppe unterhalb.

Der formale Homogenitätstest für die Low-Gruppe besagt, dass keine Homogenität der Korrelationsmatrix vorliegt, da die Nullhypothese mit  $Q$  (df = 76) = 447 mit .0001,  $p < 0.00$  verworfen wird. Auch in der High-Gruppe liegt signifikante Heterogenität zwischen den Studien vor ( $Q$  (df = 105) = 553 mit .0001,  $p < 0.00$ ).

In Tabelle 44 sind die kumulierten Korrelationskoeffizienten für beide Gruppen ersichtlich. Am ähnlichsten zwischen den beiden Gruppen ist der Zusammenhang zwischen sozialer Unterstützung und Lebenszufriedenheit. In der Low-Gruppe beträgt dieser .318 und in der High-Gruppe .335. Unterschiedlicher sind die Werte bei der Beziehung zwischen Religiosität und sozialer Unterstützung (.077 vs. .141) sowie bei der Hauptbeziehung (.246 vs. .153).

**Tabelle 44: Erste Phase der HDI-Moderatoranalyse (getrennt)**

	Religiosität	Soz. Unterstützung	Lebenszufriedenheit
Religiosität		.141	.153
Soz. Unterstützung	.077		.335
Lebenszufriedenheit	.246	.318	

**Anmerkung:** Low-Gruppe oberhalb der Diagonale und High-Gruppe unterhalb.

In der Low-Gruppe ist der Zusammenhang zwischen Religiosität und sozialer Unterstützung schwächer und der Zusammenhang zwischen Religiosität und Lebenszufriedenheit stärker ausgeprägt als in der High-Gruppe. Aufgrund der geringen Anzahl an Effektstärken je Beziehung sollte das Hauptaugenmerk jedoch auf Unterschiede bei der Hauptbeziehung (Religiosität und Lebenszufriedenheit) gesetzt werden.

Nun wird im nächsten Schritt des TSSEM-Verfahrens das Strukturgleichungsmodell in beiden Gruppen getestet (zweite Phase). In Tabelle 45 ist das Ergebnis dieser getrennten Schätzung zu sehen. An den Regressionskoeffizienten lassen sich deutliche Unterschiede zwischen den Gruppen erkennen. Während der Effekt von Religiosität auf soziale Unterstützung (b31) in der Low-HDI-Gruppe .077 beträgt, ist er in der High-HDI-Gruppe .141.

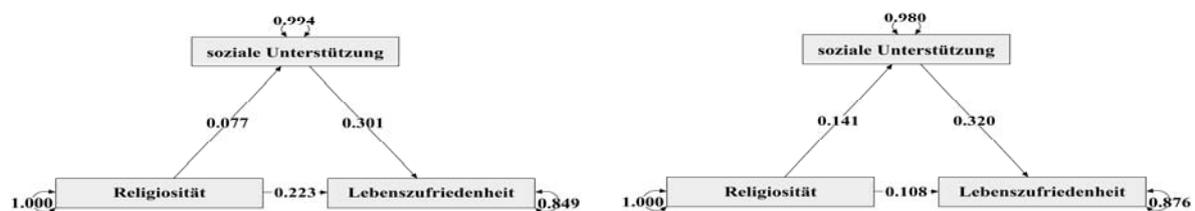
**Tabelle 45: Zweite Phase in beiden Gruppen (HDI-Datensatz)**

	Low-HDI	High-HDI
b31	0.0771 [0.0239; 0.1303]	0.1412 [0.0944; 0.1879]
b34	0.2226 [0.1877; 0.2567]	0.1079 [0.0709; 0.1439]
b35	0.3012 [0.2128; 0.3895]	0.3201 [0.2496; 0.3906]
p22	0.9940 [0.9830; 0.9994]	0.9801 [0.9647; 0.9911]
p33	0.8494 [0.7874; 0.8968]	0.8761 [0.8267; 0.9162]
Chi-Quadrat Nullmodell	317.7936	213.8123
Fallzahl	34.819	142.269

Auch beim Haupteffekt (b34) gibt es einen Unterschied. In der Gruppe der Länder mit unterdurchschnittlichem HDI ist der Effekt mehr als doppelt so hoch als in den Ländern mit hohem HDI (.222 vs. .108). Der Effekt von sozialer Unterstützung auf Lebenszufriedenheit ist hingegen in beiden Gruppen ähnlich. Aufgrund der exakten Passung des Modells können die Fit-Indizes in den Gruppen nicht berichtet werden, da aber die LBCIs bei keinem Schätzer die „0“ einschließen, kann davon ausgegangen werden, dass eine angemessene Passung des Modells

vorliegt. Die LBCIs sind in Bezug auf ihre Spannweite in beiden Gruppen ähnlich. Der indirekte Effekt und sein 95%iges LBCI von Religiosität über soziale Unterstützung auf Lebenszufriedenheit beträgt in der Low-HDI-Gruppe .023 (.007 - .041). In der High-HDI-Gruppe ist der indirekte Effekt hingegen .045 (.029 - .065). In beiden Gruppen ist der indirekte Effekt somit signifikant und es liegt eine partielle Mediation durch die Variable soziale Unterstützung vor. In Abbildung 29 ist zur Übersicht das Schätzergebnis in beiden Gruppen als Pfadmodell dargestellt.

**Abbildung 29: Low- und High-HDI-Gruppe**



**Anmerkung:** Eigene Abbildung; Links Low-HDI-Gruppe und rechts High-HDI-Gruppe.

In der Low-HDI-Gruppe werden 15,1 % der Varianz von Lebenszufriedenheit und in der High-HDI-Gruppe 12,4 % der Varianz von Lebenszufriedenheit statistisch erklärt.

Anschließend werden die Modelle mit einer eigenen A-Matrix für die HDI-High-Gruppe und gleicher S-Matrix in beiden Gruppen erneut geschätzt. Danach wird ein Multigruppenmodell erstellt, das sowohl die Low- als auch die High-HDI-Gruppe enthält. Die Regressionskoeffizienten ( $b_{31}$ ,  $b_{34}$  und  $b_{35}$ ) werden mittels Constraints in beiden Modellen gleichgesetzt. Das Ergebnis des gleichgesetzten Modells wird mittels Chi-Quadrat-Differenzentest mit dem freigeschätzten Modell verglichen. Dies entspricht zwar nicht der formellen Überprüfung eines spezifischen Regressionskoeffizienten wie im OSMASEM, stattdessen wird überprüft, ob es einen Unterschied in einem der drei vorliegenden Regressionskoeffizienten gibt. Eine vollständige Überprüfung aller Koeffizienten gleichzeitig ist aufgrund der genauen Identifikation des Modells nicht möglich, da ansonsten keine Aussagen mehr zur Modellverschlechterung getroffen werden können. Lediglich der Haupteffekt  $b_{34}$  wird einzeln getestet, da dieser im OSMA-SEM-Ansatz signifikant durch den HDI moderiert wird.

In Tabelle 46 ist ein Vergleich des Modellfits der Modelle zu sehen. Im frei geschätzten Modell können die herkömmlichen Fit-Indizes aufgrund der exakten Passung des Modells nicht berechnet werden. Im Modell mit Gleichheitsconstraints werden jedoch durch die Gleichsetzung drei Freiheitsgrade gewonnen, so dass für dieses Modell Fit-Indizes berechnet werden. Der CFI

dieses Modell beträgt .966 und der RMSEA .008 (.004 - .012). Das SRMR beträgt .033. Insgesamt kann von einer guten Modellpassung ausgegangen werden. Die Parameterschätzungen für die b-Koeffizienten des Constraint-Modells ähneln dem Modell über beide Gruppen hinweg. Der Koeffizient b31 für die Beziehung zwischen Religiosität und sozialer Unterstützung beträgt .114 (.079 - .149), der Haupteffekt b34 .166 (.141 - .191) und der Effekt b35 für die Beziehung zwischen sozialer Unterstützung und Lebenszufriedenheit .310 (.255 - .365). Der Chi-Quadrat-Differenzentest ergibt bei drei Freiheitsgraden eine signifikante Verschlechterung der Modellpassung des Constraint-Modells im Vergleich zum frei geschätzten Modell.

**Tabelle 46: Vergleich freies Modell mit Constraint-Modell (HDI-Datensatz)**

	Freies Modell	Constraint-Modell	Constraint (nur b34 frei)
DF	0	3	2
Chi-Quadrat	0	20.907	3.341
p	.000	.000	.188
RMSEA	-	.008	.003
RMSEA unt. 95%iges KI	-	.004	.000
RMSEA ober. 95%iges KI	-	.012	.009
CFI	1.000	.966	.997
TLI	-	.932	.992
AIC	24	38.907	23.341
BIC	145.013	129.666	124.185
SRMR	-	.033	.018

Deshalb muss die Null-Hypothese von gleichen direkten Effekten zwischen den Subgruppen zurückgewiesen werden. Im OSMASEM ist lediglich der Effekt auf b34 signifikant. Deshalb wird das Modell auch in der Subgruppenanalyse erneut nur mit dem gleichgesetzten Haupteffekt b34 geschätzt. Dort zeigt sich keine signifikante Verschlechterung des Modells und somit muss die Null-Hypothese von gleichen direkten Effekten zwischen den Subgruppen beibehalten werden. Dies zeigt, dass die Subgruppenanalyse eine geringere Teststärke aufweist als der OSMASEM-Ansatz. Nur auf Basis der OSMASEM-Ansatzes kann im verringerten Untersuchungsmodell nachgewiesen werden, dass HDI als Moderator wirkt.

Im nächsten Schritt wird dasselbe Vorgehen für die zweite Kulturmakrovariable „Wichtigkeit von Religion“ wiederholt, um zu überprüfen, ob sich die Effekte des Untersuchungsmodells in stark religiösen Staaten von stärker säkularen Staaten unterscheiden (Low-WR vs. High-WR). Zunächst wird abermals das reduzierte Untersuchungsmodell (siehe Abbildung 27) über beide Gruppen hinweg geprüft. In der nachfolgenden Tabelle 47 ist das Ergebnis der ersten Phase zu sehen.

**Tabelle 47: Erste Phase reduziertes Untersuchungsmodell (WR-Datensatz)**

	Religiosität	Soz. Unterstützung	Lebenszufriedenheit
Religiosität		0.128 (16.974)	0.206 (69.797)
Soz. Unterstützung	34 (0.886)		0.332 (105.689)
Lebenszufriedenheit	117 (0.917)	29 (0.942)	

**Anzahl:** Oberhalb der Diagonale sind die kumulierten Effektstärken und in Klammer die jeweilige Fallzahl zu sehen. Unterhalb der Diagonale sind die Anzahl der Effektstärken und in Klammer die  $I^2$ -Werte zu sehen.

Dieses Untersuchungsmodell basiert auf 180 Effektstärken von 142 Stichproben mit insgesamt 172.746 Befragten. Die kumulierten Effektstärken unterscheiden sich kaum von denjenigen des HDI-Datensatzes. Lediglich die Effektstärke von Lebenszufriedenheit ist leicht erhöht (.204 vs. .206). In Klammern sind über der Diagonale die Fallzahlen zu sehen, auf denen die jeweilige Beziehung beruht. Die Anzahl der Stichproben hat sich im Vergleich zum vollständigen Datensatz lediglich bei der Beziehung zwischen Religiosität und Lebenszufriedenheit um sieben Stichproben auf 117 reduziert. Auch die  $I^2$ -Werte unterscheiden sich nur in der dritten Nachkommastelle. Der formale Homogenitätstest besagt, dass keine Homogenität der Korrelationsmatrix vorliegt, da die Nullhypothese mit  $Q$  ( $df = 180$ ) = 1124 mit .0001,  $p < 0.00$  verworfen wird. Auch im WR-Datensatz liegt eine erhebliche Variabilität zwischen den Studien vor und die Schätzung eines Random-Effects-Modells ist notwendig.

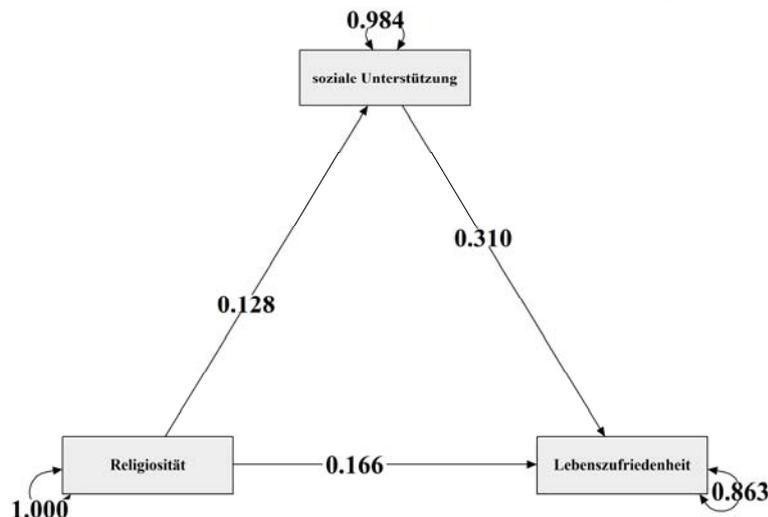
In Tabelle 48 ist das Ergebnis der zweite Phase der Schätzung des reduzierten Modells mit den WR-Daten über beide Gruppen hinweg zu sehen. Vergleicht man die geschätzten b-Koeffizienten aus dem WR-Datensatz mit denjenigen auf Basis des HDI-Datensatzes ergeben sich keine bedeutsamen Unterschiede. Das Modell ist ebenfalls genau identifiziert, so dass eine Betrachtung der Fit-Indizes nicht möglich ist.

**Tabelle 48: Zweite Phase reduziertes Untersuchungsmodell (WR-Datensatz)**

	Ohne Optimismus und Selbstwert
b31	0.1279 [0.0886; 0.1673]
b34	0.1665 [0.1384; 0.1939]
b35	0.3103 [0.2529; 0.3676]
p22	0.9836 [0.9720; 0.9922]
p33	0.8628 [0.8238; 0.8957]
Chi-Quadrat Nullmodell	457.6374
Fallzahl	172.746

In Abbildung 30 ist das Schätzergebnis in Form eines Pfaddiagramms zu sehen. Der indirekte Effekt und sein 95%iges LBCI von Religiosität über soziale Unterstützung auf Lebenszufriedenheit beträgt .040 (.027 - .054). Auch im reduzierten Modell des WR-Datensatzes wirkt soziale Unterstützung somit als partieller Mediator. Im Vergleich zum Gesamtmodell hat sich der indirekte Effekt leicht erhöht. 13,7 % der Varianz der abhängigen Variable Lebenszufriedenheit werden dadurch statistisch erklärt.

**Abbildung 30: Pfadmodell reduziertes Untersuchungsmodell (WR-Datensatz)**



**Anmerkung:** Eigene Abbildung; n = 177.746.

Im nächsten Schritt wird die erste Phase getrennt für beide Gruppen (Low- und High-WR) durchgeführt. In Tabelle 49 sind nun die  $I^2$ -Werte in Klammern und die zugehörige Anzahl von Effektstärken je Beziehung für die beiden Gruppen zu sehen. Über der Diagonale sind die Werte für die Low-WR-Gruppe und unterhalb für die High-WR-Gruppe zu sehen. Deutlich erkennbar ist, dass die Beziehung zwischen Religiosität und sozialer Unterstützung sowie zwischen sozialer Unterstützung und Lebenszufriedenheit in der Low-WR-Gruppe auf nur wenigen Effektstärken beruht.

**Tabelle 49: Erste Phase der WR-Moderatoranalyse ( $I^2$ -Werte)**

	Religiosität	Soz. Unterstützung	Lebenszufriedenheit
Religiosität		15 (.845)	42 (.906)
Soz. Unterstützung	19 (.866)		7 (.939)
Lebenszufriedenheit	75 (.848)	22 (.911)	

**Anmerkung:** Low-Gruppe oberhalb der Diagonale und High-Gruppe unterhalb.

Die Heterogenität zwischen den Studien hat sich durch die Aufteilung der Stichproben in zwei Gruppen verringert. Der formale Homogenitätstest für die Low-Gruppe besagt, dass keine Homogenität der Korrelationsmatrix vorliegt, da die Nullhypothese mit  $Q(df = 61) = 377$  mit  $.0001, p < 0.00$  verworfen wird. Auch in der High-Gruppe zeigt der Test der  $Q$ -Statistik signifikante Heterogenität zwischen den Studien an ( $Q(df = 113) = 526$  mit  $.0001, p < 0.00$ ). Deshalb muss auch in der zweiten Phase jeweils ein Random-Effects-Modell spezifiziert werden.

In Tabelle 50 sind die kumulierten Korrelationskoeffizienten für beide WR-Gruppen zu sehen. Dort lassen sich Unterschiede zwischen den Gruppen feststellen.

**Tabelle 50: Erste Phase der WR-Moderatoranalyse (getrennt)**

	Religiosität	Soz. Unterstützung	Lebenszufriedenheit
Religiosität		.098	.144
Soz. Unterstützung	.157		.264
Lebenszufriedenheit	.242	.356	

**Anmerkung:** Low-Gruppe oberhalb der Diagonale und High-Gruppe unterhalb.

In der Low-Gruppe beträgt der Zusammenhang von Religiosität und sozialer Unterstützung  $.098$  und in der High-Gruppe  $.157$ . Bei der Hauptbeziehung zwischen Religiosität und Lebenszufriedenheit ( $.144$  vs.  $.242$ ) sowie der Beziehung zwischen sozialer Unterstützung und Lebenszufriedenheit ( $.264$  vs.  $.356$ ) liegen ebenfalls deutliche Unterschiede vor. Alle drei Beziehungen sind in der Low-WR-Gruppe deutlich schwächer ausgeprägt.

Nun wird in der zweiten Phase des TSSEM-Verfahrens in beiden Gruppen das gleiche Strukturgleichungsmodell getestet. In Tabelle 51 ist das Ergebnis der separaten Schätzungen zu sehen. Es zeigen neben den Regressionskoeffizienten auch bei den Residuen deutliche Unterschiede zwischen den Gruppen. In der Low-WR-Gruppe beträgt der Effekt von Religiosität auf soziale Unterstützung  $b_{31} = .099$ , während der gleiche Effekt in der High-WR-Gruppe bei  $.157$  liegt. Der Haupteffekt  $b_{34}$ , die Beziehung zwischen Religiosität und Lebenszufriedenheit, liegt in der Low-WR-Gruppe bei  $.120$  und in der High-WR-Gruppe bei  $.191$ . Der Effekt  $b_{35}$  zwischen sozialer Unterstützung und Lebenszufriedenheit beträgt in der Low-WR-Gruppe  $.252$  und in der High-WR-Gruppe  $.326$ . Auch die beiden geschätzten Residuen  $p_{22}$  und  $p_{33}$  unterscheiden sich. Somit sind bereits deskriptiv deutliche Unterschiede zwischen den Gruppen erkennbar. Der indirekte Effekt und sein 95%iges LBCI beträgt in der Low-WR-Gruppe  $.024$  ( $.010$  -

.045). In der High-WR-Gruppe beträgt der indirekte Effekt hingegen .051 (.032 - .073). Eine partielle Mediation durch die Variable soziale Unterstützung liegt somit in beiden Gruppen vor.

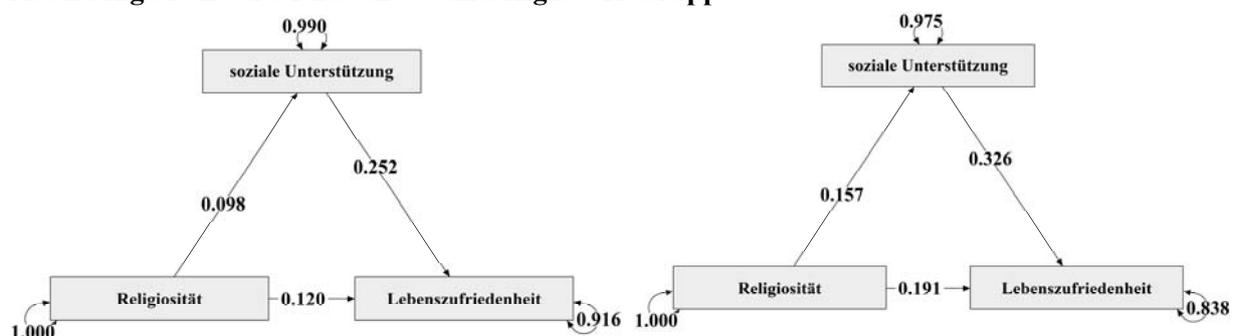
**Tabelle 51: Schätzergebnis der zweiten Phase in beiden Gruppen (WR-Datensatz)**

	Low-WR	High-WR
b31	0.0985 [0.0456; 0.1513]	0.1572 [0.1000; 0.2144]
b34	0.1198 [0.0757; 0.1626]	0.1905 [0.1565; 0.2232]
b35	0.2519 [0.1333; 0.3705]	0.3257 [0.2627; 0.3886]
p22	0.9903 [0.9771; 0.9979]	0.9753 [0.9540; 0.9900]
p33	0.9163 [0.8427; 0.9630]	0.8381 [0.7931; 0.8759]
Chi-Quadrat Nullmodell	84.3659	477.2548
Fallzahl	49.374	123.372

In Abbildung 31 sind die geschätzten Koeffizienten als Pfadmodelle dargestellt. Was die Varianzaufklärung des Modells innerhalb der beiden Gruppen betrifft, gibt es stärkere Unterschiede zwischen den beiden Median-Split-Gruppen als beim HDI-Moderator. In der Low-WR-Gruppen können lediglich 8,4 % der Varianz der abhängigen Variable Lebenszufriedenheit statistisch erklärt werden, während in der High-WR-Gruppe 16,2 % der abhängigen Variable erklärt werden können.

Dies kann als Indiz für die Wirksamkeit des Moderators „Wichtigkeit von Religion“ gedeutet werden, da das gleiche Erklärungsmodell in religiöseren Stichproben (Wichtigkeit von Religion  $\geq 47\%$ ) eine höhere Varianzaufklärung hat. So ist etwa der Haupteffekt von Religiosität auf Lebenszufriedenheit in religiösen Staaten höher als in stärker säkularen Ländern (mit niedrigeren WR-Werten). Auch die Effekte b31 und b35 unterscheiden sich zwischen den beiden Gruppen. Die individuelle Religiosität nimmt demzufolge in stark religiösen Gesellschaften auch empirisch eine entscheidendere Rolle zur Erklärung der Lebenszufriedenheit ein als in weniger religiösen Gesellschaften. In säkularen Ländern hingegen muss davon ausgegangen werden, dass konkurrierende Erklärungsvariablen einen höheren Einfluss haben.

**Abbildung 31: Zweite Phase Low- und High-WR-Gruppe**



**Anmerkung:** Eigene Abbildung; Links Low-WR-Gruppe und rechts High-WR-Gruppe.

Im nächsten Schritt folgt der formale Test des Multigruppenmodells. Dazu werden die Modelle mit einer eigenen A-Matrix für die High-WR-Gruppe und gleicher S-Matrix in beiden Gruppen

erneut geschätzt. Danach wird ein Multigruppenmodell erstellt, das sowohl die Low- als auch die High-WR-Gruppe enthält und die Regressionskoeffizienten (b31, b34 und b35) werden mittels Constraints gleichgesetzt. Tabelle 52 zeigt den Vergleich dieser beiden Modelle. Auf eine Einschränkung der gleichgesetzten Effekte, wie zuvor beim HDI-Moderator, wird an dieser Stelle verzichtet, da sich im OSMASEM-Ansatz kein signifikanter Effekt auf einen der drei Regressionskoeffizienten gezeigt hat.

**Tabelle 52: Vergleich freies Modell und Constraint Modell (WR-Moderator)**

	Freies Modell	Constraint-Modell
DF	0	3
Chi-Quadrat	0	20.061
p	.000	.000
RMSEA	-	.008
RMSEA unt. 95%iges KI	-	.004
RMSEA ober. 95%iges KI	-	.012
CFI	1.000	.969
TLI	-	.939
AIC	24	38.061
BIC	145	128.597
SRMR	-	.039

Im frei geschätzten Modell können aufgrund der exakten Passung des Modells keine Fit-Indizes, berechnet werden. Dies gilt jedoch nicht für das Constraintmodell, so dass für dieses Modell Fit-Indizes berechnet werden können. Der CFI von diesem Modell beträgt .969 und der RMSEA .008 (.004 - .012). Das SRMR beträgt .039. Dies entspricht einer guten Modellpassung.

Die Parameterschätzungen für die b-Koeffizienten des Constraint-Modells ähneln dem Modell über beide Gruppen hinweg. Der Koeffizient b31 für die Beziehung zwischen Religiosität und sozialer Unterstützung beträgt .125 (.086 - .163), der Haupteffekt b34 .172 (.146 - .198) und der Effekt b35 für die Beziehung zwischen sozialer Unterstützung und Lebenszufriedenheit .315 (.259 - .369). Der Chi-Quadrat-Differenzentest ergibt auch bei der zweiten Moderatorvariable „Wichtigkeit von Religion“ bei drei Freiheitsgraden eine signifikante Verschlechterung der Modellpassung des Constraint-Modells im Vergleich zum frei geschätzten Modell. Deshalb wird die Null-Hypothese von gleichen direkten Effekten zwischen den Subgruppen zurückgewiesen und ein moderierender Effekt der Variable Wichtigkeit von Religion angenommen.

Beide Kultur-Proxy-Variablen, also zum einen der „Human Development Index“ und zum anderen die „Wichtigkeit von Religion“ sind wirksam und haben einen Einfluss auf die Regressionskoeffizienten. Allerdings zeigen sich Unterschiede zwischen den beiden zur Moderatoranalyse verwendeten Verfahren. Wird im reduzierten HDI-Untersuchungsmodell nur einer der drei

Regressionskoeffizienten ( $b_{34}$ ) getestet, da im OSMASEM der Interaktionsterm als einziger signifikant ist, zeigt sich keine signifikante Verschlechterung des gleichgesetzten Modells. Die Teststärke reicht in diesem Fall nicht aus, um im Subgruppendesign einen Moderationseffekt auf den Haupteffekt  $b_{34}$  aufzudecken. Beim WR-Moderator gibt es ebenfalls Unterschiede zwischen den Verfahren. Während das TSSEM-Verfahren mit Subgruppendesign zeigt, dass die A-Matrix mit den darin befindlichen Regressionskoeffizienten in beiden Gruppen unterschiedlich ist, ist im OSMASEM keiner der drei Interaktionsterme statistisch signifikant, obwohl der Omnibustest eine signifikante Modellverbesserung anzeigt. Im vollständigen Modell beeinflusst der WR-Moderator jedoch auch andere Regressionskoeffizienten, die im reduzierten Modell nicht integriert sind ( $b_{31}$ ,  $b_{33}$  und  $b_{37}$ ).

## 9.5 Verfahren zur Berücksichtigung abhängiger Effektstärken

Im Anschluss an die Durchführung des Mittelwert-TSSEM-Ansatzes und den Moderatoranalysen wird ein Vergleich der verschiedenen Verfahren zur Zusammenfassung abhängiger Effektstärken vorgenommen. Dies kann als weitere Sensitivitätsanalyse, neben der Analyse des Grundmodells mittels OSMASEM und der Überprüfung der Moderatoren, der bislang gewonnenen Ergebnisse betrachtet werden.

Die bisherigen Analysen wurden auf Basis des Mittelwert-TSSEM-Verfahrens durchgeführt. Dem Ansatz von Wilson/Polanin/Lipsey (2016) folgend werden die Three-Level-Meta-Analyse mit dem TSSEM-Verfahren kombiniert, um zu überprüfen, ob es durch die Berücksichtigung aller vorliegenden Effektstärken zu anderen Ergebnissen kommt. Ebenfalls wird der robuste Ansatz nach Tanner-Smith/Tipton (2014) angewandt (siehe 9.5.1) und am Ende des Kapitels die Ergebnisse dieser drei unterschiedlichen Verfahren miteinander verglichen (siehe 9.5.2).

### 9.5.1 Wilson-Ansatz und robuster Ansatz

Als Ausgangsmodell der robusten Modellierungsstrategie wird das Three-Level-Modell des Wilson-Ansatzes gewählt.<sup>129</sup> Dazu wird zunächst mit dem R-Paket *metafor* ein Three-Level-Modell auf Studien- und Effektstärkeebene (K-ID und ESID) mit dem HDI-Datensatz geschätzt. Die Ergebnisse dieser Schätzung entsprechen den kumulierten Korrelationskoeffizienten aus der ersten Phase der TSSEM-Analyse sowie der asymptotischen Varianz-Kovarianz-Matrix, die für die Schätzung eines Strukturgleichungsmodells in der zweiten Phase notwendig ist. Das Ergebnis dieser Schätzung ist in Tabelle 53 zu sehen.

Über der Diagonale finden sich die kumulierten Effektstärken mit dem zugehörigen Standardfehler und unterhalb der Diagonale die Anzahl der Effektstärken, auf der die jeweilige Beziehung basiert.  $I^2$ -Werte können mithilfe dieses Ansatzes für die einzelnen Koeffizienten nicht berechnet werden. Dennoch wird ein Heterogenitätstest auf Basis der  $Q$ -Statistik durchgeführt. Dieser zeigt an, dass keine Homogenität der Korrelationsmatrix vorliegt, da die Nullhypothese mit  $Q$  ( $df = 446$ ) = 3392 mit .0001,  $p < 0.00$  verworfen wird.

---

<sup>129</sup> Diese Modellierungsstrategie basiert auf einer Korrespondenz mit Wolfgang Viechtbauer (Autor von *metafor*) über die E-Mailliste „R-sig-meta-analysis“ vom 19.12.2018. Seiner Empfehlung nach sollte auch eine robuste Modellierung mit möglichst realistischen Annahmen zum Modell starten, um ein effizientes Schätzmodell zu konstruieren.

**Tabelle 53: (Random-Effects-) Ergebnis der ersten Phase des Wilson-Ansatzes**

	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		0.150 (0.020)	0.151 (0.023)	0.136 (0.019)	0.197 (0.011)
Soz. Unterstützung	53		0.390 (0.050)	0.207 (0.079)	0.363 (0.025)
Optimismus	43	7		0.325 (0.043)	0.372 (0.034)
Selbstwert	62	2	12		0.356 (0.027)
Lebenszufriedenheit	219	31	16	23	

**Anmerkung:** n = 205.483; k = 456; Überhalb der Diagonale sind die kumulierten Korrelationskoeffizienten aus der ersten Phase sowie den zugehörigen Standardfehlern und unterhalb die Anzahl der Effektstärken je Beziehung zu sehen.

In Tabelle 54 ist das Ergebnis der ersten Phase, unter Verwendung des robusten Schätzers, zu sehen. Oberhalb der Diagonale befinden sich die kumulierten Effektstärken mit zugehörigem Standardfehler und unterhalb die Anzahl von Effektstärken je Beziehung.

**Tabelle 54: Erste Phase robuster Ansatz**

	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		0.150 (0.018)	0.151 (0.020)	0.136 (0.020)	0.197 (0.011)
Soz. Unterstützung	53		0.390 (0.028)	0.207 (0.019)	0.363 (0.040)
Optimismus	43	7		0.325 (0.051)	0.372 (0.053)
Selbstwert	62	2	12		0.356 (0.051)
Lebenszufriedenheit	219	31	16	23	

**Anmerkung:** n = 205.483; k = 456; Überhalb der Diagonale sind die kumulierten Korrelationskoeffizienten aus der ersten Phase sowie den zugehörigen Standardfehlern und unterhalb die Anzahl der Effektstärken je Beziehung zu sehen.

In Bezug auf die kumulierten Effektstärken lassen sich keine Unterschiede zwischen den Verfahren feststellen. Lediglich die Standardfehler unterscheiden sich zum Teil voneinander. Eine eindeutige Tendenz ist dabei jedoch nicht feststellbar. Die Anzahl der Effektstärken je Beziehung variiert von 1 bis 15, wobei der Mittelwert bei 2,59 und der Median bei 2 liegt. Auch hier sollte in der zweiten Analysephase ein Random-Effects-Modell geschätzt werden. Nun folgt zunächst ein Vergleich von Mittelwert-TSSEM und dem Wilson-Ansatz im Hinblick auf

die kumulierten Korrelationskoeffizienten und die Anzahl der Effektstärken für die entsprechenden Beziehungen (siehe Tabelle 55). Über der Diagonale befinden sich die Ergebnisse der ersten Phase des Wilson-Ansatzes und unterhalb der Diagonale die entsprechenden Ergebnisse des Mittelwert-TSSEM-Ansatzes. Jeweils sind dort in Klammern die Anzahl der Effektstärken für die jeweilige Beziehung abgetragen.

Die Ergebnisse der robusten Schätzung werden hier nicht ausgewiesen, da die kumulierten Effektstärken identisch mit den Ergebnissen des Wilson-Ansatzes sind. Beim Vergleich der Ansätze wird deutlich, dass sich die Anzahl der Effektstärken je Beziehung voneinander unterscheiden. Dies trifft vorwiegend auf die Beziehungen zwischen Religiosität und den Mediatoren sowie zwischen Religiosität und Lebenszufriedenheit zu. Im Gegensatz dazu sind die Anzahl der Effektstärken für die Kovarianzen zwischen den Mediatoren und die Beziehungen zwischen Mediatoren und Lebenszufriedenheit in beiden Ansätzen ähnlich. Deutliche Unterschiede in den kumulierten Korrelationskoeffizienten lassen sich nicht feststellen. Die betragsmäßig größeren Unterschiede entfallen auf die Beziehungen, bei denen die Anzahl der Effektstärken zwischen den beiden Verfahren sehr unterschiedlich ist.

**Tabelle 55: Vergleich erste Phase Wilson-Ansatz und Mittelwert-TSSEM-Ansatz**

	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		0.150 (53)	0.150 (41)	0.136 (60)	0.197 (219)
Soz. Unterstützung	0.128 (34)		0.390 (6)	0.207 (2)	0.363 (31)
Optimismus	0.151 (21)	0.384 (6)		0.325 (8)	0.372 (15)
Selbstwert	0.149 (32)	0.180 (2)	0.288 (8)		0.356 (21)
Lebenszufriedenheit	0.204 (125)	0.332 (29)	0.377 (15)	0.391 (21)	

**Anmerkung:**  $n = 205.483$ ;  $k$  (Mittelwert-TSSEM) = 293;  $k$  (Wilson et al. 2016 Ansatz) = 456. Über der Diagonale sind die Ergebnisse des Wilson-Ansatzes und unterhalb der Diagonale die Ergebnisse des Mittelwert-TSSEM-Ansatzes zu sehen. In Klammern sind jeweils die Anzahl der Effektstärken für die jeweilige Beziehung abgetragen.

Im nächsten Schritt wird nun die kumulierte Korrelationsmatrix sowie die asymptotische Varianz-Kovarianz-Matrix dazu genutzt, das bereits bekannte Untersuchungsmodell zu testen. In der nachfolgenden Tabelle 56 ist das Schätzergebnis des Untersuchungsmodells zu sehen.

Im Unterschied zum Mittelwert-TSSEM-Ansatz kommt es bei der vorliegenden Schätzung der Kovarianz  $k_2$  zu keinen Problemen, obwohl die Anzahl der Effektstärken, auf denen die Beziehung basiert, in beiden Modellen gleich ist. Möglicherweise ist der Wilson-Ansatz weniger

anfällig für eine niedrige Anzahl von Effektstärken je Beziehung. Zur Vergleichbarkeit wird ein Modell ohne die Kovarianz k2 spezifiziert. Beide Modelle unterschieden sich kaum voneinander. Das reduzierte Modell weist einen Freiheitsgrad auf, so dass dort Fit-Indizes berechnet werden können. Es liegt eine gute Passung des Modells vor. Der CFI beträgt .998, der RMSEA .0047 (.0016 - .0088) und das SRMR .0591.

Im vollständigen Modell mit Kovarianz k2 betragen die indirekten Effekte von sozialer Unterstützung .034 (.021 - .050), von Optimismus .029 (.014 - .047) und von Selbstwert .032 (.020 - .046). Im reduzierten Modell ohne Kovarianz k2 beträgt der indirekte Effekt von sozialer Unterstützung .041(.028 - .059), von Optimismus .023 (.008 - .041) und Selbstwert .038 (.025 - .054). In Abbildung 32 ist das Ergebnis dieser Schätzung als Pfadmodell zu sehen.

**Tabelle 56: Gesamtmodell mit k2 und ohne k2 (Wilson-Ansatz)**

	Mit k2	Ohne k2
b31	0.1498 [0.1111; 0.1886]	0.1491 [0.1103; 0.1879]
b32	0.1505 [0.1059; 0.1951]	0.1499 [0.1052; 0.1945]
b33	0.1363 [0.0986; 0.1740]	0.1357 [0.0980; 0.1735]
b34	0.1031 [0.0757; 0.1292]	0.0933 [0.0658; 0.1197]
b35	0.2247 [0.1485; 0.2997]	0.2782 [0.2105; 0.3456]
b36	0.1928 [0.0964; 0.2812]	0.1568 [0.0548; 0.2497]
b37	0.2324 [0.1621; 0.3031]	0.2817 [0.2175; 0.3461]
p22	0.9775 [0.9644; 0.9877]	0.9778 [0.9647; 0.9878]
k1	0.3678 [0.2717; 0.4638]	0.3673 [0.2713; 0.4634]
k2	0.1865 [0.0317; 0.3412]	
p33	0.9774 [0.9619; 0.9888]	0.9775 [0.9622; 0.9889]
k3	0.3049 [0.2227; 0.3871]	0.3040 [0.2219; 0.3862]
p44	0.9814 [0.9697; 0.9903]	0.9816 [0.9699; 0.9904]
p55	0.7437 [0.6941; 0.7874]	0.7249 [0.6741; 0.7708]
Ind1 (b31*b35)	0.0337 [0.0206; 0.0500]	0.0415 [0.0276; 0.0586]
Ind2 (b32*b36)	0.0290 [0.0141; 0.0469]	0.0235 [0.0082; 0.0409]
Ind3 (b33*b37)	0.0317 [0.0198; 0.0465]	0.0382 [0.0253; 0.0540]
Chi-Quadrat-Veränderung	-	5.5753
DF	-	1.0000
p	-	0.0182
RMSEA	-	0.0047
RMSEA unt. 95%iges KI	-	0.0016
RMSEA ober. 95%iges KI	-	0.0088
CFI	-	0.9929
SRMR	-	0.0591
Fallzahl	205.483	205.483

In der nachfolgenden Tabelle 57 ist das Ergebnis der zweiten Phase des robusten Ansatzes zu sehen. Unterschiede lassen sich dabei vor allem bei der Kovarianz k2 ausmachen. Wird die Kovarianz aus dem Modell entfernt, verschlechtert sich das Modell, so dass der Koeffizient b37 (Effekt von Selbstwert auf Lebenszufriedenheit) nicht signifikant und negativ wird. Auf Basis der Fit-Indizes müsste das Modell mit einem Freiheitsgrad (ohne k2) verworfen werden. Die

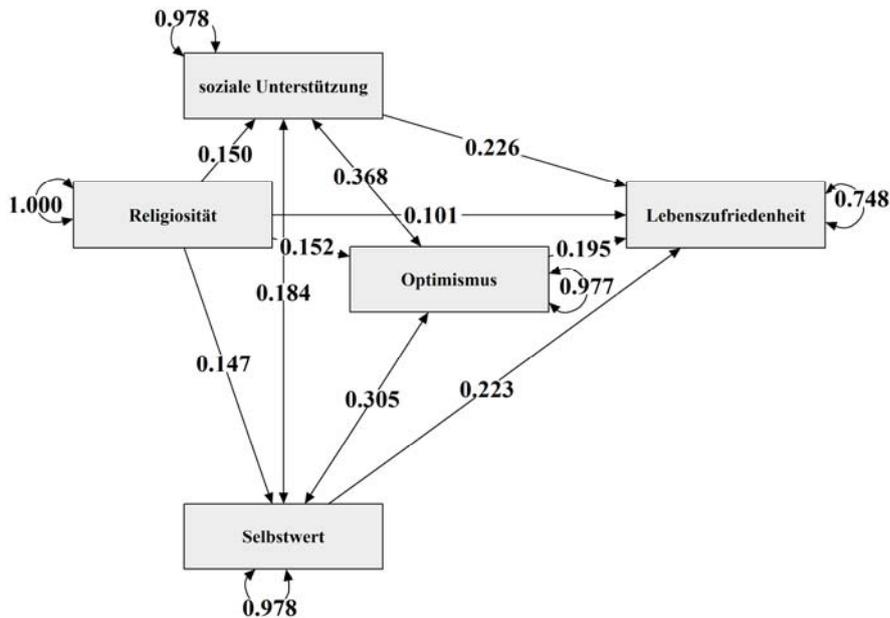
Punktschätzer der beiden Ansätze sind beinahe identisch. Die Unterschiede liegen in den geschätzten LBCIs, die sich teilweise stark voneinander unterscheiden. Auch der indirekte Effekt Ind1 wird nicht signifikant, da das LBCI die „0“ einschließt.

**Tabelle 57: Gesamtmodell mit k2 und ohne k2 (robuster Ansatz)**

	<b>Robuster Ansatz mit k2</b>	<b>Robuster Ansatz ohne k2</b>
b31	0.1498 [0.1137; 0.1860]	0.1792 [0.1410; 0.2173]
b32	0.1505 [0.1120; 0.1890]	0.1425 [0.1027; 0.1823]
b33	0.1363 [0.0974; 0.1752]	0.1235 [0.0820; 0.1650]
b34	0.1031 [0.0728; 0.1315]	0.1694 [0.1403; 0.1993]
b35	0.2247 [0.1211; 0.3262]	0.0159 [-0.0833; 0.1107]
b36	0.1928 [0.0581; 0.3224]	0.2277 [0.0996; 0.3575]
b37	0.2324 [0.1342; 0.3325]	-0.1037 [-0.1761; -0.0348]
p22	0.9775 [0.9654; 0.9871]	0.9679 [0.9528; 0.9801]
k1	0.3678 [0.3134; 0.4221]	0.3622 [0.3080; 0.4163]
k2	0.1865 [0.1500; 0.2229]	
p33	0.9774 [0.9643; 0.9874]	0.9797 [0.9668; 0.9894]
k3	0.3049 [0.2056; 0.4042]	0.2493 [0.1504; 0.3481]
p44	0.9814 [0.9693; 0.9905]	0.9848 [0.9728; 0.9933]
p55	0.7437 [0.6550; 0.8154]	0.9107 [0.8552; 0.9447]
Ind1 (b31*b35)	0.0337 [0.0200; 0.0510]	0.0029 [-.0165; 0.0179]
Ind2 (b32*b36)	0.0290 [0.0089; 0.0509]	0.0324 [0.0139; 0.0553]
Ind3 (b33*b37)	0.0317 [0.0167; 0.0510]	-0.0128 [-0.0237; -0.0430]
Chi-Quadrat-Veränderung	-	101.1027
DF	-	1.0000
p	-	0.0000
RMSEA	-	0.0221
RMSEA unt. 95%iges KI	-	0.0186
RMSEA ober. 95%iges KI	-	0.0258
CFI	-	0.8570
SRMR	-	0.1595
Fallzahl	205.483	205.483

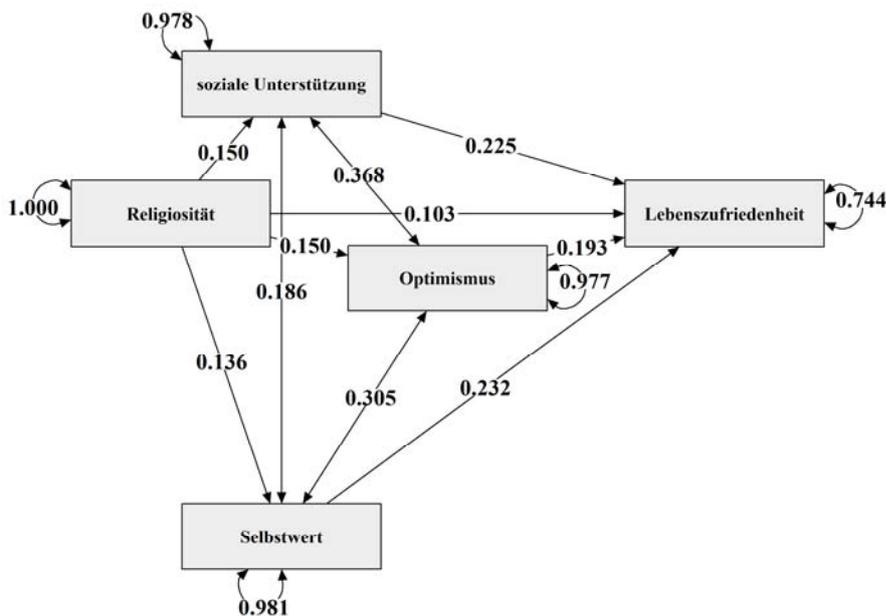
In den nachfolgenden Abbildungen (Abbildung 32 und Abbildung 33) sind zur graphischen Veranschaulichung die beiden Modelle mit k2 zu sehen.

**Abbildung 32: Zweite Phase Schätzergebnis des Wilson-Ansatzes (mit k2)**



**Anmerkung:** Eigene Abbildung; n = 205.483.

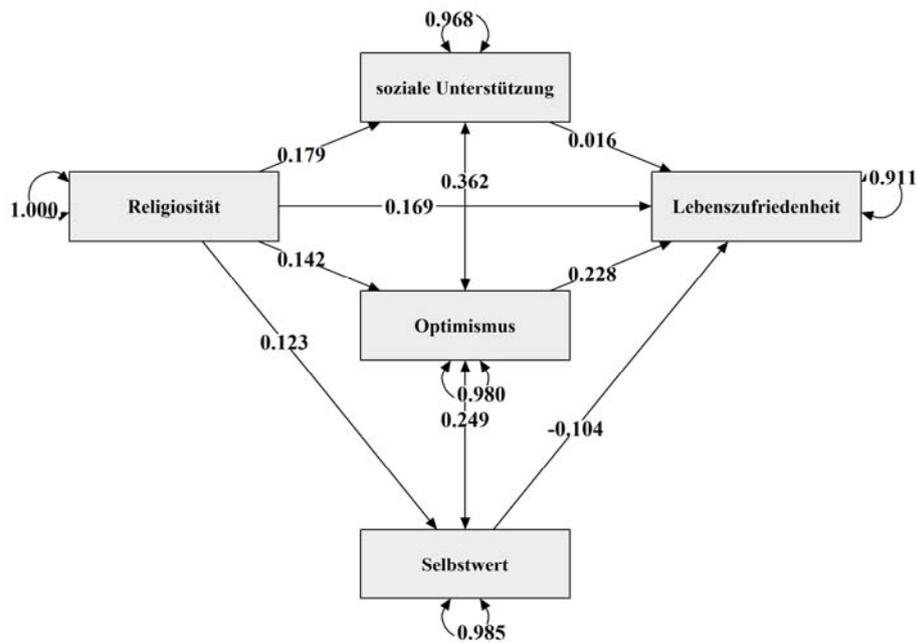
**Abbildung 33: Zweite Phase Schätzergebnis des robusten Ansatzes (mit k2)**



**Anmerkung:** Eigene Abbildung; n = 205.483.

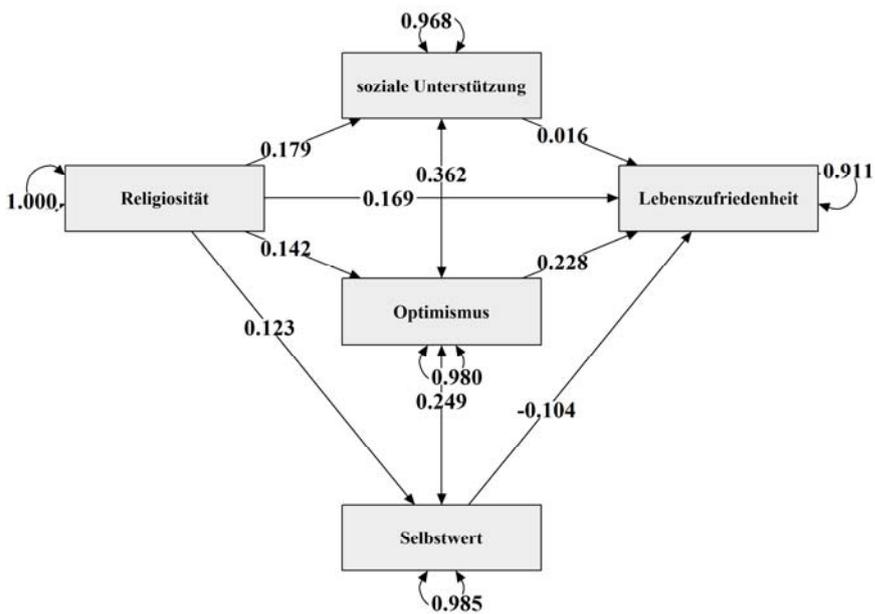
Daran anknüpfend ist in Abbildung 34 das gleiche Modell auf Basis des Wilson-Ansatzes, lediglich ohne die Kovarianz k2, abgebildet. Die Unterschiede der Koeffizienten lassen sich bei der zweiten oder dritten Nachkommastelle finden. Die statistische Varianzaufklärung der abhängigen Variable Lebenszufriedenheit des Modells mit k2 beträgt 27 % und beim Modell ohne k2 sogar 29 %. In Abbildung 35 ist zum direkten Vergleich das Modell aus dem robusten Ansatz ohne k2 zu sehen.

Abbildung 34: Zweite Phase Schätzergebnis des Wilson-Ansatzes ohne k2



Anmerkung: Eigene Abbildung; n = 205.483.

Abbildung 35: Zweite Phase Schätzergebnis des robusten Ansatzes (ohne k2)



Anmerkung: Eigene Abbildung; n = 205.483.

## 9.5.2 Vergleich der Ansätze

In Tabelle 58 folgt der Vergleich der Schätzergebnisse der zweiten TSSEM-Phase der drei unterschiedlichen Ansätze. Dabei lassen sich leichte Unterschiede feststellen. Die Fit-Indizes können nicht miteinander verglichen werden, da das Untersuchungsmodell jeweils genau identifiziert ist.

**Tabelle 58: Vergleich der Ansätze (Zweite Phase mit k2)**

	Mittelwert-TSSEM	Wilson-Ansatz	Robuster Ansatz
b31	0.1277 [0.0882; 0.1672]	0.1498 [0.1111; 0.1886]	0.1498 [0.1137; 0.1860]
b32	0.1513 [0.1065; 0.1961]	0.1505 [0.1059; 0.1951]	0.1505 [0.1120; 0.1890]
b33	0.1485 [0.1032; 0.1938]	0.1363 [0.0986; 0.1740]	0.1363 [0.0974; 0.1752]
b34	0.1071 [0.0704; 0.1407]	0.1031 [0.0757; 0.1292]	0.1031 [0.0728; 0.1315]
b35	0.1874 [0.0822; 0.2948]	0.2247 [0.1485; 0.2997]	0.2247 [0.1211; 0.3262]
b36	0.2077 [0.0721; 0.3322]	0.1928 [0.0964; 0.2812]	0.1928 [0.0581; 0.3224]
b37	0.2817 [0.1504; 0.4113]	0.2324 [0.1621; 0.3031]	0.2324 [0.1342; 0.3325]
p22	0.9837 [0.9721; 0.9922]	0.9775 [0.9644; 0.9877]	0.9775 [0.9654; 0.9871]
k1	0.3643 [0.3094; 0.4192]	0.3678 [0.2717; 0.4638]	0.3678 [0.3134; 0.4221]
k2	0.1614 [-0.0371; 0.3598]	0.1865 [0.0317; 0.3412]	0.1865 [0.1500; 0.2229]
p33	0.9771 [0.9616; 0.9887]	0.9774 [0.9619; 0.9888]	0.9774 [0.9643; 0.9874]
k3	0.2659 [0.1572; 0.3746]	0.3049 [0.2227; 0.3871]	0.3049 [0.2056; 0.4042]
p44	0.9779 [0.9624; 0.9893]	0.9814 [0.9697; 0.9903]	0.9814 [0.9693; 0.9905]
p55	0.7274 [0.6376; 0.7967]	0.7437 [0.6941; 0.7874]	0.7437 [0.6550; 0.8154]
Ind1 (b31*b35)	0.0315 [0.0191; 0.0469]	0.0337 [0.0206; 0.0500]	0.0337 [0.0200; 0.0510]
Ind2 (b32*b36)	0.0269 [0.0059; 0.0501]	0.0290 [0.0141; 0.0469]	0.0290 [0.0089; 0.0509]
Ind3 (b33*b37)	0.0501 [0.0274; 0.0780]	0.0317 [0.0198; 0.0465]	0.0317 [0.0167; 0.0510]
Fallzahl	205.483	205.483	205.483

Dennoch lassen sich die geschätzten Koeffizienten sowie deren LBCIs vergleichen. Grundsätzliche Unterschiede in der Tendenz der Ergebnisse sind nicht feststellbar, außer bei der Kovarianz k2, die im Mittelwert-TSSEM-Ansatz nicht signifikant ist, da dort das LBCI die „0“ einschließt. Es zeigt sich jedoch, dass die LBCIs bei den Beziehungen, die auf wenigen Effektstärken beruhen, wie der Effekt von Selbstwert zu Lebenszufriedenheit (b37), im Wilson-Ansatz ein deutlich schmaleres Konfidenzintervall als im Mittelwert-TSSEM-Ansatz haben, obwohl die Beziehung im Wilson-Ansatz lediglich auf Basis von zwei weiteren Effektstärken berechnet wird. Ähnlich verhält es sich auch bei den Regressionskoeffizienten b31, b35 und b36. In allen genannten Fällen sind die LBCIs im Wilson-Ansatz schmalere und das Ergebnis somit präziser. Bei b32, dem Effekt von Religiosität auf Optimismus, und b33, dem Effekt von Religiosität auf Selbstwert, sind hingegen trotz unterschiedlicher Anzahl an Effektstärken keine offensichtlichen Unterschiede zwischen den Ansätzen erkennbar.

Auch im Vergleich zum robusten Ansatz ergeben sich Unterschiede bezüglich der Konfidenzintervalle. Die Punktschätzer entsprechen den Wilson-Ergebnissen. Bei den Effekten b35, b36 und b37 sind die Konfidenzintervalle des robusten Ansatzes, im Vergleich zum Wilson-Ansatz, breiter und entsprechen annähernd den LBCIs des Mittelwert-TSSEM-Ansatzes. Bei der Kovarianz k1 ist das LBCI im Wilson-Ansatz sogar breiter als im herkömmlichen TSSEM-Ansatz. Bei k2 ist das LCBI im Wilson-Ansatz vollständig im positiven Bereich und deutlich schmaler als im Mittelwert-TSSEM-Modell. Bei der Kovarianz k3 verhält es sich ähnlich. Die vier geschätzten Residuen unterscheiden sich zwischen den Modellen minimal, aber ohne erkennliche Tendenz, was die Höhe der Koeffizienten betrifft. Die LBCIs der geschätzten Residuen sind im Wilson-Ansatz und im robusten Ansatz ebenfalls schmaler.

Auch die geschätzten indirekten Effekte können zwischen den drei Methoden zur Zusammenfassung abhängiger Effektstärken verglichen werden. Im Mittelwert-TSSEM-Verfahren beträgt der indirekte Effekt und sein 95%iges LBCI von Religiosität über soziale Unterstützung auf Lebenszufriedenheit .032 (.019 - .047), der indirekte Effekt von Religiosität über Optimismus auf Lebenszufriedenheit .026 (.006 - .050). und der indirekte Effekt von Religiosität über Selbstwert .048 (.027 - .074). Im reduzierten Modell, ohne die Kovarianz k2 auf Basis des Wilson-Ansatzes, beträgt der indirekte Effekt von sozialer Unterstützung .041 (.028 - .059), von Optimismus .023 (.008 - .041) und von Selbstwert .038 (.025 - .054). Bei beiden Verfahren zeigen die signifikanten indirekten Effekte eine partielle Mediation an. Die statistische Varianzaufklärung der abhängigen Variable Lebenszufriedenheit ist im Mittelwert-TSSEM-Ansatz (27,3 %) etwas höher als im Wilson-Verfahren bzw. dem robusten Ansatz (25,6 %). Die Unterschiede zwischen Wilson-Ansatz und robustem Ansatz sind in Bezug auf die indirekten Effekte sehr gering. Auch die Ergebnisse des robusten Ansatzes zeigen partielle Mediationen an. Insgesamt betrachtet scheint die Verwendung des Wilson-Ansatzes vor allem dann relevant zu sein, wenn für bestimmte Beziehungen des Modells nur sehr wenige Effektstärken oder spezifischere Forschungsfragen bezüglich der Messinstrumente vorliegen. Wenn es jedoch um die Generalisierung von Effekten, wie in der vorliegenden Arbeit, geht, ist dies nicht notwendig. Grundsätzliche, auf das Untersuchungsmodell und seine inhaltlichen Aussagen wie etwa die Signifikanz einzelner Effekte bezogene Unterschiede lassen sich zwischen den Verfahren nicht finden. Empfehlungen zur Verwendung eines konkreten Verfahrens können auf Basis der hier gewonnen Erkenntnisse nur zurückhaltend formuliert werden. Zur Entscheidungsfindung sollten zukünftige Anwender der TSSEM-Methode genau überlegen, ob sie den zusätzlichen Aufwand beim Datenhandling, die der Wilson-Ansatz mit sich bringt, in Kauf nehmen wollen. Bei einer ausreichenden Anzahl von Effektstärken für jede Beziehung scheint es aus Sicht dieses

Vergleichs einen leichten Vorteil für den Mittelwert-TSSEM-Ansatz zu geben, da dieser in der Anwendung leichter ist. Lediglich bei konkreten Forschungsfragen bezüglich der Messinstrumente und/oder wenn unterschiedliche Fallzahlen für unterschiedliche Korrelationen berichtet werden (für eine potenzielle Gewichtung), könnte es sein, dass der Wilson-Ansatz bessere Ergebnisse liefert. Anhand der Unterschiede bezüglich der Kovarianz  $k^2$  zeigt sich die Notwendigkeit, verschiedene Verfahren im Sinne einer Sensitivitätsanalyse einzusetzen.

Durch die Berücksichtigung aller Effektstärken kann die Präzision der Schätzung jedoch erhöht werden. Schon ein Unterschied von wenigen Effektstärken für spezifische Beziehungen hat einen relativ großen Einfluss auf das LBCI. Die Leistungsfähigkeit des Wilson-Ansatzes könnte noch größer sein, wenn Studien unterschiedliche Fallzahlen für die einzelnen Beziehungen berichten, da der Ansatz eine Gewichtung anhand der Fallzahl bei der Zusammenfassung der Effektstärken vornimmt. In der vorliegenden Studie haben die erhobenen Studien für die einzelnen Beziehungen bzw. Effektstärken jedoch keine Fallzahlen angegeben. Der robuste Ansatz hingegen kann keine weitere Verbesserung des Modells bewirken. Es scheint hingegen, dass die LBCIs unter Verwendung des robusten Schätzers breiter werden und die Präzision der Ergebnisse dadurch geringer wird. Der große Vorteil des robusten Ansatzes könnte sich allerdings auch erst bei multiplen Abhängigkeiten zeigen.

Die Erkenntnisse dieser Arbeit bezüglich der Verfahren zur Berücksichtigung abhängiger Effektstärken sollten in weiteren Anwendungen und Simulationsstudien systematisch überprüft und ausgeweitet werden. Besonders relevant ist es, die unterschiedlichen Verfahren zum Umgang mit abhängigen Effektstärken (Multiple-Endpoints) im Kontext des OSMASEM zu evaluieren, da sie dort noch nicht behandelt wurden.

## 9.6 Überprüfung des Publikationsbias

Nun werden die meta-analytischen Ergebnisse im Hinblick auf das Vorliegen eines Publikationsbias überprüft. Da es bislang keine etablierten Methoden zur Überprüfung des Publikationsbias für modellbasierte Meta-Analysen gibt, wird die Überprüfung anhand der bivariaten Beziehungen durchgeführt. Für jede Beziehung wird auf Grundlage des Mittelwert-TSSEM-Datensatzes ein Random-Effects-Modell in metafor zur Schätzung der Effektstärken spezifiziert. Anschließend werden die Koeffizienten z-transformiert und fünf unterschiedliche Verfahren zur Überprüfung des Publikationsbias durchgeführt. Zunächst werden die mittels Trim-and-Fill-Methode erweiterten Funnel-Plots mit 95%igem Konfidenzintervall betrachtet. Dann folgen die fail-safe-Ns (gewichtete Variante) sowie die Rangkorrelationsmethode. Den Abschluss des Kapitels bildet der Egger-Regressionstest. In der nachfolgenden Tabelle 59 sind die (bivariat) kumulierten Korrelationskoeffizienten und die Anzahl der Effektstärken, auf der die jeweilige Beziehung beruht, zu sehen. Vergleicht man die bivariat geschätzten kumulierten Korrelationskoeffizienten (oberhalb der Diagonale) mit den multivariat geschätzten Koeffizienten aus der ersten Phase des TSSEM-Verfahrens (siehe Tabelle 34) zeigen sich keine grundlegenden Unterschiede. Unterhalb der Diagonale ist die Anzahl von Effektstärke je Beziehung zu sehen.

**Tabelle 59: Bivariat kumulierte Korrelationskoeffizienten**

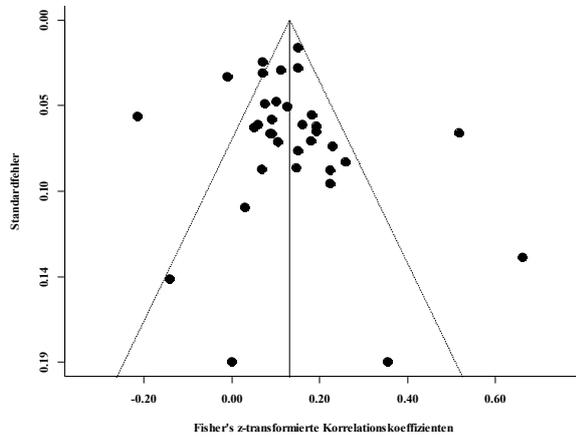
	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		.131	.154	.153	.211
Soz. Unterstützung	34		.413	.195	.357
Optimismus	21	6		.304	.414
Selbstwert	32	2	8		.435
Lebenszufriedenheit	125	29	15	21	

**Anmerkung:** Über der Diagonale sind die bivariat kumulierten Korrelationskoeffizienten und unterhalb der Diagonale die jeweilige Anzahl der Effektstärken zu sehen.

Das Trim-and-Fill-Verfahren wird an dieser Stelle mit der metafor Spezifikation „Null“ durchgeführt. Es wird zunächst keine Seite festgelegt, auf der „fehlende“ Studien eingesetzt werden. Die neu eingesetzten Studien sind als weiße Kreise in den Abbildungen zu sehen. Der kumulierte Korrelationskoeffizient wird dabei für jede Beziehung vor und nach dem Trim-and-Fill-Verfahren miteinander verglichen, sofern das Verfahren weitere Studien einfügt. In Abbildung

36 ist der Funnel-Plot der Zelle 1, also der Beziehung von Religiosität und sozialer Unterstützung, zu sehen. Trim-and-Fill schlägt für diese Beziehung keine weiteren Studien vor. Es sind zwar Ausreißer außerhalb des Trichters zu erkennen, aber augenfällige Tendenzen lassen sich dabei nicht feststellen. Es liegt eine symmetrische Verteilung der Effektstärken vor.

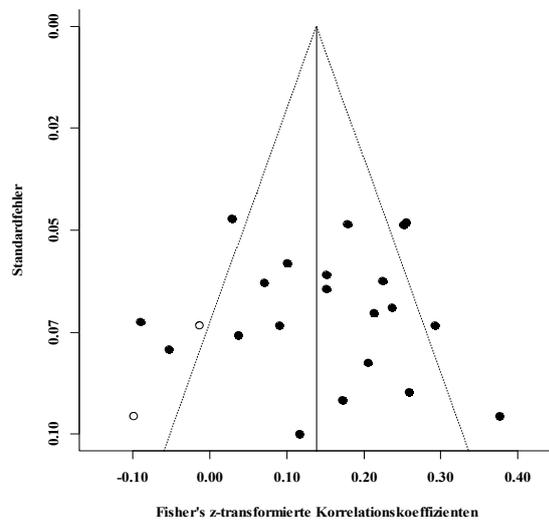
**Abbildung 36: Funnel-Plot Zelle 1**



**Anmerkung:** Eigene Abbildung; 34 Effektstärken; Trim-and-Fill-Methode fügt keine Studien ein.

In der zweiten Zelle hingegen werden zwei weitere Studien auf der linken Seite eingefügt (siehe Abbildung 37).

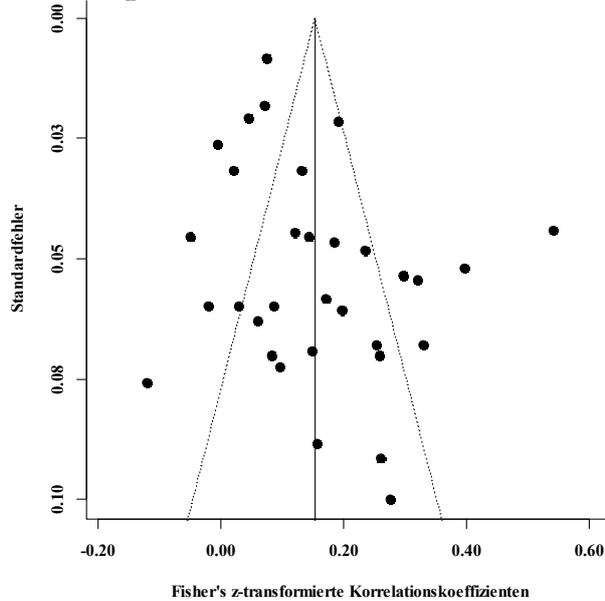
**Abbildung 37: Funnel-Plot Zelle 2**



**Anmerkung:** Eigene Abbildung; 21 Effektstärken; Trim-and-Fill-Methode fügt zwei Studien auf der linken Seite ein.

Durch diese zusätzlichen Studien würde sich die kumulierte Effektstärke von .153 auf .137 verringern. Dies ist ein Hinweis, dass ein leichter Publikationsbias vorliegt. Insgesamt liegt dennoch eine relativ symmetrische Verteilung vor.

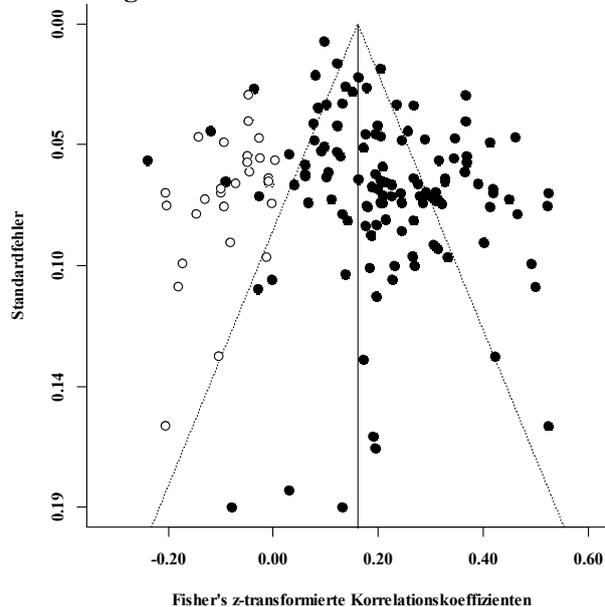
**Abbildung 38: Funnel-Plot Zelle 3**



**Anmerkung:** Eigene Abbildung; 32 Effektstärken; Trim-and-Fill-Methode fügt keine Studien ein.

Die Beziehung zwischen Religiosität und Optimismus (Zelle 3) ist ebenfalls unbedenklich (siehe Abbildung 38). Interessanterweise zeigt sich in Abbildung 39 für Zelle 4 der Korrelationsmatrix, also der Hauptbeziehung zwischen Religiosität und Lebenszufriedenheit, eine leichte Asymmetrie in der Verteilung.

**Abbildung 39: Funnel-Plot Zelle 4**



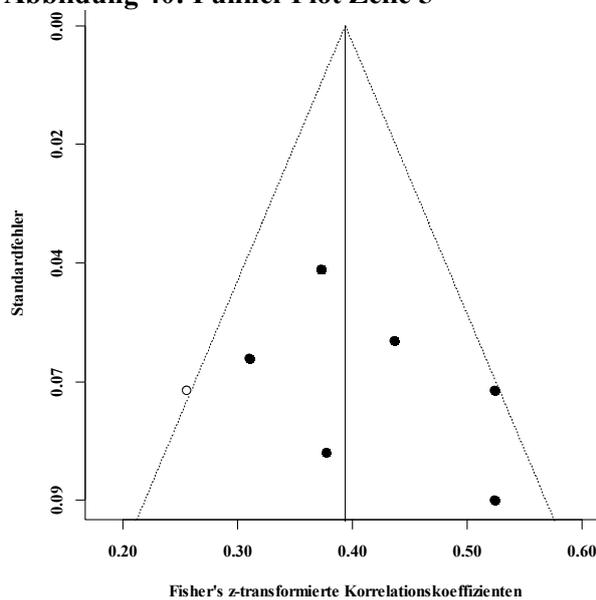
**Anmerkung:** Eigene Abbildung; 125 Effektstärken; Trim-and-Fill-Methode fügt 27 Studien auf der linken Seite ein.

Auf der linken Seite sind weniger Studien zu finden, als bei einer symmetrischen Verteilung zu erwarten wären. Das Trim-and-Fill-Prozedere fügt auf dieser Seite entsprechend 27 neue Studien ein. Die Hinzunahme dieser Studien würde zu einer Verringerung der Effektstärke von

.209 auf .159 führen. Dies ist ein Hinweis dafür, dass vermutlich Studien mit einem niedrigen bzw. negativen Effekt weniger publiziert bzw. in der Literatursuche gefunden wurden. Die kumulierte Effektstärke ist somit leicht nach oben verzerrt.

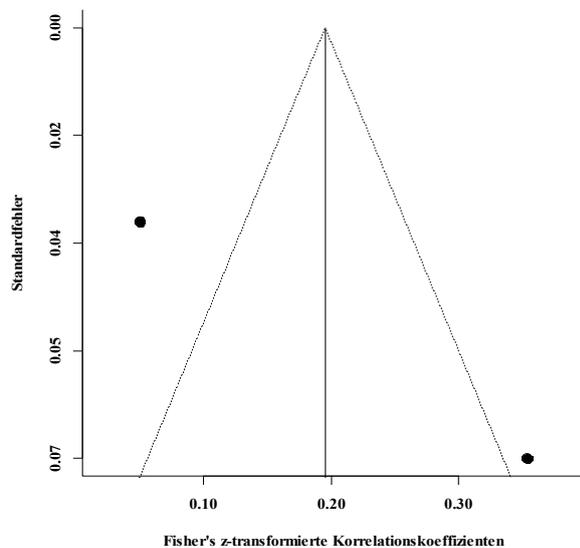
In Abbildung 40 ist Zelle 5 der Korrelationsmatrix, also die Beziehung der Variablen soziale Unterstützung und Optimismus, zu sehen. Die wenigen vorliegenden Studien befinden sich innerhalb des 95%igen Konfidenzintervalls. Eine Asymmetrie lässt sich in der Verteilung der Studien nicht erkennen. Trim-and-Fill fügt auf der linken Seite eine weitere Studie hinzu. Durch die Hinzunahme dieser Studie würde die kumulierte Effektstärke von .391 auf .374 sinken. Da dieser Plot jedoch auf sehr wenigen Effektstärken beruht, sollte dies nicht überbewertet werden.

**Abbildung 40: Funnel-Plot Zelle 5**



**Anmerkung:** Eigene Abbildung; 6 Effektstärken; Trim-and-Fill-Methode fügt 1 Studie auf der linken Seite ein.

**Abbildung 41: Funnel-Plot Zelle 6**

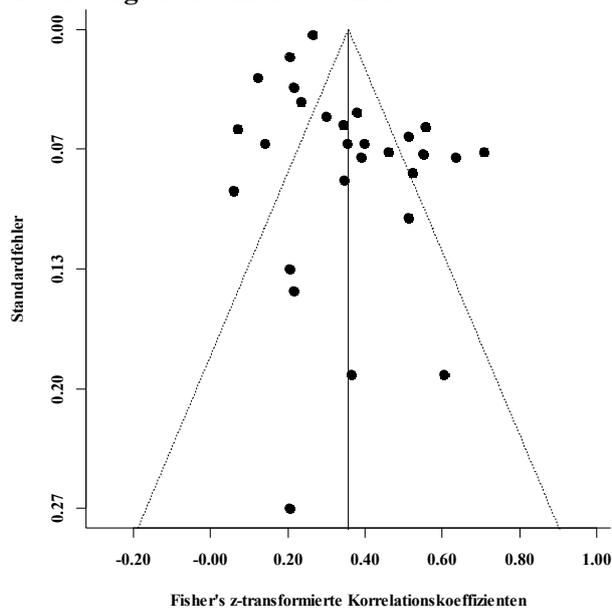


**Anmerkung:** Eigene Abbildung; 2 Effektstärken; Trim-and-Fill-Methode kann nicht durchgeführt werden.

Ähnlich verhält es sich mit der Beziehung zwischen sozialer Unterstützung und Selbstwert (Kovarianz k2), die auf lediglich zwei Studien beruht. In Abbildung 41 ist trotz sehr geringer Studienanzahl der Funnel-Plot abgetragen. Eine Beurteilung der Symmetrie kann nicht durchgeführt werden. Die Trim-and-Fill-Methode kann nicht ausgeführt werden, da nicht genügend empirische Informationen vorliegen.

Die Beziehung zwischen sozialer Unterstützung und Lebenszufriedenheit basiert auf mehr Studien (siehe Abbildung 42).

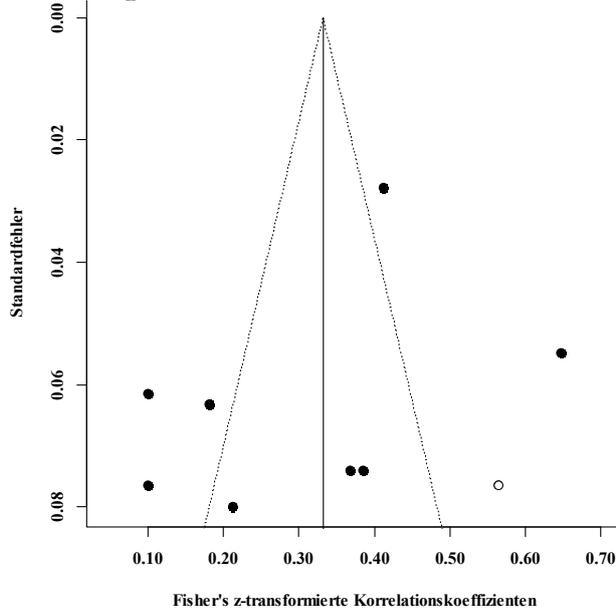
**Abbildung 42: Funnel-Plot Zelle 7**



**Anmerkung:** Eigene Abbildung; 29 Effektstärken; Trim-and-Fill-Methode fügt keine Studien ein.

Zwar befinden sich einige Studien außerhalb des Konfidenzintervalls, Asymmetrie ist optisch jedoch nicht feststellbar. Auch das Trim-and-fill-Verfahren berechnet keine weiteren Studien, so dass ein Publikationsbias auf Grundlage des Funnel-Plots bei dieser Beziehung unwahrscheinlich ist. In Abbildung 43 ist der Funnel-Plot der Beziehung (Kovarianz k3) zwischen Optimismus und Selbstwert zu sehen (Zelle 8). Durch die geringe Anzahl an Studien sind Aussagen über die Symmetrie der Verteilung schwierig. Die Trim-and-Fill-Methode fügt eine Studie auf der rechten Funnelseite ein, so dass die kumulierte Effektstärke leicht von .295 auf .320 ansteigt. Dieses Ergebnis sollte jedoch nicht überinterpretiert werden.

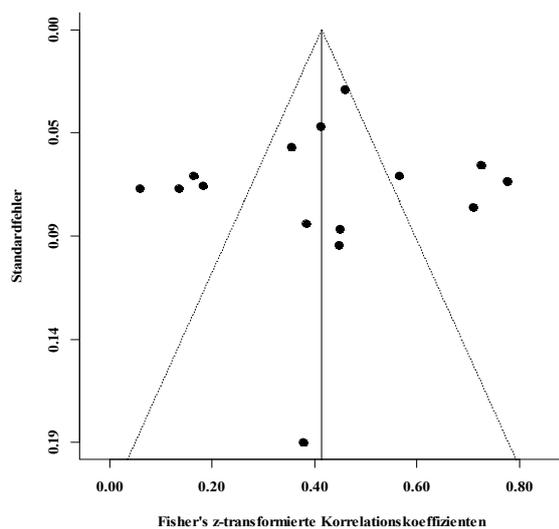
**Abbildung 43: Funnel-Plot Zelle 8**



**Anmerkung:** Eigene Abbildung; 8 Effektstärken; Trim-and-Fill-Methode fügt eine Studie auf der rechten Seite ein.

Abbildung 44 zeigt den Funnel-Plot der Beziehung zwischen Optimismus und Lebenszufriedenheit (Zelle 9). Es befinden sich einige Effektstärken außerhalb des Konfidenzintervalls, eine Asymmetrie ist allerdings visuell nicht feststellbar. Der Trim-and-Fill-Schätzer findet ebenfalls keine fehlenden Studien, so dass das Vorliegen eines Publikationsbias auf Basis der visuellen Inspektion unwahrscheinlich ist.

**Abbildung 44: Funnel-Plot Zelle 9**

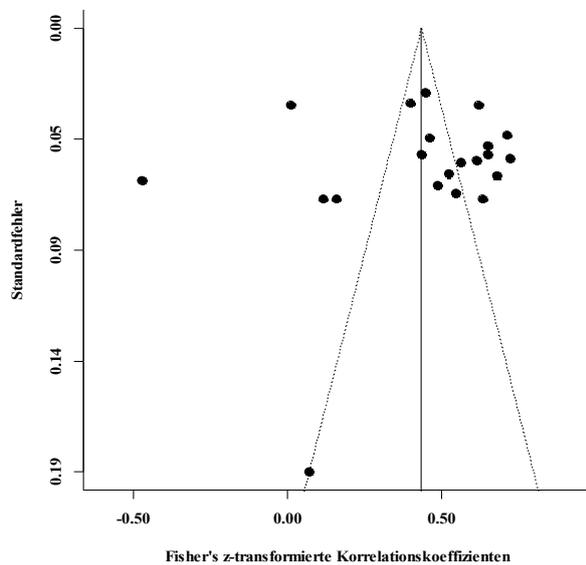


**Anmerkung:** Eigene Abbildung; 15 Effektstärken; Trim-and-Fill-Methode fügt keine Studien ein.

Der Funnel-Plot der Beziehung zwischen Selbstwert und Lebenszufriedenheit (Zelle 10) ist in

Abbildung 45 zu sehen. Optisch zeigt sich eine leichte Asymmetrie. Es scheint, als ob einige Studien auf der linken Seite fehlen. Das Trim-and-Fill-Verfahren fügt jedoch keine weiteren Studien ein, so dass es unwahrscheinlich ist, dass ein Publikationsbias vorliegt.

**Abbildung 45: Funnel-Plot Zelle 10**



**Anmerkung:** Eigene Abbildung; 21 Effektstärken; Trim-and-Fill-Methode fügt keine Studien ein.

Nach der visuellen Inspektion der Funnel-Plots und der Betrachtung der Auswirkungen des Trim-and-Fill-Verfahrens folgen nun die formellen statistischen Verfahren zur Überprüfung eines Publikationsbias. In der nachfolgenden Tabelle 60 ist eine Übersicht über Rosenbergs (gewichtete) fail-safe-Ns unterhalb der Diagonale und Begg/Mazumdar Rangkorrelationsmethode über der Diagonale zu sehen.

**Tabelle 60: Rosenbergs fail-safe Ns und Begg/Mazumdar Rangkorrelation**

	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		.193	.976	.338	.294
Soz. Unterstützung	2.023		.469	1.000	.320
Optimismus	669	388		.383	.692
Selbstwert	2.721	5	799		.193
Lebenszufriedenheit	58.689	55.571	3.098	9.908	

**Anmerkung:** Überhalb der Diagonale ist der Rangkorrelationskoeffizient nach Begg/Mazumdar und unterhalb der Diagonale Rosenbergs fail-safe-Ns zu sehen. Die grau hinterlegten Wert sollten inhaltlich nicht interpretiert werden (siehe 5.5.1).

Die Rosenbergs fail-safe-N-Methode dient lediglich zur Illustration und beschreibt, wieviel nicht signifikante Studien in die Meta-Analyse eingeschlossen werden müssten, um den Gesamteffekt nicht signifikant werden zu lassen. Der Wert ist für die schwach besetzten Zellen

eher gering, während er für die häufiger untersuchten Beziehungen deutlich höhere Werte aufweist. Vor allem die Kovarianz  $k_2$ , also die Beziehung zwischen sozialer Unterstützung und Selbstwert, hat einen problematisch niedrigen Wert. Die Beziehungen, die auf mehr Effektstärken beruhen, haben einen deutlich höheren Wert. Oberhalb der Diagonale sind die p-Werte der Rangkorrelation (Kendalls Tau) zwischen Effektstärke und Studiengröße abgetragen. Wie dort zu sehen ist, ist keiner der Werte signifikant ( $p < .05$ ), was gegen das Vorliegen eines Publikationsbias spricht. Die grau hinterlegten Felder dürfen nicht interpretiert werden, da für die entsprechenden Beziehungen keine ausreichende Anzahl von Effektstärken ( $>40$ ) aus unabhängigen Stichproben vorliegt.

Darüber hinaus sind die Ergebnisse des Egger-Tests mit den zugehörigen p-Werten der jeweiligen Regressionsschätzung in Tabelle 61 zu sehen. Die grau hinterlegten Felder markieren die Regressionsschätzungen, für die zu wenige Effektstärken/Stichproben vorliegen ( $<17$ ). Es liegt in keinem Fall ein signifikantes Testergebnis vor, so dass auch hier keine bedeutenden Anzeichen eines Publikationsbias zu finden sind.

**Tabelle 61: Egger-Testergebnisse**

	Religiosität	Soz. Unterstützung	Optimismus	Selbstwert	Lebenszufriedenheit
Religiosität		.172	.820	.253	.064
Soz. Unterstützung			.284	NA	.399
Optimismus				.284	.887
Selbstwert					.242
Lebenszufriedenheit					

**Anmerkung:** Überhalb der Diagonale sind die p-Werte von Egger-Test für die jeweiligen Beziehungen zu sehen. Die grau hinterlegten Wert sollten inhaltlich nicht interpretiert werden (siehe 5.5.1).

Insgesamt scheinen die Ergebnisse der Meta-Analyse, zumindest bei bivariater Betrachtung, keinen ausgeprägten Publikationsbias aufzuweisen. Keiner der Tests zeigt einen signifikanten Zusammenhang zwischen Effektstärke und Standardfehler. Lediglich die Beziehungen zwischen den Mediatoren sind mit Vorsicht zu interpretieren, da sie auf sehr wenigen Studien beruhen. Allerdings überrascht dies kaum, da etwa die Kovarianz  $k_2$  bereits bei der Analyse problematisch ist. Bei Zelle 4, der Beziehung zwischen Religiosität und Lebenszufriedenheit, zeigen sich gewisse Indizien für das Vorliegen eines Publikationsbias. Dies ist zum einen ein fast signifikanter (p-Wert = .064) Egger-Test sowie eine erkennbar asymmetrische Verteilung des Funnel-Plots. Da der Egger-Test jedoch nicht signifikant ist, wird davon ausgegangen, dass eine gewisse Verzerrung vorliegt, diese aber nicht ausgeprägt genug ist, um einen Publikationsbias

nachzuweisen. Demzufolge sind die in dieser Arbeit gewonnen Erkenntnisse trotz leichter Verzerrungen durchaus valide. Darüber hinaus macht die starke Variation der Effektstärken die Notwendigkeit von Random-Effects-Schätzungen deutlich.



## **10 Schlussfolgerung**

Im abschließenden Kapitel werden nun die Ergebnisse der durchgeführten modellbasierten Meta-Analyse systematisch aufbereitet. Daran anknüpfend folgt eine Sammlung der Schwächen der Arbeit, bevor Ansatzpunkte für weitere Forschungsarbeiten formuliert werden.

### **10.1 Systematisierung der Ergebnisse**

Es liegt derzeit keine aktuelle meta-analytische Arbeit über den zugrunde gelegten Zusammenhang vor (siehe die bekannten Arbeiten in Kapitel 3 und 4). Die inzwischen etwas in die Jahre gekommenen Erkenntnisse vorheriger Meta-Analysen konnten in dieser Arbeit somit aktualisiert werden. Die Durchführung einer modellbasierten Meta-Analyse, unter Berücksichtigung konkreter Erklärungsmechanismen, die für den positiven Zusammenhang von Religiosität und Lebenszufriedenheit verantwortlich sind, ist bislang einzigartig. Ein Vergleich mit anderen (modellbasierten) Studienergebnissen ist daher nicht möglich. Die Anwendung des OSMA-SEM-Ansatzes zur Sensitivitätsüberprüfung und vor allem zur Untersuchung kontinuierlicher Moderatoren ist eine der ersten Anwendungen dieses Verfahrens nach Jak/Cheung (2018a) überhaupt.

Die Arbeit hat drei unterschiedliche Fragestellungen an einer empirischen Exemplifikation untersucht. In der vorliegenden Studie ging es zunächst einmal darum, zu evaluieren, ob es möglich ist, eine explorative modellbasierte Meta-Analyse ohne etabliertes Erklärungsmodell durchzuführen. Darüber hinaus beschäftigt sich die Arbeit mit den inhaltlichen Fragen bezüglich des empirischen Zusammenhangs der beiden Variablen Religiosität und Lebenszufriedenheit und welche Mechanismen zur Erklärung herangezogen werden können. Weiterhin werden die ungelösten methodischen Fragen des modellbasierten TSSEM-Ansatzes in den Fokus genommen und verschiedene Verfahren exemplarisch angewandt.

Die bislang existierenden modellbasierten Meta-Analysen basieren auf etablierten theoretisch fundierten Erklärungsmodellen wie der „Theory of Planned Behavior“. Die vorliegende Studie zeigt, dass es möglich ist, ein Erklärungsmodell für eine modellbasierte Meta-Analyse analytisch-empirisch herzuleiten und dieses meta-analytisch zu überprüfen. Eine explorative Anwendung des Verfahrens ist somit durchführbar, auch wenn dies mit bestimmten Problemen, wie etwa Konvergenz- und Schätzproblemen bei nur wenig untersuchten Teilbereichen des Untersuchungsmodells, verbunden ist. Dennoch ergibt sich dadurch die Möglichkeit, die modellbasierte Meta-Analyse in neuen Anwendungsbereichen einzusetzen. Dadurch rückt diese Form der Meta-Analyse näher an die empirische Praxis in den Sozialwissenschaften heran, wo häufig

kein feststehendes Erklärungsmodell in Form einer etablierten Theorie vorliegt und zur Erklärung spezifischer sozialer Phänomene zunächst ein empirisch-analytisches Modell konstruiert werden muss.

Inhaltlich haben sich, neben der bereits genannten erstmaligen Durchführung einer modellbasierten Meta-Analyse zu dieser Forschungsfrage, eine Reihe von interessanten Ergebnissen gezeigt. In Tabelle 62 sind diese modellinhärenten Ergebnisse der Analyse in Bezug auf die in Kapitel 8 formulierten Erwartungen zusammenfassend zu sehen. Die konkreten Anmerkungen zu den jeweiligen Erwartungen folgen in den nachfolgenden Ausführungen.

**Tabelle 62: Erwartungen und Ergebnis**

	Beschreibung	Falsifikation
E1	Untersuchungsmodell konvergiert	0
E2	Random-Effects-Modell	0
E3.1	b34: positiver Effekt (Religiosität → Lebenszufriedenheit)	0
E3.2	b31: positiver Effekt (Religiosität → soziale Unterstützung)	0
E3.3	b32: positiver Effekt (Religiosität → Optimismus)	0
E3.4	b33: positiver Effekt (Religiosität → Selbstwert)	0
E3.5	b35: positiver Effekt (soziale Unterstützung → Lebenszufriedenheit)	0
E3.6	b36: positiver Effekt (Optimismus → Lebenszufriedenheit)	0
E3.7	b37: positiver Effekt (Selbstwert → Lebenszufriedenheit)	0
E4.1	ind1: partieller positiver Mediator (soziale Unterstützung)	0
E4.2	ind2: partieller positiver Mediator (Optimismus)	0
E4.3	ind3: partieller positiver Mediator (Selbstwert)	0
E5.1	k1: positive Kovariation (soziale Unterstützung und Optimismus)	0
E5.2	k2: positive Kovariation (soziale Unterstützung und Selbstwert)	X
E5.3	k3: positive Kovariation (Optimismus und Selbstwert)	0
E6.1	Negativer Moderator HDI → b34; weitere Interaktionsterme möglich	0
E6.2	Positiver Moderator WR → b34; weitere Interaktionsterme möglich	(0)
E6.3	Kleinere Effektstärken bei Studien mit großer Fallzahl	X
E6.4	Kein Publikationsbias (herkömmliche Literatur vs. graue Literatur)	X
E6.5	Kleinere Effektstärken bei repräsentativen Stichproben	X
E6.6	Kleinere Effektstärken bei schlechter bewerteten Studien	X
E6.7	Kein Einfluss des Publikationsjahrs	X
E7.1	Konvergenzprobleme beim Subgruppendesign	0
E7.2	Präzisere Schätzung (schmalere LBCIs) beim Wilson- und robusten Ansatz	0

**Anmerkung:** 0 = konnte nicht falsifiziert werden; (0) = teilweise falsifiziert; X = konnte falsifiziert werden.

Das Untersuchungsmodell ist in allen regulären Modellierungsschritten konvergiert (E1). Lediglich beim Subgruppendesign im vollständigen Untersuchungsmodell konnte, aufgrund fehlender Werte, keine gültige Lösung gefunden werden. Nach dem unerwarteten Ausschluss der Kovarianz k2 und der Gewinnung eines Freiheitsgrades haben die Fit-Werte jeweils eine sehr gute Passung des Modells an die empirischen Daten gezeigt. In allen Analyseschritten haben die entsprechenden statischen Kennwerte die Nutzung eines Random-Effects-Modells nahegelegt (E2). Selbst bei den, anhand des Medians, getrennten Gruppen im Subgruppendesign war

in den Gruppen die Heterogenität jeweils höher als die notwendigen Grenzwerte für die Schätzung eines Fixed-Effects-Modells.

Der bereits vielfach empirisch und auch meta-analytisch überprüfte schwach positive Zusammenhang von Religiosität und Lebenszufriedenheit (siehe Kapitel 3.3) hat sich in dieser Studie erneut gezeigt (E3.1). Die Analyse des Publikationsbias dieser Beziehung zeigt jedoch schwache Hinweise für eine gewisse Überschätzung des Effekts  $b_{34}$ . Auch die positiven Wirkungen von Religiosität auf die drei Mediatorvariablen (E3.2, E3.3 und E3.4) und die positive Wirkung der Mediatoren auf Lebenszufriedenheit (E3.5, E3.6 und E3.7) konnten nicht falsifiziert werden. Die Stärke der Effekte ist auch über die verschiedenen Ansätzen hinweg kaum unterschiedlich.

Auch die oftmals vermuteten und auch teilweise empirisch präsentierten Erklärungsmechanismen haben einer meta-analytischen Überprüfung standgehalten. Durchweg zeigen sich sowohl signifikante direkte als auch signifikante indirekte Effekte für alle drei Mediatorvariablen (E4.1, E4.2 und E4.3). Dies entspricht den vermuteten partiellen Mediationen. Vergleicht man das verringerte Untersuchungsmodell mit nur einem Mediator (soziale Unterstützung) mit dem vollständigen Modell, zeigt sich, dass im verringerten Modell der direkte Effekt ( $b_{34}$ ) stärker ausgeprägt ist. Es ist deshalb davon auszugehen, dass sich bei Berücksichtigung weiterer Erklärungsmechanismen (siehe Kapitel 3.4) der direkte Effekt weiter verringern würde und vermutlich irgendwann nicht mehr signifikant werden würde. Deshalb liegt die Vermutung nahe, dass eigentlich kein direkter, oder zumindest ein nur sehr schwacher, Effekt zwischen den beiden zentralen Variablen des Untersuchungsmodells vorliegt und der zumeist positive Zusammenhang aus der Beziehung mit anderen Variablen resultiert. Ebenfalls zeigt sich, dass die Effekte von den Mediatorvariablen auf die abhängige Variable Lebenszufriedenheit deutlich stärker ausgeprägt sind als der direkte Effekt von Religiosität. Zur umfassenden Erklärung des Phänomens Lebenszufriedenheit kann die Variable Religiosität deshalb vergleichsweise wenig beitragen.

Bei den drei postulierten Kovarianzen zwischen den Mediatoren sieht es anders aus (E5.1, E5.2 und E5.3). Die Kovariation  $k_2$  ist nicht signifikant (E5.2) und wird deshalb in weiteren Modellschätzungen ausgeschlossen, was den bereits erwähnten positiven Nebeneffekt hat, dass dann, aufgrund des gewonnenen Freiheitsgrades, Fit-Indizes zur Verfügung stehen. Interessanterweise ist die Kovarianz jedoch im Wilson- und im robusten Ansatz jeweils signifikant. Dies zeigt bereits die höhere Robustheit dieser Methoden gegenüber einer geringen Anzahl von Effektstärken für spezifische Beziehungen (E7.2).

Auch die Überprüfung der verschiedenen postulierten Moderatorvariablen, ein im Kontext von modellbasierten Meta-Analysen bislang kaum angewandter Aspekt, zeigen interessante Ergebnisse. Die Wirksamkeit der beiden vorgeschlagenen Moderatoren zur Überprüfung der kulturellen Bedingtheit des Erklärungsmodells auf der Makroebene (HDI und Wichtigkeit von Religion) konnte nicht vollständig falsifiziert werden (E6.1 und E6.2). Während die einzige ursprünglich für den TSSEM-Ansatz vorliegende Methode zur Untersuchung von Moderatorvariablen (Subgruppenanalyse) nur grobe Erkenntnisse generiert, zeigt der OSMASEM detailliert, welche Teile des Erklärungsmodells durch die Moderatorvariablen beeinflusst werden. Die Subgruppenanalyse hat im vorliegenden Untersuchungsmodell mit großen Konvergenzproblemen aufgrund vieler fehlender Werte zu kämpfen (E7.1).

Eine Durchführung im vollständigen Untersuchungsmodell ist somit nicht möglich und die beiden unterschiedlichen Methoden zur Durchführung von Moderatoranalysen können nur im stark reduzierten Untersuchungsmodell miteinander verglichen werden. Die Subgruppenanalyse, auf Basis eines Mediansplits, zeigt für beide Kulturmoderatorvariablen im stark verringerten Erklärungsmodell einen signifikanten moderierenden Effekt auf die A-Matrix. Allerdings lässt sich aus dieser Untersuchung nicht präzise ableiten, auf welche Koeffizienten dieser konkret wirksam ist. Mit dem neuen OSMASEM-Ansatz ist es im vollständigen Modell andererseits möglich, eine Untersuchung der kontinuierlichen Moderatorvariablen durchzuführen.

Das Ergebnis dieser Analyse zeigt eindeutig, auf welche Parameter des Erklärungsmodells (b-Koeffizienten) der entsprechende Moderator wirkt. Im Fall des HDI-Moderators zeigt sich etwa ein signifikanter, moderierender, negativer Effekt auf den Haupteffekt (b34). Dieser besagt, dass in höher entwickelten Ländern (höherer HDI) der Haupteffekt von Religiosität auf die Lebenszufriedenheit schwächer ausgeprägt ist (E6.1). Im verringerten Subgruppenmodell kann dieser Effekt jedoch nicht repliziert werden, da die Teststärke nicht ausreichend ist. Im OSMA-SEM-Ansatz zeigt sich dieser Effekt jedoch auch im verringerten Modell.

Die Überprüfung des zweiten Makromoderators (E6.2), Wichtigkeit von Religion, zeigt hingegen signifikante Moderatoreffekte auf die Regressionskoeffizienten b32, b33 und b37. Allerdings wird nicht wie vermutet der Haupteffekt b34 durch die Variable moderiert.

Auch eine Reihe methodischer Moderatoren wurde an dieser Stelle überprüft (E6.3 – E6.7). Als einzige der dort überprüften Variablen ergibt das Publikationsjahr einen signifikanten Effekt auf Teilbereiche der Erklärungsmodells (E6.7). Dies entspricht somit Periodeneffekte für die entsprechenden Teile des Untersuchungsmodells.

Die empirische Überprüfung der unterschiedlichen Verfahren zur Berücksichtigung abhängiger Effektstärken hat ergeben, dass, wenn die Generalisierung von Ergebnissen im Fokus steht, kaum Unterschiede zwischen den Methoden bestehen. Was die Punktschätzungen der Verfahren betrifft, zeigt sich zwischen Wilson-Ansatz und robustem Ansatz kaum ein Unterschied. Lediglich im Vergleich zum Mittelwertansatz lassen sich beachtenswerte Unterschiede erkennen (E7.2). Bei einer ausreichend hohen Anzahl von Effektstärken für alle Beziehungen kann durchaus die pragmatische Lösung via Mittelwertbildung gewählt werden. Diese hat jedoch den Nachteil, dass sie zu eher breiten LBCIs führt. Liegen lediglich sehr wenige Effektstärken für bestimmte Bereiche vor, sollte deshalb auf den Wilson-Ansatz bzw. den robusten Ansatz zurückgegriffen werden, um die Präzision der Schätzung zu erhöhen. Auch wenn einzelne schwach besetzte Beziehungen, wie etwa im vorliegenden Fall die Kovarianz  $k_2$ , nicht signifikant werden, sollte eines der anderen Verfahren angewandt werden, um die Robustheit der Schätzung zu erhöhen. Sinnvoll ist sicherlich die Anwendung mehrerer Verfahren, um die Robustheit der Ergebnisse zu überprüfen.

Die Überprüfung des Publikationsbias ergibt keine besonderen Hinweise auf schwerwiegende Verzerrungen aufgrund von nicht berücksichtigten empirischen Studien, so dass den Analyseergebnissen, auch im Hinblick auf die Sensitivitätsanalysen durch die Variation der verschiedenen Analysespielarten, uneingeschränkt Vertrauen entgegengebracht werden kann.

## **10.2 Grenzen der vorliegenden modellbasierten Meta-Analyse**

Gegen jede empirische Arbeit können eine Reihe von Einwände geltend gemacht werden. Dies gilt auch für die vorliegende modellbasierte Meta-Analyse, die in Anlehnung an die methodologische empirische Praxis in dem Forschungszweig und den Empfehlungen der Methodensliteratur entstanden ist. Diese Schwächen lassen sich zunächst einmal auf die grundlegenden gegen die Methodik angeführten Argumente zurückführen (siehe Kapitel 2.3). Diese sollen an dieser Stelle jedoch nicht weiter thematisiert werden, da sie bereits diskutiert wurden. Darüber hinaus lässt sich jedoch auch die konkrete Ausgestaltung der Analyse kritisieren. Diese Kritik bezieht sich von den zugrundeliegenden Definitionen der Kernkonstrukte bis hin zur Darstellung der Ergebnisse über alle Phasen der Meta-Analyse hinweg. Die meisten dieser selbstverursachten Schwächen resultieren dabei aus einer Reihe kontingenter methodischer Entscheidungen und sind nicht zuletzt in Hinblick auf beschränkte zeitliche und materielle Ressourcen entstanden. Die aus Sicht der vorliegenden Arbeit wichtigsten Punkte werden nun in der chronologischen Abfolge einer Meta-Analyse thematisiert.

Die Definition der Kernkonstrukte ist im Hinblick auf die unabhängige Variable Religiosität bzw. Spiritualität kritisch zu betrachten. Durch die Berücksichtigung des Spiritualitätskonstrukts wird das Religiositätskonzept aufgeweicht. Allerdings orientiert sich dieses Vorgehen an der aktuellen Praxis in diesem Forschungsbereich, die von einer großen Übereinstimmung der beiden Konstrukte ausgeht (siehe Kapitel 3.1). Darüber hinaus können durch dieses Vorgehen weitere Studien für die Meta-Analyse berücksichtigt werden, was die Teststärke, besonders im Blick auf die Mediatormechanismen, erhöht. Auch die Beschränkung der abhängigen Variable auf die Lebenszufriedenheit lässt sich kritisieren, da eine Überprüfung der affektiven kurzfristigen state-Komponente andere Ergebnisse und Schlüsse erwarten lässt.

Auch das Erklärungsmodell und das daraus abgeleitete Untersuchungsmodell kann im Hinblick auf die Vollständigkeit beanstandet werden. So fehlen etwa negative Wirkpfade sowie reziproke Beziehungen, die aus dem Modell ausgeschlossen wurden. Andererseits sollte das Erklärungsmodell nicht überkomplex werden, da ansonsten eine sinnvolle empirische Anwendung meta-analytisch nicht möglich erscheint. Es geht also vor allem darum, eine Balance zwischen Komplexität und Erklärungskraft des meta-analytischen Modells zu finden.

Die Literatursuche zum Auffinden relevanter Primärliteratur lässt sich aus verschiedenen Blickwinkeln angreifen. Zum einen wurde nur Literatur berücksichtigt, die deutsch- oder englischsprachig ist. Es ist durchaus plausibel, dass durch diese Beschränkungen bestimmte Ergebnisse in der Meta-Analyse systematisch nicht berücksichtigt werden konnten. Darüber hinaus wurden keine weiteren Suchstrategien neben der systematischen Suche in Literaturdatenbanken und den Literaturverzeichnissen bekannter Reviews und Meta-Analysen angewandt. Weitere Suchmethoden wie das Footnote-Chasing könnten ebenfalls eingesetzt werden. Auch das Spezifikum des multivariaten Erklärungsmodells wurde in der Literatursuche nicht angemessen berücksichtigt, da lediglich Studien gesucht wurden, die über die Keywords von abhängiger und unabhängiger Variable gefunden wurden. Dies führt zu dem Problem, dass viele fehlende Werte vorliegen. Im häufig relevanteren Random-Effects-Modell mag dies zunächst noch unproblematisch erscheinen, da das Modell gut damit umgehen kann. Muss allerdings ein Fixed-Effects-Modell geschätzt werden, dann sind nach Jak/Cheung (2017) komplexere Modellierungsschritte notwendig und deshalb eine aufwendigere Literatursuche angemessen.

Auch das anschließende Screening und die Kodierung können kritisiert werden, da im Idealfall in beiden Arbeitsschritten zur Reliabilitätssicherung eine vollständige Kodierung durch zwei Autoren angeraten wird. Darüber hinaus wurden im Falle von fehlenden statistischen Informationen die Studienautoren aufgrund beschränkter personeller Ressourcen nicht kontaktiert, so

dass manche Informationen, wie etwa Korrelationen, fehlen und dadurch manche Studien ausgeschlossen werden mussten. Auch eine formelle Überprüfung der Reliabilität mit den entsprechenden statistischen Koeffizienten hätte durchgeführt werden können.

Daneben kann die Auswahl der konkreten Analysemethoden beanstandet werden. Der Ansatz von Yu et al. (2018) zur verbesserten Quantifizierung der Heterogenität zwischen den Studien hätte, zumindest experimentell, angewandt werden können. Ebenso wäre ein Vergleich der Ergebnisse weiterer Methoden zum Umgang mit fehlenden Effektstärken, wie etwa eine Zufallsauswahl, aber auch die Reliabilitätskorrektur im Sinne von Schmidt/Hunter (2015) möglich gewesen. Alle Modellierungsschritte hätten, soweit möglich, auch mittels der verschiedenen Variationen des GLS-Ansatzes nach Becker (2009) oder des Bayes-Ansatzes weiter auf Sensitivität hin überprüft werden können. Die Anwendung des OSMASEM-Ansatzes zur Untersuchung der Moderatoren ist sicherlich riskant, da noch keine Anwendungen des Ansatzes außer den Beispielen von Jak/Cheung (2018a) vorliegen. Inwiefern die Ergebnisse korrekt und robust sind, muss die zukünftige Anwendung und der Vergleich mit ursprünglichen bislang vorliegenden modellbasierten Methoden zeigen. Dieser Ansatz kann jedoch dem Hauptnachteil des TSSEM-Ansatzes bezüglich der Untersuchung von Moderatoren entgegenwirken. Dies ermöglicht in der vorliegenden Untersuchung auch den Vergleich beider Methoden zur Moderatorenuntersuchung (OSMASEM vs. Subgruppen in TSSEM). Beim Subgruppendesign bleibt die Frage, welche Kategorien zur Untersuchung verwendet werden sollen. So ist stets zu klären, welche Kategorien inhaltlich gehaltvoll und empirisch umsetzbar sind. In der vorliegenden Arbeit wurde dazu lediglich ein analytisch nicht rechtfertigbarer Mediansplit durchgeführt.

Auch was die Anzahl der Moderatoren im Vergleich zu herkömmlichen Mixed-Effects Meta-Regressionen betrifft, bleibt die vorliegende Arbeit auf jeweils eine Moderatorvariable je Modell beschränkt. Die Leistungsfähigkeit des OSMASEM zur Evaluierung multipler Moderatoren muss sich in Zukunft noch zeigen. Bislang ist die Anwendung des OSMASEM-Ansatzes nur bei vollständigen Daten der Moderatorvariable und bezüglich aller Elemente einer Matrix möglich. Auch ein Test von einzelnen Pfaden des Untersuchungsmodells wäre sicherlich interessant, kann aber derzeit noch nicht durchgeführt werden, da unklar ist, ob die Modelle dann korrekt geschätzt werden. Auch die Überprüfung der unterschiedlichen Verfahren zur Berücksichtigung abhängiger Effektstärken steht im OSMASEM noch aus, da die Anwendung des Wilson-Ansatzes und des robusten Ansatzes bislang im metaSEM nicht vollständig realisiert werden können. Ebenfalls muss die Überprüfung des Publikationsbias kritisch gesehen werden, da lediglich die bivariaten Zusammenhänge, aufgrund des Fehlens von Verfahren für die modellbasierte Meta-Analyse, überprüft wurden.

Aber auch bezüglich der Durchführung und was den Vergleich der verschiedenen Ansätze betrifft, muss bedacht werden, dass es sich lediglich um eine empirische Anwendung handelt. Die gewonnenen Erkenntnisse sind nicht verallgemeinerbar. Sie dienen lediglich als Hinweise für weitere empirische Studien und können die Ergebnisse von Monte-Carlo Simulationsstudien keineswegs ersetzen. Die Erkenntnisse können jedoch Anhaltspunkte weiterer Untersuchungen darstellen.

### **10.3 Ausblick**

Aus den Erfahrungen, die durch die Exemplifikation dieser modellbasierten Meta-Analyse gewonnen wurden, können verschiedene Schlussfolgerungen für weitere Forschungsansätze abgeleitet werden. Zunächst einmal folgen dazu einige auf den Inhalt bezogene Perspektiven, bevor die methodischen Anknüpfungspunkte für weitere Forschungsansätze beleuchtet werden.

Auch wenn die inhaltlichen Ergebnisse in dieser Exemplifizierungsarbeit etwas im Hintergrund stehen, werden hier einige Anknüpfungspunkte präsentiert. Bei weiteren sozialwissenschaftlichen Untersuchungen muss zunächst eine spezifischere Einbettung in den Theoriekanon erfolgen (siehe dazu die Anmerkungen in Kapitel 1). Was die grundsätzliche Fragestellung betrifft, sollten zukünftige Untersuchungen von der Frage abrücken, ob es einen Zusammenhang zwischen den Variablen gibt oder nicht, und vielmehr die spezifischen Wirkursachen für den Zusammenhang sowie die Effektrichtungen in den Fokus nehmen. Weitere Forschungsarbeiten können sich an dem in dieser Arbeit entwickelten Erklärungsmodell orientieren und den Zusammenhang zwischen Religiosität bzw. Spiritualität und Lebenszufriedenheit nicht direkt, sondern mithilfe der in dieser Arbeit systematisierten Erklärungsmechanismen in Form von Mediatoren untersuchen und auch bestimmte Moderatoren ins Untersuchungsmodell aufnehmen. Dabei sind sicherlich die Mediatoren aus den Bereichen der kognitiven Orientierung und positiven Gefühle besonders für die weitere Untersuchung interessant, weil hier vor allem positive Auswirkungen auf die Lebenszufriedenheit zu erwarten sind. Auch negative Wirkpfade und reziproke Beziehungen könnten berücksichtigt werden.

Möglicherweise sollten dabei auch weitere sozialwissenschaftliche Mechanismen ins Modell aufgenommen werden, um die humanwissenschaftlich dominierte Sichtweise des Erklärungsmodells zu ergänzen. So wäre etwa eine differenzierte Untersuchung des Mediators „soziale Unterstützung“ durch den Vergleich der beiden Teilkomponenten emotionale und materielle Unterstützung aufschlussreich. Weiterhin wäre zu überlegen, ob die Mediatoren aus dem Bereich Lebensstil (bspw. gesunde Ernährung) bei gleichzeitiger Kontrolle von Gesundheit weitere Erklärungsleistung liefern. So sollten auch weitere, in der empirischen Untersuchung dieser

Arbeit nicht berücksichtigte Mediatoren, möglichst vergleichend, in zukünftigen Untersuchungen des Zusammenhangs berücksichtigt werden. Dies alles gilt sowohl für neue Primär- bzw. Sekundärforschung als auch für weitere meta-analytische Arbeiten. In der bisherigen Forschungspraxis ist dies bislang kaum der Fall (siehe Kapitel 3).

Die Klassifizierung der Messinstrumente könnte als Grundlage für weitere Moderationsanalysen dienen. Allerdings muss diese Klassifizierung zunächst im Hinblick auf Trennschärfe weiterentwickelt werden. Auch die kulturelle Bedingtheit des Zusammenhangs sollte zukünftig verstärkt in den Fokus genommen werden. Die Ergebnisse dieser Arbeit können erste Einsichten liefern, welche Moderatoren auf Makroebene sinnvollerweise untersucht werden sollen. Es ist dabei sicher notwendig, alternative Messungen für Religiosität auf Makroebene zu finden. Auch eine Kombination aus beiden kulturellen Moderatorvariablen wäre interessant. Mit einem solchen Modell könnte überprüft werden, ob sich das Erklärungsmodell in hoch entwickelten (hoher HDI) und dennoch stark religiösen Gesellschaften, wie beispielsweise den USA, von hoch entwickelten stärker säkularen Staaten, wie den skandinavischen Ländern, unterscheidet. Auch eine Überprüfung der S-Matrix, also der Varianzen bzw. Fehler und Kovarianzen, wäre möglich. Darüber hinaus könnten etwa Unterschiede zwischen eher kollektivistischen und individualistischen und unterschiedlich religiös geprägten Gesellschaften in den Blick genommen werden. Auch eine Überprüfung weiterer Moderatoren wie der differenzierten Untersuchung von Stichprobenalter und Art der Stichprobe könnte hilfreiche Erkenntnisse generieren. Ob auch einzelne Elemente des Modells (bspw. ein einzelner Regressionskoeffizient) mittels OS-MASEM getestet werden können, müssen zukünftige methodische Studien zeigen.

Bei zukünftigen Anwendungen der Methodik im explorativen Kontext muss verstärkt darauf geachtet werden, dass alle Beziehungen in ausreichendem Maße vorliegen, um die entsprechenden Moderatoranalysen und auch die Kombination verschiedener Moderatoren durchführen zu können. Es sollte eine Literatursuchstrategie spezifisch für den MASEM-Bereich entwickelt werden, da die herkömmliche Literatursuche für die (explorative) modellbasierte Meta-Analyse eher ungeeignet erscheint, da Mediatorvariablen lediglich als Zufallstreffer berücksichtigt werden. Am zielführendsten erscheint es dabei, einzelne Literatursuchen für die unterschiedlichen Pfade des Modells durchzuführen. Die Gefahr besteht jedoch darin, dass sich die gefundenen Studien noch stärker voneinander unterscheiden als im vorliegenden Ansatz, bei dem nur nach dem Haupteffekt gesucht wurde. Möglicherweise ist auch eine Kombination beider Strategien zielführend. Dies war allerdings in der exemplarischen Anwendung des Verfahrens im vorliegenden Fall nicht möglich, da das abschließende Untersuchungsmodell erst nach der Literatursuche anhand der am häufigsten untersuchten Mediatoren festgelegt wurde.

Die vorliegenden empirischen Erkenntnisse basieren lediglich auf korrelativen Daten im Querschnitt. Diese Datenbasis ermöglicht keine kausalen Schlüsse im formellen Sinn. Auch umgekehrte Kausalrichtungen sind möglich (siehe die Ausführungen in Kapitel 3.4). So sind etwa Wechselwirkungen zwischen unabhängiger und abhängiger Variable und den Mediatoren denkbar. Nur eine sorgfältige Prüfung im Längsschnittdesign und mittels (quasi-) experimenteller Arbeiten kann Klarheit darüber schaffen. Der Zusammenhang sollte zukünftig längsschnittlich in unterschiedlichen Stichproben (Kranke, Gesunde, Studierende etc.) unter Berücksichtigung der Erklärungsmechanismen untersucht werden.

Weitere methodische Arbeiten sollten zunächst auch die Vorarbeiten einer modellbasierten Meta-Analyse verstärkt in den Fokus nehmen und sich mit Fragen der adäquaten explorativen Entwicklung eines multivariaten Erklärungsmodells (wenn kein etabliertes vorliegt) und einer gegebenenfalls angepassten Erhebungsstrategie beschäftigen. Dabei ist es sicherlich sinnvoll, auch die existierenden R-Pakete bzw. weitere Softwarelösungen für die meta-analytischen Vorarbeiten (Erhebung und Kodierung) einzusetzen.

Eine Vielzahl der hier aufgeworfenen methodischen Fragen sollte zukünftig mit Simulationsstudien (siehe dazu ausführlich Kapitel 5.4.2) und weiteren empirischen Anwendungen überprüft werden. Die Simulationsparameter sollten in neuen Studien nicht willkürlich festgelegt werden, sondern auf einem fundierten systematischen Review, wie etwa bei Fernández et al. (2019), der modellbasierten meta-analytischen Praxis und den bisherigen Simulationsstudien basieren. In den bisherigen Simulationsarbeiten wurden häufig unrealistische Simulationsparameter, wie etwa die ausschließliche Verwendung von Fixed-Effects-Modellen, angewandt, die in der empirischen Praxis nur wenig Relevanz besitzen. Die Aussagen solcher Studien bleiben dadurch begrenzt. Zu den zukünftig zu berücksichtigenden Parametern zählen etwa die Variation des Anteils an fehlenden Werten, die Art der Mess- und Strukturmodelle, die Verwendung von Reliabilitätskorrekturen sowie der generelle Vergleich der unterschiedlichen modellbasierten Verfahren (TSSEM, GLS, OSMASEM etc.) in Hinblick auf alle genannten Aspekte.

Besonders die hier exemplarisch angewandten Verfahren zur Berücksichtigung abhängiger Effektstärken müssen mittels Simulationen auf ihre Leistungsfähigkeit hin überprüft werden, da sie in der Praxis oft vernachlässigt werden (siehe Kapitel 4.1). Alle grundlegenden Modellierungsformen sollten insbesondere im Hinblick auf die Moderatoranalyse untersucht werden. Dies gilt insbesondere für den OSMASEM-Ansatz, da für dieses neue Verfahren neben den Erkenntnissen von Jak/Cheung (2018a) und der vorliegenden empirischen Anwendung noch keine Simulationsergebnisse vorliegen. Sollte der OSMASEM-Ansatz seine Versprechungen

in weiteren Anwendungen und Simulationsstudien erfüllen können, dann wird sich dieser Ansatz sicher aufgrund seiner Vorteile durchsetzen. Wünschenswert wäre darüber hinaus eine Erweiterung bzw. Verknüpfung der bestehenden R-Pakete *metafor* und *metaSEM*, um in einem einheitlichen Framework eine vollständige modellbasierte Meta-Analyse mit verschiedenen Möglichkeiten zur Berücksichtigung abhängiger Effektstärken, Ergebnisdiagnostik sowie Visualisierungen durchführen zu können.

Ebenfalls relevant sind bei explorativen Erklärungsmodellen, wie im vorliegenden Fall, sicherlich auch die Ausarbeitung von Mindestanzahlen für Teile des Untersuchungsmodells. Im vorliegenden Fall hat sich gezeigt, dass es bei nur sehr wenigen vorliegenden Effektstärken für spezifische Beziehungen ( $<10$ ) zu Konvergenz- und Schätzproblemen (bspw. sehr breiten LBCIs) kommen kann. Interessant wären hier sicherlich Anwendungsregeln, ab welcher Anzahl von Effektstärken eine Beziehung sinnvollerweise im Untersuchungsmodell berücksichtigt werden sollte. Wie bereits angesprochen, könnte es sinnvoll sein, getrennte Literatursuchen für unterschiedliche Bereiche des Untersuchungsmodells durchzuführen. Bei etablierten Erklärungsmodellen könnte dieses Problem durch eine Mindestanzahl von Beziehungen als Selektionskriterium sichergestellt werden, die in einer Untersuchung vorliegen müssen (bspw. 50 %), um die entsprechende Studie auch in die Meta-Analyse aufzunehmen. Dies hätte allerdings den Nachteil, dass nicht alle empirischen Informationen bei der meta-analytischen Zusammenfassung genutzt werden können.

Insgesamt bedarf es einer Anpassung und Erweiterung der Leitlinien zur Anfertigung von Meta-Analysen um die modellbasierte Perspektive, da die bisherigen vorwiegend humanwissenschaftlichen Ansätze zumeist im Hinblick auf bivariate Beziehungen bzw. Gruppenunterschiede entwickelt wurden und multivariate Erklärungsmodelle kaum berücksichtigen. Dies hat zur Folge, dass bei der Anfertigung einer modellbasierten meta-analytischen Studie eine Vielzahl von methodischen ad-hoc Entscheidungen getroffen werden müssen. Die bisherigen methodischen Arbeiten fokussieren sich auf die statistische und programmtechnische Umsetzung dieser Analyseverfahren und vernachlässigen die Vorarbeiten. Diese Arbeit konnte an einigen Stellen erste Hinweise für die konsequente Weiterentwicklung der bisherigen methodischen Standards liefern, die nun konsequent aufgegriffen und weiterverfolgt werden müssen, um sie dem methodischen Kanon der Meta-Analyse anzufügen.

## Anhang A: Kodierschema

Das folgende Kodierschema wurde zur Extrahierung der benötigten Informationen genutzt. Lediglich die Teile zur Erfassung von mehrfachen Messinstrumenten werden aus Platzgründen nicht dargestellt.

Meta-Daten	
<b>Fortlaufende ID (globale ID – 3 Ziffern)</b>	<var>
	V0
<b>Studiennummer</b>	
	V1
<b>Autoren</b>	
	V2
<b>Titel</b>	
	V3
<b>DOI</b>	
	V4
<b>Verkoderkürzel</b>	
<b>MW</b>	V5
<b>Kodier-Datum</b>	
	V6
<b>Versionsnummer des Koding-Schemas</b>	
	V7

Angaben zur Studie	
<b>H2.1 Stichprobengröße</b>	
<input type="checkbox"/> Konkrete Angabe: <input type="checkbox"/> Keine Angabe (-99)	V8
<b>H2.2 Publikationsjahr (siehe Bibliographie)</b>	
	V9
<b>H2.3 Land</b>	
<b>Freitext:</b>	V10
<input type="checkbox"/> mehrere Länder – deshalb Exklusion - (88)	
<b>H2.4 Publikationsstatus (siehe Bibliographie)</b>	
<b>herkömmliche Literatur</b>	V11
<input type="checkbox"/> Journal (1)	
<input type="checkbox"/> Monographie (2)	
<input type="checkbox"/> Beitrag im Sammelband (3)	
<b>graue Literatur</b>	V12
<input type="checkbox"/> Dissertation (4)	
<input type="checkbox"/> Stud. Abschlussarbeit (BA/MA) (5)	
<input type="checkbox"/> Forschungsbericht (6)	
<input type="checkbox"/> Schriftenreihe (7)	
<input type="checkbox"/> Konferenzbeitrag (8)	
<input type="checkbox"/> etwas anderes (9) und zwar:	
<b>H2.5 Welches Journal?</b>	
	V13

<b>H2.6 Datenquelle</b>	
Name des Datensatzes:	V14
<input type="checkbox"/> eigene Erhebung (2)	V15
<input type="checkbox"/> Nicht angegeben (99)	
Daten aus dem Jahr:	
<input type="checkbox"/> Nicht angegeben (99)	
<b>H2.7 Wurden Korrelationen zwischen den Untersuchungsvariablen und/oder potenziellen Mediatoren berichtet?</b>	
<input type="checkbox"/> Ja (1)	V16
<input type="checkbox"/> Nein (2)	
<input type="checkbox"/> Nur Beziehung zwischen einer Untersuchungsvariable und einem Mediator und nicht zwischen UV und AV (3)	
<b>H2.8 Wurde eine Messung von R/S für die Untersuchung genutzt?</b>	
<input type="checkbox"/> Ja (1)	V17
<input type="checkbox"/> Nein (2) – weiter zu H2.10	
<b>H2.10 Einordnung Messinstrument R/S</b>	
Wieviele Messungen von R/S wurden vorgenommen? Anzahl:	V18
<hr/>	
Welcher Art ist R/S Messung 1?	V19
<input type="checkbox"/> rein öffentliche religiöse/spirituelle Praxis	V20 V21 V22
<input type="checkbox"/> rein private religiöse/spirituelle Praxis	
<input type="checkbox"/> rein kognitive Dimensions (Belief, Orientierung, Einstellung etc.)	
<input type="checkbox"/> Mixed/Mehrdimensional	
<input type="checkbox"/> Ein-Item-Messung (1)	V20 V21 V22
<input type="checkbox"/> Multiple-Item-Messung (2)	
Anzahl der Items:	V20 V21 V22
Skalenbreite:	
Nähere Einordnung R/S Messung 1:	
<b><u>RS/Verhalten:</u></b>	V23
<input type="checkbox"/> Anzahl Besuch r/s Dienst (Kirchgang etc.) (1)	
<input type="checkbox"/> Teilnahme an anderem r/s Verhalten (öffentlich) (2)	
<input type="checkbox"/> R/S Verhalten - privat (Beten, Meditation, etc.) (3)	
<input type="checkbox"/> gemischtes r/s Verhalten (privat + öffentlich) (4)	
<input type="checkbox"/> gemischtes religiöses Verhalten (privat + öffentlich - Islam) (5)	
<input type="checkbox"/> gemischtes religiöses Verhalten (privat + öffentlich – Christ.) (6)	
<input type="checkbox"/> gemischtes religiöses Verhalten (privat + öffentlich – Jüd.) (7)	
<b><u>Einstellung:</u></b>	
<input type="checkbox"/> religiöse Selbsteinschätzung (8)	
<input type="checkbox"/> religiöse Erfahrung (9)	
<input type="checkbox"/> religiöse Salienz (10)	
<input type="checkbox"/> religiöser Glaube (belief) (11)	

religiöse Bindung (commitment) (12)

Intrapersonell – kognitiv (13)

Interpersonell - verhalten (14)

Religiöse Orientierung / Einstellung

intrinsisch (15)

extrinsisch (16)

intrinsisch-rev. (17)

extrinsisch-rev. persönlich (18)

extrinsisch-rev. sozial (19)

quest Orientierung (20)

religiöse Interaktion (21)

religiöse Werte (22)

Zentralität von Religion (23)

Involvement (24)

Wichtigkeit von Religion (25)

Wichtigkeit von Glauben (faith) (26)

Wichtigkeit von Gott (27)

**Spiritualitätsbereich:**

Wichtigkeit von Spiritualität (28)

selbsteingeschätzte Spiritualität (29)

spirituelle Erfahrung (30)

spirituelle Reife (31)

spirituelle Salienz (32)

spirituelle Aktivitäten / Praxis (33)

Spirituelle Orientierung / Einstellung (34)

Unterdimension Spiritualität n. n. b. (35)

**Multidimensionale M.:**

multidimensionale R/S-Messung (36)

multidimensionale R-Messung (37)

multidimensionale S-Messung (38)

**Anderes:**

Messinstrument spez. christ. Religiosität (39)

Messinstrument spez. muslimische Religiosität (40)

Messinstrument spez. jüdische Religiosität (41)

Messinstrument spez. Budh. Religiosität (42)

Mystizismus (43)

eine andere Form (44) und zwar:

V24

**Freie Bemerkung:**

V25

**Konkretes Messinstrument:**

V26

<b>H2.11 Wurde eine Messung von SWB für die Untersuchung genutzt?</b>	
<input type="checkbox"/> Ja (1) <input type="checkbox"/> Nein (2) – weiter zu H3	V70
<b>Mit wie vielen Messinstrumenten wurde Lebenszufriedenheit erfasst?</b> <b>Anzahl:</b>	V71
<b>H2.13 Einordnung Messinstrument Lebenszufriedenheit</b>	
<b>Wie wurde Lebenszufriedenheit 1 erfasst?</b> <input type="checkbox"/> 3-item version of the Satisfaction with Life Scale (SWLS) <input type="checkbox"/> 5-item version of the Satisfaction with Life Scale (SWLS) <input type="checkbox"/> Satisfaction with Life Scale (SWLS) – ohne nähere Angabe <input type="checkbox"/> Life Satisfaction Index-Z (LSI-Z 12) <input type="checkbox"/> Cantril Ladder / Ladder Scale of General Well-Being (1 Item -10er Skala) <input type="checkbox"/> Anderes Messinstrument: <b>Anzahl der Items:</b>	V72
<input type="checkbox"/> Eigenes Messinstrument: <b>Anzahl der Items:</b>	V73
<input type="checkbox"/> Ein-Item-Messung (1) <input type="checkbox"/> Multiple-Item-Messung (2) <input type="checkbox"/> keine Angabe (99)	V74
<b>Anzahl der Items:</b> <b>Skalenbreite:</b>	V75
<b>Anzahl der Items:</b> <b>Skalenbreite:</b>	V76 V77
<b>Freie Bemerkung LZ Messung 1:</b>	V91

Study-Level Moderatoren	
<b>H3.1 Alter</b>	
Arithmetisches Mittel in Jahren: <input type="checkbox"/> Keine Angabe (999)	V100
<b>H3.2 Geschlecht</b>	
Prozent männlich: <input type="checkbox"/> Keine Angabe (999)	V101
Prozent weiblich: <input type="checkbox"/> Keine Angabe (999)	V102
<b>H3.3 Makromerkmale</b>	
Human Development Index aus Erhebungsjahr: <input type="checkbox"/> Keine Angabe (999)	V103
Wichtigkeit von Religion nach Land aus Erhebungsjahr (WVS) <input type="checkbox"/> Keine Angabe (999)	V104
<b>H3.4 Stichprobeneigenschaften</b>	
Repräsentativität Bevölkerung eines Landes: <input type="checkbox"/> Ja (1) <input type="checkbox"/> Nein (2) <input type="checkbox"/> Keine Angabe (99)	V105
Ggf. Spezialpopulation: <input type="checkbox"/> Schüler (1) <input type="checkbox"/> Studenten (2) <input type="checkbox"/> Mitarbeiter (3) <input type="checkbox"/> Senioren (4) <input type="checkbox"/> willkürliche Stichprobe (5) <input type="checkbox"/> Strafgefangene (6) <input type="checkbox"/> Pflegende (7) und zwar von: <input type="checkbox"/> spezielle ethnische Gruppe (8) und zwar: <input type="checkbox"/> klinische Stichprobe (9) und zwar: <input type="checkbox"/> etwas anderes (10) und zwar: <input type="checkbox"/> etwas anderes (11)	V106 V107 V108 V109 V110
<b>H3.5 Art des Untersuchungsmodells</b>	
<input type="checkbox"/> Querschnittsmodell <input type="checkbox"/> Längsschnittmodell	V111
Anmerkung Längsschnitt:	V112

Mediatoren	
<b>Aus welchem Bereich stammen die Mediatoren? (Mehrfachwahl möglich)</b>	
<input type="checkbox"/> Theorie der sozialen Unterstützung/ Kohäsionshypothese <b>Wie viele Mediatorvariablen:</b>	M1 M2
<input type="checkbox"/> Theorie der kognitiven Orientierung / der Kohärenz <b>Wie viele Mediatorvariablen:</b>	M3 M4
<input type="checkbox"/> Theorie der positiven Gefühle / Emotionen / des Selbstwerts <b>Wie viele Mediatorvariablen:</b>	M5 M6
<input type="checkbox"/> Theorie der alternativen Werte <b>Wie viele Mediatorvariablen:</b>	M7 M8
<input type="checkbox"/> in der Empirie entdeckte <b>Wie viele Mediatorvariablen:</b>	M9 M10
<b>H4.1 Theorie der sozialen Unterstützung/ Kohäsionshypothese</b>	
<b>Welcher Bereich wurde erfasst?</b> <input type="checkbox"/> Soziale Unterstützung (1) =M1	M11
<b>Wie wurde soziale Unterstützung erfasst?</b> <input type="checkbox"/> Anzahl enger Freunde (1) <input type="checkbox"/> Duke Social Support Index – Subskalen (2) <input type="checkbox"/> Interpersonal Support Evaluation List (ISEL) – ohne self-esteem (3) <input type="checkbox"/> Multidimensional Scale of Perceived Social Support (MSPSS: Zimet et al. 1988) (4) <input type="checkbox"/> Inventory of Socially Supportive Behaviors (ISSB) (5) <input type="checkbox"/> Frequency of social interaction (6) <input type="checkbox"/> Network Analysis Profile (NAP; Cohen & Sokolovsky, 1979; Sokolovsky&Cohen, 1981) (7) <input type="checkbox"/> 12-item Brief Interpersonal Support Evaluation List (8) <input type="checkbox"/> Lubben Social Network Scale (LSNS-R; Lubben et al., 2001 ) (9) <input type="checkbox"/> Sense of Support Scale (SSS) (Dolbier and Steinhardt 2000). (10) <input type="checkbox"/> Social Provisions Scale (SPS; Cutrona & Russell, 1987) (11) <input type="checkbox"/> Social Support Questionnaire + SSQ6 (Sarason, Sarason, & Shearin, 1986) (12) <input type="checkbox"/> The Social Support Appraisals Scale (Vaux et al., 1986) (13) <input type="checkbox"/> Adult Social Support Questionnaire (14) <input type="checkbox"/> Personal Resource Questionnaire—PRQ2000 (Weinert 2003) (15) <input type="checkbox"/> Strong Tie Support Scale (Lin et al. 1986) (16) <input type="checkbox"/> Social Support for Recovery Scale (SSRS) (17) <input type="checkbox"/> APGAR scale – family and friends (18) <input type="checkbox"/> Kang’s (1985) Family Support Scale (19) <input type="checkbox"/> Social Support Survey (MOS) (20) <input type="checkbox"/> 2-Way Social Support Survey (Shakespeare-Finch & Obst, 2011) (21) <input type="checkbox"/> Perceived social support- family/friends (PSS-Fr) Procidano and Heller (1983) (22) <input type="checkbox"/> eigene Messung mehrere Items (23)	M12
<b>Anmerkungen:</b>	M13

<b>H4.2 Theorie der kognitiven Orientierung / der Kohärenz</b>	
<p><b>Welcher Bereiche wurden erfasst?</b></p> <p><input type="checkbox"/> Sense of Coherence / Kohärenzgefühl (1) =M2</p> <p><input type="checkbox"/> Hope / Hoffnung (1) =M3</p> <p><input type="checkbox"/> Optimism / Optimismus (1) = M4</p> <p><input type="checkbox"/> Purpose in Life (1) =M5</p> <p><input type="checkbox"/> Meaning in Life (1) = M6</p> <p><b>Anmerkungen:</b></p> <hr/>	<p>M14</p> <p>M15</p> <p>M16</p> <p>M17</p> <p>M18</p> <p>M19</p>
<p><b>Wie wurde Sense of Coherence / Kohärenzgefühl erfasst?</b></p> <p><input type="checkbox"/> Antonovsky (1987) - Orientation to Life Scale</p> <p><b>Anmerkungen:</b></p> <hr/>	<p>M20</p> <p>M21</p>
<p><b>Wie wurde Hope / Hoffnung erfasst?</b></p> <p><input type="checkbox"/> The Hope Scale (Snyder et al. 1991) (1)</p> <p><input type="checkbox"/> Herth Hope Index (HHI) (2)</p> <p><input type="checkbox"/> Children's Hope Scale (Snyder et al., 2002) (3)</p> <p><input type="checkbox"/> Positive Psychological Strengths Questionnaire (PPSQ); (Luthans et al. 2007) Unterskala Hoffnung (4)</p> <p><b>Anmerkungen:</b></p> <hr/>	<p>M22</p> <p>M23</p>
<p><b>Wie wurde Optimism / Optimismus erfasst?</b></p> <p><input type="checkbox"/> Arabic Scale of Optimism and Pessimism (ASOP) (1)</p> <p><input type="checkbox"/> The Life Orientation Test-Revised (LOT-R) (2)</p> <p><input type="checkbox"/> Positive Psychological Strengths Questionnaire (PPSQ) (3)</p> <p><input type="checkbox"/> General optimism -- &gt; eigene 10er Skala (4)</p> <p><b>Anmerkungen:</b></p> <hr/>	<p>M23_1</p> <p>M23_2</p>
<p><b>Wie wurde Purpose in Life erfasst?</b></p> <p><input type="checkbox"/> Purpose in Life Factor von: Ryff's(1989) Scales of Psychological Well-Being (1)</p>	<p>M24</p>



<input type="checkbox"/> Generosity / Großzügigkeit (1) = M9 <input type="checkbox"/> Gratitude / Dankbarkeit (1) = M10	M32 M33
<hr/>	
<p><b>Wie wurde Forgiveness / Vergebung erfasst?</b></p> <input type="checkbox"/> Brief Multidimensional Measure of Religiousness/Spirituality – Subscale – Forgiveness (1) <input type="checkbox"/> Forgiving Personality Inventory (FP; Jones, Kamat, Hopko, & Lawler-Row, in press) (2) <input type="checkbox"/> Heartland Forgiveness Scale (HFS; Thompson et al., 2005; Zhang, 2009) 3) <input type="checkbox"/> Forgiveness of Self (FOS) scale (Mauger et al., 1992) (4) <input type="checkbox"/> eigene Messung – mehrere Items (5) <p><b>Anmerkungen:</b></p>	M34
<hr/>	
<p><b>Wie wurde Generosity / Großzügigkeit erfasst?</b></p> <input type="checkbox"/> 10-item Interpersonal Generosity Scale (IGS) (1)	M35
<p><b>Anmerkungen:</b></p>	M36
<hr/>	
<p><b>Wie wurde Gratitude / Dankbarkeit erfasst?</b></p> <input type="checkbox"/> 6-item Gratitude Questionnaire (GQ) (1) <input type="checkbox"/> Gratitude, Resentment and Appreciation Test (GRAT; Gan, 2009; Watkins et al., 2003) (2)	M37
<p><b>Anmerkungen:</b></p>	M38
<hr/>	
<p><b>Anmerkungen:</b></p>	M39
<hr/>	
<b>H4.7 In der Empirie gefunden Mediatoren</b>	
<p><b>Welcher Bereiche wurde erfasst?</b></p> <input type="checkbox"/> Self-efficacy / Selbstwirksamkeitserwartung (1) =M11 <input type="checkbox"/> sense of control (Kontrollüberzeugung) / Mastery (Selbstbeherrschung) / Perceived Control of internal states (1) = M12	M40 M41
<hr/>	
<p><b>Wie wurde self efficacy / Selbstwirksamkeit erfasst?</b></p> <input type="checkbox"/> Arabic scale of self-efficacy (1) <input type="checkbox"/> Generalized Self-Efficacy Scale (GSES) / General Self-Efficacy Scale (SES) (Zhang & Schwarzer, 1995) (2)	M42

<p><input type="checkbox"/> Self-Efficacy Instrument developed by Sherer et al. (1982) - subskalen general self-efficacy and social (3)</p> <p><input type="checkbox"/> Positive Psychological Strengths Questionnaire (PPSQ) – self efficacy Unterskala (Luthans et al. 2007) (4)</p> <p><input type="checkbox"/> AMHS-IV includes a nine-item self-efficacy scale (5)</p> <p><input type="checkbox"/> eigene Messung – mehrere Items (6)</p> <p><b>Anmerkungen:</b></p> <hr/>	M43
<p><b>Wie wurde Self-Actualization / Mastery / sense of control erfasst?</b></p> <p><input type="checkbox"/> Adolescence Self-Control Scale (1)</p> <p><input type="checkbox"/> Pearlin Mastery Model Scale (Pearlin, Menghan, Lieberman, &amp; Mullan, 1981) (2)</p> <p><input type="checkbox"/> Locus of Control (Rotter 1966) (3)</p> <p><input type="checkbox"/> the control subscale of the Dispositional Resilience measure (Bartone, Ursano, Wright, &amp; Ingraham, 1989) (4)</p> <p><input type="checkbox"/> Environmental Mastery subscale of the Psychological Well-Being measure and the Perceived Social Control scale (5)</p> <p><input type="checkbox"/> perceived control of internal states scale (PCOISS; Pallant 2000) (6)</p> <p><input type="checkbox"/> eigene Messung – mehrere Items (7)</p> <p><input type="checkbox"/> Life Attitude Profile Scale (LAPS) Reker and Peacock (1981) – Unterskala life control</p> <p><b>Anmerkungen:</b></p>	M44
<p><b>Anmerkungen:</b></p>	M45

5. Korrelationsmatrix

a. Korrelationsmatrix mit Mediatoren

V	RS1	RS2	RS3	RS4	RS5	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	SWB <sub>1</sub>	SWB <sub>2</sub>	SWB <sub>3</sub>
RS1	1																			
RS2		1																		
RS3			1																	
RS4				1																
RS5					1															
M1						1														
M2							1													
M3								1												
M4									1											
M5										1										
M6											1									
M7												1								
M8													1							
M9														1						
M10															1					
M11																1				
M12																	1			
SWB <sub>1</sub>																		1		
SWB <sub>2</sub>																			1	
SWB <sub>3</sub>																				1

<b>Kodierüberprüfung</b>	
<b>Schweregrad der Kodierung</b>	
<input type="checkbox"/> Sehr schwer <input type="checkbox"/> Schwer <input type="checkbox"/> Eher schwer <input type="checkbox"/> Eher leicht <input type="checkbox"/> Leicht <input type="checkbox"/> Sehr leicht	Q1
<b>Zusammenfassende subjektive methodische Güte der Studie (sehr gut – sehr schlecht)</b>	
<input type="checkbox"/> Sehr gut <input type="checkbox"/> gut <input type="checkbox"/> eher gut <input type="checkbox"/> eher schlecht <input type="checkbox"/> schlecht <input type="checkbox"/> sehr schlecht	Q2
<b>Allgemeine Anmerkungen</b>	
	Q3

## Anhang B: Studienübersicht

K-ID	P-ID	S-ID	Autoren	Jahr	Land	N	M1	M4	M7	ES
1	001	001_1	Abdel-Khalek	2008	Kuwait	219				1
2	001	001_2	Abdel-Khalek	2008	Kuwait	205				1
3	002	002_1	Abdel-Khalek	2010	Kuwait	224				2
4	003	003_1	Abdel-Khalek	2010	Kuwait	487				2
5	004	004_1	Abdel-Khalek	2011	Kuwait	499			1	5
6	005	005_1	Abdel-Khalek	2011	Ägypten	224				2
7	006	006_1	Abdel-Khalek	2012	Ägypten	214			1	3
8	006	006_2	Abdel-Khalek	2012	Ägypten	363			1	3
9	006	006_3	Abdel-Khalek	2012	Kuwait	332			1	3
10	006	006_4	Abdel-Khalek	2012	Kuwait	342			1	3
11	007	007_1	Abdel-Khalek	2012	Kuwait	242				1
12	007	007_2	Abdel-Khalek	2012	Kuwait	235				1
13	007	007_3	Abdel-Khalek	2012	Kuwait	251				1
14	007	007_4	Abdel-Khalek	2012	Kuwait	258				1
15	007	007_5	Abdel-Khalek	2012	Kuwait	226				1
16	007	007_6	Abdel-Khalek	2012	Kuwait	208				1
17	008	008_1	Abdel-Khalek	2013	Katar	113				2
18	008	008_2	Abdel-Khalek	2013	Katar	133				2
19	009	009_1	Abdel-Khalek	2014	Libanon	111				2
20	009	009_2	Abdel-Khalek	2014	Libanon	128				2
21	010	010_1	Abdel-Khalek	2014	Katar	194				1
22	010	010_2	Abdel-Khalek	2014	Katar	178				1
23	011	011_1	Abdel-Khalek	2015	Libanon	261				2
24	011	011_2	Abdel-Khalek	2015	Libanon	215				2
25	012	012_1	Abdel-Khalek/Eid	2011	Kuwait	673				1
26	012	012_2	Abdel-Khalek/Eid	2011	Kuwait	1264				1
27	012	012_3	Abdel-Khalek/Eid	2011	Palä. Auto.	452				1
28	012	012_4	Abdel-Khalek/Eid	2011	Palä. Auto.	557				1
29	013	013_1	Abdel-Khalek/Lester	2007	Kuwait	460		1		1
30	013	013_2	Abdel-Khalek/Lester	2007	USA	274		1		1
31	014	014_1	Abdel-Khalek/Lester	2010	Kuwait	234		1	1	6
32	015	015_1	Abdel-Khalek/Lester	2012	USA	158				2
33	015	015_2	Abdel-Khalek/Lester	2012	Kuwait	192				2
34	017	017_1	Abdel-Khalek/Maltby	2008	Kuwait	271		1		3
35	017	017_2	Abdel-Khalek/Maltby	2008	UK	205		1		3
36	018	018_1	Abdel-Khalek/Naceur	2007	Algerien	109		1		3
37	018	018_2	Abdel-Khalek/Naceur	2007	Algerien	135		1		3
38	019	019_1	Abdel-Khalek/Pratap Singh	2014	Indien	200		1	1	9
39	019	019_2	Abdel-Khalek/Pratap Singh	2014	Indien	200		1	1	9
40	021	021_1	Abu-Raiya et al.	2015	USA	2208				1
41	022	022_1	Abu-Raiya et al.	2016	Australien	200				2
42	024	024_1	Aghababaei/Blachnio	2014	Polen	208				3
43	025	025_1	Aghababaei et al.	2016	Iran	474			1	9
44	026	026_1	Ai et al.	2004	USA	246		1		1
45	027	027_1	Ai et al.	2013	USA	484	1	1		7
46	028	028_1	Ai et al.	2005	USA	453		1		1
47	029	029_1	Ajibade et al.	2016	USA	199				1
48	030	030_1	Allen/Wang	2014	USA	267			1	5
49	032	032_1	Ayyash-Abdo/Sanchez-Ruiz	2012	Libanon	1401		1	1	3
50	033	033_1	Baetz et al.	2002	Kanada	88				3
51	034	034_1	Baker et al.	2013	Malaysia	1003				1
52	035	035_1	Baker et al.	2011	USA	247	1			1
53	036	036_1	Ball et al.	2003	USA	492			1	1
54	037	037_1	Bamonti et a.	2015	USA	50	1			1
55	038	038_1	Baroun	2006	Kuwait	941				3
57	040	040_1	Beach	2016	USA	193				1
58	041	041_1	Bergan/McConatha	2001	USA	320				1
59	042	042_1	Bernardo/Estrellado	2014	Philippinen	362			1	3
60	043	043_1	Bradshaw et al.	2015	USA	1024			1	1
61	044	044_1	Burroughs/Rindfleisch	2002	USA	373				1

62	045	045_1	Burton/Plaks/Peterson	2015	USA	619				1
63	046	046_1	Büssing et al.	2012	Deutschland	191				3
64	047	047_1	Byrd/Lear/Schwenka	2000	USA	150				3
65	049	049_1	Clark et al.	2012	USA	4128	1			3
66	050	050_1	Consedine/Magai/King	2004	USA	1118	1			1
67	051	051_1	Cotton et al.	2006	USA	450		1	1	7
68	052	052_1	Cowlshaw et al.	2013	Australien	324	1			3
69	054	054_1	Davis	2005	USA	130				1
70	056	056_1	Dengate	2009	Kanada	1245	1			2
71	058	058_1	Dezutter et al.	2010	Belgien	207				1
72	059	059_1	Dezutter/Soenens/Hutsebaut	2006	Belgien	472			1	11
74	061	061_1	Duffy et al.	2012	USA	472				1
75	062	062_1	Duffy/Sedlacek	2010	USA	2432				1
76	063	063_1	Edara	2013	USA	637				2
77	064	064_1	Ekas et. al	2009	USA	119		1	1	9
78	066	066_1	Fabricatore/Handal/Fenzel	2000	USA	120				1
79	067	067_1	Fiala/Bjoerck/Gorsuch	2002	USA	249	1			3
80	067	067_2	Fiala/Bjoerck/Gorsuch	2002	USA	93	1			3
81	068	068_1	Fiori et al.	2006	USA	397				3
82	070	070_1	Fitzell/Pakenham	2010	USA	622	1			1
83	072	072_1	Froehlich et al.	2006	USA	251	1			3
84	073	073_1	Galea et al.	2007	Malta	312				2
85	074	074_1	Gall et al.	2000	Kanada	32		1	1	6
86	077	077_1	Gernat	2007	USA	200				2
87	078	078_1	Gillum/Sullivan/Bybee	2006	USA	151	1		1	3
88	079	079_1	Halama/Martos/Adamovová	2010	Ungarn	249			1	3
89	079	079_2	Halama/Martos/Adamovová	2010	Slowakei	274			1	3
90	080	080_1	Han	2010	USA	274			1	3
91	080	080_2	Han	2010	USA	127			1	3
92	081	081_1	Hayward/Elliot	2011	USA	177				1
93	082	082_1	Headey et al.	2010	Deutschland	20000				2
94	083	083_1	Heaven/Ciarrochi	2007	Australien	289	1		1	3
95	083	083_2	Heaven/Ciarrochi	2007	Australien	274	1		1	3
96	084	084_1	Hebert et al.	2009	USA	284				3
97	085	085_1	Holder et al.	2016	Sambia	391				3
98	085	085_2	Holder et al.	2016	Sambia	902				3
99	088	088_1	Jackson/Bergeman	2011	USA	529				1
100	090	090_1	Jang/Johnson	2003	USA	2107			1	1
101	091	091_1	Jang et al.	2006	USA	255				1
102	092	092_1	Jaramillo	2007	USA	215				2
103	093	093_1	Joshanloo	2011	Iran	292				2
104	094	094_1	Joshanloo/Daemi	2015	Iran	322			1	1
105	095	095_1	Kamm-Steigelman et al.	2006	USA	59	1			3
106	096	096_1	Karaminia/Hashemian	2014	Iran	2010	1			1
107	097	097_1	Kelley/Miller	2007	USA	615				4
108	099	099_1	Kimwelli/Stilwell	2002	USA	470	1			1
109	100	100_1	Koenig et al.	2014	USA	129	1	1		3
110	101	101_1	Kogan et al.	2005	USA	318		1		4
111	102	102_1	Lager	2001	USA	247	1			1
112	103	103_1	Laudet/Morgan/White	2006	USA	353	1			3
113	104	104_1	Lawler-Row/Elliot	2009	USA	425	1			5
114	106	106_1	Lechner et al.	2013	Polen	1581				2
115	107	107_1	Lee	2007	USA	145	1			5
116	108	108_1	Lee/Woo	2013	USA	147	1			1
117	109	109_1	Leondari/Gialamas	2009	Griechenland	363				3
118	110	110_1	Levin	2002	USA	205			1	3
119	111	111_1	Linnabery et al.	2014	USA	188	1			1
120	114	114_1	Lynch	2012	Irland	100			1	1
121	115	115_1	Maheshwari/Singh	2009	Indien	94				1
122	115	115_2	Maheshwari/Singh	2009	Indien	60				1
123	116	116_1	Maltby/Day	2000	UK	172		1	1	7
124	116	116_2	Maltby/Day	2000	UK	188		1	1	7
125	117	117_1	Marques/Lopez/Mitchell	2013	Portugal	227				2

128	120	120_1	McGowan et al.	2016	USA	143	1			2
129	121	121_1	Meulemann	2001	Deutschland	1596				2
130	123	123_1	Moore	2013	Irland	100				2
131	124	124_1	Morgan Consoli et al.	2015	USA	121	1			1
132	125	125_1	Williams Morris et al.	2010	USA	235		1		1
133	127	127_1	Chi Wai Ng/Fisher	2016	Hongkong	268				1
134	128	128_1	Omu	2010	Kuwait	40				1
135	129	129_1	Osborne/Milojev/Sibley	2016	Neuseeland	1486				3
136	131	131_1	Park	2005	USA	169				1
137	132	132_1	Park/Bauer/Oescher	2001	USA	1406			1	2
138	132	132_2	Park/Bauer/Oescher	2001	USA	2922			1	2
139	132	132_3	Park/Bauer/Oescher	2001	USA	2260			1	2
140	132	132_4	Park/Bauer/Oescher	2001	USA	14024			1	2
141	133	133_1	Park/Roh/Yeo	2012	USA	200	1			5
142	134	134_1	Perkins et al.	2007	USA	127		1		3
143	135	135_1	Poulin et al.	2009	USA	1382	1			5
144	136	136_1	Prasad	2003	USA	30	1			3
145	136	136_2	Prasad	2003	USA	30	1			3
146	138	138_1	Reed/Neville	2014	USA	137				3
147	139	139_1	Robins/Fiske	2009	USA	454	1			4
148	140	140_1	Rogers et al.	2012	USA	167				1
149	141	141_1	Roh et al.	2014	USA	200				2
150	142	142_1	Roscoe et al.	2009	USA	17	1			1
151	143	143_1	Salsman et al.	2005	USA	217	1	1		15
152	143	143_2	Salsman et al.	2005	USA	159	1	1		15
153	144	144_1	Schuurmans-Stekhoven	2005	Australien	265				1
156	146	146_1	Smith et al.	2010	USA	289		1		3
157	146	146_2	Smith et al.	2010	USA	259		1		3
158	147	147_1	Snowshoe et al.	2017	Kanada	290				1
159	148	148_1	Ahmadi et al.	2014	Iran	210	1			1
160	149	149_1	Steenwyk et al.	2010	USA	254				4
161	150	150_1	Steger/Frazier	2008	USA	508				1
162	152	152_1	Steger et al.	2010	USA	284				4
163	153	153_1	Suhail/Chaudhry	2004	Pakistan	973	1		1	6
164	154	154_1	Tae/Heitkemper/Kim	2012	Südkorea	214			1	1
165	156	156_1	Tiliouine	2012	Algerien	3173				1
166	157	157_1	Thomas et al.	2016	Spanien	224				3
167	160	160_1	Vitale	2015	USA	340	1			5
168	161	161_1	Wachholtz/Rogoff	2013	USA	259				2
169	162	162_1	Wahlen	2013	Irland	111				1
170	163	163_1	Walsh et al.	2007	USA	100				4
171	164	164_1	Wang et al.	2016	China	1812	1			2
172	165	165_1	Warren et al.	2015	USA	553		1		1
173	167	167_1	Weinert-Moreno	2012	USA	38				3
174	167	167_2	Weinert-Moreno	2012	USA	104				3
175	168	168_1	White/Driver/Warren	2010	USA	42				1
176	169	169_1	Wilkins et al.	2012	USA	558				4
177	170	170_1	Wilt et al.	2016	USA	418			1	3
178	170	170_2	Wilt et al.	2016	USA	965			1	3
179	171	171_1	Wnuk et al.	2012	Polen	50	1			1
180	173	173_1	Wyshak	2016	USA	93679	1			1
181	174	174_1	Yoon/Lee	2006	USA	215	1			5
182	175	175_1	Zullig/Ward/Horn	2006	USA	522				5

**Anmerkungen:** P-ID=Publikationsidentifikationsnummer; S-ID=Stichprobenidentifikationsnummer M1=soziale Unterstützung; M4=Optimismus; M7=Selbstwert.

## Anhang C: Studien in der Meta-Analyse

- Abdel-Khalek, Ahmed M. 2008: Religiosity, health, and well-being among Kuwaiti personnel. In: *Psychological reports* 102 (1), 181–184.
- Abdel-Khalek, Ahmed M. 2010: Quality of life, subjective well-being, and religiosity in Muslim college students. In: *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation* 19 (8), 1133–1143.
- Abdel-Khalek, Ahmed M. 2010: Religiosity, subjective well-being, and neuroticism. In: *Mental Health, Religion & Culture* 13 (1), 67–79.
- Abdel-Khalek, Ahmed M. 2011: Religiosity, subjective well-being, self-esteem, and anxiety among Kuwaiti Muslim adolescents. In: *Mental Health, Religion & Culture* 14 (2), 129–140.
- Abdel-Khalek, Ahmed M. 2011: Subjective well-being and religiosity in Egyptian college students. In: *Psychological reports* 108 (1), 54–58.
- Abdel-Khalek, Ahmed M. 2012: Associations between religiosity, mental health, and subjective well-being among Arabic samples from Egypt and Kuwait. In: *Mental Health, Religion & Culture* 15 (8), 741–758.
- Abdel-Khalek, Ahmed M. 2012: Subjective well-being and religiosity. A cross-sectional study with adolescents, young and middle-age adults. In: *Mental Health, Religion & Culture* 15 (1), 39–52.
- Abdel-Khalek, Ahmed M. 2013: The relationships between subjective well-being, health, and religiosity among young adults from Qatar. In: *Mental Health, Religion & Culture* 16 (3), 306–318.
- Abdel-Khalek, Ahmed M. 2014: Happiness, health, and religiosity. Significant associations among Lebanese adolescents. In: *Mental Health, Religion & Culture* 17, 30–38.
- Abdel-Khalek, Ahmed M. 2014: Religiosity, health and happiness. Significant relations in adolescents from Qatar. In: *The International journal of social psychiatry* 60 (7), 656–661.
- Abdel-Khalek, Ahmed M./Duregger, Cornelia 2015: Happiness, health, and religiosity among Lebanese young adults. In: *Cogent Psychology* 2 (1), 233.
- Abdel-Khalek, Ahmed M./Eid, Ghada K. 2011: Religiosity and its association with subjective well-being and depression among Kuwaiti and Palestinian Muslim children and adolescents. In: *Mental Health, Religion & Culture* 14 (2), 117–127.
- Abdel-Khalek, Ahmed M./Lester, David 2007: Religiosity, health, and psychopathology in two cultures. Kuwait and USA. In: *Mental Health, Religion & Culture* 10 (5), 537–550.
- Abdel-Khalek, Ahmed M./Lester, David 2010: Personal and Psychological Correlates of Happiness Among a Sample of Kuwaiti Muslim Students. In: *Journal of Muslim Mental Health* 5 (2), 194–209.
- Abdel-Khalek, Ahmed M./Lester, David 2012: Constructions of religiosity, subjective well-being, anxiety, and depression in two cultures. Kuwait and USA. In: *The International journal of social psychiatry* 58 (2), 138–145.
- Abdel-Khalek, Ahmed M./Maltby, John 2008: The comparison of predictors of death obsession within two cultures. In: *Death studies* 32 (4), 366–377.
- Abdel-Khalek, Ahmed M./Naceur, Farida 2007: Religiosity and its association with positive and negative emotions among college students from Algeria. In: *Mental Health, Religion & Culture* 10 (2), 159–170.
- Abdel-Khalek, Ahmed M./Singh, Ajai P. 2014: Religiosity , Subjective Well-Being and Anxiety in a Sample of Indian University Students = التدين و الرفاه و القلق لدى عينة من طلاب الجامعة الهنود. In: *The Arab Journal of Psychiatry* 25 (2), 201–208.

- Abu-Rayya, Hisham M./Almoty, Shayma/White, Fiona A./Abu-Rayya, Maram H. 2016: The Interconnection Between Islamic Religiosity and Deviancy Among Australian Muslim Youth. A Partial Mediation Role of Life Satisfaction. In: *International Journal for the Psychology of Religion* 26 (4), 337–347.
- Abu-Rayya, Hisham M./Pargament, Kenneth I./Krause, Neal/Ironson, Gail 2015: Robust links between religious/spiritual struggles, psychological distress, and well-being in a national sample of American adults. In: *The American journal of orthopsychiatry* 85 (6), 565–575.
- Aghababaei, Naser/Błachnio, Agata 2014: Purpose in life mediates the relationship between religiosity and happiness. Evidence from Poland. In: *Mental Health, Religion & Culture* 17 (8), 827–831.
- Aghababaei, Naser/Sohrabi, Faramarz/Eskandari, Hossein/Borjali, Ahmad/Farrokhi, Noorali/Chen, Zhuo J. 2016: Predicting subjective well-being by religious and scientific attitudes with hope, purpose in life, and death anxiety as mediators. In: *Personality and Individual Differences* 90, 93–98.
- Ai, Amy L./Hall, Daniel/Pargament, Kenneth I./Tice, Terrence N. 2013: Posttraumatic growth in patients who survived cardiac surgery. The predictive and mediating roles of faith-based factors. In: *Journal of behavioral medicine* 36, 186–198.
- Ai, Amy L./Peterson, Christopher/Tice, T. N./Bolling, S. F./Koenig, H. G. 2004: Faith-based and secular pathways to hope and optimism subconstructs in middle-aged and older cardiac patients. In: *Journal of health psychology* 9 (3), 435–450.
- Ai, Amy L./Tice, Terrence N./Peterson, Christopher/Huang, Bu 2005: Prayers, spiritual support, and positive attitudes in coping with the September 11 national crisis. In: *Journal of personality* 73 (3), 763–791.
- Ajibade, Adebayo/Hook, Joshua N./Utsey, Shawn O./Davis, Don E./van Tongeren, Daryl R. 2016: Racial/Ethnic Identity, Religious Commitment, and Well-Being in African Americans. In: *Journal of black psychology* 42 (3), 244–258.
- Al Zaben, Faten/Khalifa, Doaa A./Sehlo, Mohammad G./Al Shohaib, Saad/Binzaqr, Salma A./Badreg, Alae M./Alsaadi, Rawan A./Koenig, Harold G. 2015: Religious Involvement and Health in Dialysis Patients in Saudi Arabia. In: *Journal of religion and health* 54 (2), 713–730.
- Allen, G. E. K./Wang, Kenneth T. 2014: Examining Religious Commitment, Perfectionism, Scrupulosity, and Well-Being Among LDS Individuals. In: *Psychology of religion and Spirituality* 6 (3), 257–264.
- Ayyash-Abdo, Huda/Sanchez-Ruiz, Maria-Jose 2012: Subjective wellbeing and its relationship with academic achievement and multilinguality among Lebanese university students. In: *International journal of psychology* 47 (3), 192–202.
- Baetz, Marilyn/Larson, David B./Marcoux, Gene/Bowen, Rudy/Griffin, Ron 2002: Canadian psychiatric inpatient religious commitment. An association with mental health. In: *Canadian journal of psychiatry* 47 (2), 159–166.
- Baker, Andrew M./Moschis, George P./Ong, Fon S./Pattanapanyasat, Ra-pee 2013: Materialism and Life Satisfaction. The Role of Stress and Religiosity. In: *Journal of consumer affairs* 47 (3), 548–563.
- Baker, Tamara A./Buchanan, NiCole T./Small, Brent J./Hines, Resche D./Whitfield, Keith E. 2011: Identifying the Relationship Between Chronic Pain, Depression, and Life Satisfaction in Older African Americans. In: *Research on Aging* 33 (4), 426–443.
- Ball, J./Armistead, L./Austin, B. J. 2003: The relationship between religiosity and adjustment among African-American, female, urban adolescents. In: *Journal of Adolescence* 26 (4), 431–446.

- Bamonti, Patricia/Lombardi, Sarah/Duberstein, Paul R./King, Deborah A./van Orden, Kimberly A. 2016: Spirituality attenuates the association between depression symptom severity and meaning in life. In: *Aging & Mental Health* 20 (5), 494–499.
- Baroun, Khader A. 2006: Relations among religiosity, health, happiness, and anxiety for Kuwaiti adolescents. In: *Psychological reports* 99 (3), 717–722.
- Beach, Victoria L. 2016: Religiosity and Prayer In Relation to Health and Life Satisfaction In Older Adults. Abschlussarbeit. Saint John's University. New York.
- Bergan, Anne/McConatha, Jasmin T. 2001: Religiosity and Life Satisfaction. In: *Activities, Adaptation & Aging* 24 (3), 23–34.
- Bernardo, Allan B. I./Estrellado, Alicia F. 2014: Measuring Hope in the Philippines. Validating the Short Version of the Locus-of-Hope Scale in Filipino. In: *Social Indicators Research* 119, 1649–1661.
- Bradshaw, Matt/Ellison, Christopher G./Fang, Qijuan/Mueller, Collin 2015: Listening to Religious Music and Mental Health in Later Life. In: *The Gerontologist* 55 (6), 961–971.
- Burroughs, J. E./Rindfleisch, A. 2002: Materialism and well-being. A conflicting values perspective. In: *Journal of consumer research* 29 (3), 348–370.
- Burton, Caitlin M./Plaks, Jason E./Peterson, Jordan B. 2015: Why Do Conservatives Report Being Happier Than Liberals? The Contribution of Neuroticism. In: *Journal of Social and Political Psychology* 3 (1), 89–102.
- Büssing, Arndt/Hedtstück, Anemone/Khalsa, Sat B. S./Ostermann, Thomas/Heusser, Peter 2012: Development of Specific Aspects of Spirituality during a 6-Month Intensive Yoga Practice. In: *Evidence-based complementary and alternative medicine : eCAM* 2012, 981523.
- Byrd, Kevin R./Lear, Delbert/Schwenka, Stacy 2000: Mysticism as a Predictor of Subjective Well-Being. In: *International Journal for the Psychology of Religion* 10 (4), 259–269.
- Clark, Cari J./Henderson, Kimberly M./Leon, Carlos F. M. de/Guo, Hongfei/Lunos, Scott/Evans, Denis A./Everson-Rose, Susan A. 2012: Latent constructs in psychosocial factors associated with cardiovascular disease. An examination by race and sex. In: *Frontiers in psychiatry* 3, 5.
- Consedine, Nathan S./Magai, Carol/Conway, Francine 2004: Predicting ethnic variation in adaptation to later life: styles of socioemotional functioning and constrained heterotypy. In: *Journal of Cross-Cultural Gerontology* 19 (2), 97–131.
- Cotton, Sian/Puchalski, Christina M./Sherman, Susan N./Mrus, Joseph M./Peterman, Amy H./Feinberg, Judith/Pargament, Kenneth I./Justice, Amy C./Leonard, Anthony C./Tsevat, Joel 2006: Spirituality and religion in patients with HIV/AIDS spirituality and religion in patients with HIV/AIDS. In: *Journal of general internal medicine* 21 (5), S5-S13.
- Cowlshaw, Sean/Niele, Sylvia/Teshuva, Karen/Browning, Colette/Kendig, Hal 2013: Older adults' spirituality and life satisfaction. A longitudinal test of social support and sense of coherence as mediating mechanisms. In: *Ageing & Society* 33 (7), 1243–1262.
- Davis, Bonnie 2005: Mediators of the relationship between hope and well-being in older adults. In: *Clinical nursing research* 14 (3), 253–272.
- Dengate, Jennifer 2009: The effects of religious commitment on well-being. Masterarbeit. University of Manitoba.
- Dezutter, Jessie/Robertson, Linda A./Luyckx, Koen/Hutsebaut, Dirk 2010: Life satisfaction in chronic pain patients. The stress-buffering role of the centrality of religion. In: *Journal for the Scientific Study of Religion* 49 (3), 507–516.
- Dezutter, Jessie/Soenens, Bart/Hutsebaut, Dirk 2006: Religiosity and mental health. A further exploration of the relative importance of religious behaviors vs. religious attitudes. In: *Personality and Individual Differences* 40 (4), 807–818.

- Duffy, Ryan D./Allan, Blake A./Bott, Elizabeth M. 2012: Calling and Life Satisfaction Among Undergraduate Students. Investigating Mediators and Moderators. In: *Journal of Happiness Studies* 13, 469–479.
- Duffy, Ryan D./Sedlacek, William E. 2010: The Salience of a Career Calling Among College Students. Exploring Group Differences and Links to Religiousness, Life Meaning, and Life Satisfaction. In: *Career development quarterly* 59 (1), 27–41.
- Edara, Inna R. 2013: Spirituality's Unique Role In Positive Affect, Satisfaction With Life, And Forgiveness Over And Above Personality And Individualism-Collectivism. In: *Research in the Social Scientific Study of Religion* 24, 15–41.
- Ekas, Naomi V./Whitman, Thomas L./Shivers, Carolyn 2009: Religiosity, Spirituality, and Socioemotional Functioning in Mothers of Children with Autism Spectrum Disorder. In: *Journal of autism and developmental disorders* 39 (5), 706–719.
- Fabricatore, Anthony N./Handal, Paul J./Fenzel, L. M. 2000: Personal spirituality as a moderator of the relationship between stressors and subjective well-being. In: *Journal of Psychology and Theology* 28 (3), 221.
- Fiala, William E./Bjorck, Jeffrey P./Gorsuch, Richard 2002: The Religious Support Scale: construction, validation, and cross-validation. In: *American journal of community psychology* 30 (6), 761–786.
- Fiori, Katherine L./Brown, Edna E./Cortina, Kai S./Antonucci, Toni C. 2006: Locus of control as a mediator of the relationship between religiosity and life satisfaction. Age, race, and gender differences. In: *Mental Health, Religion & Culture* 9 (3), 239–263.
- Fitzell, Amber/Pargament, Kenneth I. 2010: Application of a stress and coping model to positive and negative adjustment outcomes in colorectal cancer caregiving. In: *Psycho-oncology* 19 (11), 1171–1178.
- Froehlich, James P./Fialkowski, Geraldine M./Scheers, N. J./Wilcox, Peter C./Lawrence, Richard T. 2006: Spiritual Maturity and Social Support in a National Study of a Male Religious Order. In: *Pastoral Psychology* 54 (5), 465–478.
- Galea, Michael/Ciarrocchi, Joseph W./Piedmont, Ralph L./Wicks, Robert J. 2007: Child abuse, personality, and spirituality as predictors of happiness in maltese college students. In: Piedmont, Ralph (Hrsg.): *Research in the social scientific study of religion. Research in the social scientific study of religion*. Leiden: Brill, 141–154.
- Gall, Terry L./Miguez de Renart, Rosa M./Boonstra, Bonnie 2000: Religious Resources in Long-Term Adjustment to Breast Cancer. In: *Journal of Psychosocial Oncology* 18 (2), 21–37.
- Gernat, Carol A. 2003: *The Relationship Between Daily Spiritual Experience and Practice, and Health and Life Satisfaction in Doctoral-Level Counselors*. Dissertation. Seton Hall University.
- Gillum, Tameka L./Sullivan, Cris M./Bybee, Deborah I. 2006: The importance of spirituality in the lives of domestic violence survivors. In: *Violence against women* 12 (3), 240–250.
- Halama, Peter/Martos, Tamas/Adamovova, Lucia 2010: Religiosity and well-being in Slovak and Hungarian student samples: The role of personality traits. In: *Studia psychologica* 52 (2), 101–115.
- Han, GiBaeg 2010: *Adult Attachment Patterns, Mental Representation of Self, and Faith*. Dissertation.
- Hayward, R. D./Elliott, Marta 2011: Subjective and objective fit in religious congregations. Implications for well-being. In: *Group processes & intergroup relations* 14 (1), 127–139.

- Headey, Bruce/Schupp, Juergen/Tucci, Ingrid/Wagner, Gert G. 2010: Authentic happiness theory supported by impact of religion on life satisfaction. A longitudinal analysis with data for Germany. In: *The Journal of Positive Psychology* 5 (1), 73–82.
- Heaven, Patrick C. L./Ciarrochi, Joseph 2007: Personality and religious values among adolescents. A three-wave longitudinal analysis. In: *British Journal of Psychology* 98, 681–694.
- Hebert, Randy/Zdaniuk, Bozena/Schulz, Richard/Scheier, Michael 2009: Positive and negative religious coping and well-being in women with breast cancer. In: *Journal of palliative medicine* 12 (6), 537–545.
- Holder, Mark D./Coleman, Ben/Krupa, Tim/Krupa, Eugene 2016: Well-Being's Relation to Religiosity and Spirituality in Children and Adolescents in Zambia. In: *Journal of Happiness Studies* 17 (3), 1235–1253.
- Jackson, Brenda R./Bergeman, C. S. 2011: How Does Religiosity Enhance Well-Being? The Role of Perceived Control. In: *Psychology of religion and Spirituality* 3 (2), 149–161.
- Jang, Sung J./Johnson, Byron R. 2003: Strain, negative emotions, and deviant coping among African Americans. A test of general strain theory. In: *Journal of quantitative criminology* 19 (1), 79–105.
- Jang, Yuri/Borenstein, Amy R./Chiriboga, David A./Phillips, Karon/Mortimer, James A. 2006: Religiosity, Adherence to Traditional Culture, and Psychological Well-Being Among African American Elders. In: *Journal of Applied Gerontology* 25 (5), 343–355.
- Jaramillo, Matthew 2007: Calling, life satisfaction and job satisfaction. Masterarbeit. Colorado State University. Fort Collins, Colorado.
- Joshanloo, Mohsen 2011: Investigation of the Contribution of Spirituality and Religiousness to Hedonic and Eudaimonic Well-Being in Iranian Young Adults. In: *Journal of Happiness Studies* 12 (6), 915–930.
- Joshanloo, Mohsen/Daemi, Fatemeh 2015: Self-esteem mediates the relationship between spirituality and subjective well-being in Iran. In: *International journal of psychology* 50 (2), 115–120.
- Kamm-Steigleman, Lucia/Kimble, Laura P./Dunbar, Sandra/Sowell, Richard L./Bairan, Annette 2006: Religion, relationships and mental health in midlife women following acute myocardial infarction. In: *Issues in mental health nursing* 27 (2), 141–159.
- Karaminia, Reza/Hashemian, Sayedeh A. 2014: Prediction of war veteran mental health based on spiritual well-being, social support and self-efficacy variables. The mediating role of life satisfaction. In: *Journal of education and health promotion* 57 (3), 35–40.
- Kelley, Brien S./Miller, Lisa 2007: Life satisfaction and spirituality in adolescents. In: Piedmont, Ralph (Hrsg.): *Research in the social scientific study of religion. Research in the social scientific study of religion*. Leiden: Brill, 233–262.
- Kimweli, David/Stilwell, Wiliam 2002: Community Subjective Well-Being, Personality Traits and Quality of Life Therapy. In: Michalos, Alex/Diener, Ed/Glatzer, Wolfgang et al. (Hrsg.): *Advances in Quality of Life Research 2001. Social Indicators Research Series*. Dordrecht: Springer, 193–226.
- Koenig, Harold G./Berk, Lee S./Daher, Noha S./Pearce, Michelle J./Bellinger, Denise L./Robins, Give J./Nelson, Bruce/Shaw, Sally F./Cohen, Harvey J./King, Michael B. 2014: Religious involvement is associated with greater purpose, optimism, generosity and gratitude in persons with major depression and chronic medical illness. In: *Journal of psychosomatic research* 77 (2), 135–143.
- Kogan, Steven M./Luo, Zupei/Murry, Velma M./Brody, Gene H. 2005: Risk and Protective Factors for Substance Use among African American High School Dropouts. In: *Psychology of Addictive Behaviors* 19, 382–391.

- Lager, Julia M. 2006: Relationship among religious coping, psychosocial factors, and quality of life in individuals with type 2 diabetes. Dissertation. Texas A&M University. College Station.
- Laudet, Alexandre B./Morgen, Keith/White, William L. 2006: The Role of Social Supports, Spirituality, Religiousness, Life Meaning & Affiliation with 12-Step Fellowships in Quality of Life Satisfaction Among Individuals in Recovery from Alcohol & Drug Problems. In: *Alcoholism Treatment Quarterly* 24, 33–73.
- Lawler-Row, Kathleen A./Elliott, Jeff 2009: The Role of Religious Activity and Spirituality in the Health and Well-being of Older Adults. In: *Journal of health psychology* 14 (1), 43–52.
- Lechner, Clemens M./Tomasik, Martin J./Silbereisen, Rainer K./Wasilewski, Jacek 2013: Exploring the Stress-Buffering Effects of Religiousness in Relation to Social and Economic Change. Evidence From Poland. In: *Psychology of religion and Spirituality* 5 (3), 145–156.
- Lee, Eun-Kyoung O. 2007: Religion and Spirituality as Predictors of Well-Being Among Chinese American and Korean American Older Adults. In: *Journal of Religion, Spirituality & Aging* 19 (3), 77–100.
- Lee, Kyoung H./Woo, Hyeyoung 2013: Stressors, social support, religious practice, and general well-being among Korean adult immigrants. In: *Journal of Evidence-Based Social Work* 10 (5), 421–434.
- Leondari, Angeliki/Gialamas, Vasilios 2009: Religiosity and psychological well-being. In: *International journal of psychology* 44 (4), 241–248.
- Levin, Jeff 2002: Is depressed affect a function of one's relationship with god? Findings from a study of primary care patients. In: *The International Journal of Psychiatry in Medicine* 32 (4), 379–393.
- Linnabery, Eileen/Stuhlmacher, Alice F./Towler, Annette 2014: From Whence Cometh Their Strength. Social Support, Coping, and Well-Being of Black Women Professionals. In: *Cultural diversity & ethnic minority psychology* 20 (4), 541–549.
- Lynch, James 2012: Religiosity. Bachelorarbeit. DBS School of Arts. Dublin.
- Maheshwari, Saurabh/Singh, Purnima 2009: Psychological well-being and pilgrimage. Religiosity, happiness and life satisfaction of Ardh -Kumbh Mela pilgrims (Kalpvasis) at Prayag, India. In: *Asian journal of social psychology* 12, 285–292.
- Maltby, John/Day, Liza 2000: Depressive symptoms and religious orientation. Examining the relationship between religiosity and depression within the context of other correlates of depression. In: *Personality and Individual Differences* 28 (2), 383–393.
- Marques, Susana C./Lopez, Shane J./Mitchell, Joanna 2013: The Role of Hope, Spirituality and Religious Practice in Adolescents' Life Satisfaction. Longitudinal Findings. In: *Journal of Happiness Studies* 14 (1), 251–261.
- McGowan, Joseph C./Midlarsky, Elizabeth/Morin, Ruth T./Graber, Liat S. 2016: Religiousness and Psychological Distress in Jewish and Christian Older Adults. In: *clinical gerontologist* 39 (5), 489–507.
- Meulemann, Heiner 2001: Religiosität, Anpassungsbereitschaft und Lebenszufriedenheit. Traditionelle und moderne Formen der Bewältigung von Misserfolg und ihr Einfluss auf die Lebenszufriedenheit ehemaliger Gymnasiasten im 43. Lebensjahr. In: *Zeitschrift für Soziologie der Erziehung und Sozialisation* 21, 227–242.
- Moore, Gavin 2013: The relationship between religious orientation, coping style, and psychological health on death anxiety and life satisfaction. Bachelorarbeit. DBS School of Arts. Dublin.

- Morgan Consoli, M. L./Delucio, K./Noriega, E./Llamas, J. 2015: Predictors of Resilience and Thriving Among Latina/o Undergraduate Students. In: *Hispanic journal of behavioral sciences* 37 (3), 304–318.
- Morris, Ruth W./Martin, Bess/Hopson, Jamal/Welch-Murphy, Kristin 2010: Besides That I'm Ok. Well-Being in Caribbean and American Adolescents and Youth. In: *Journal of Research on Christian Education* 19 (1), 56–78.
- Ng, Eddie C. W./Fisher, Adrian T. 2016: Protestant Spirituality and Well-Being of People in Hong Kong. The Mediating Role of Sense of Community. In: *Applied Research Quality Life* 11 (4), 1253–1267.
- Omu, Onutobor 2010: Life satisfaction, self-efficacy and religious faith in stroke patients living in Kuwait. Dissertation.
- Osborne, Danny/Milojev, Petar/Sibley, Chris G. 2016: Examining the indirect effects of religious orientations on well-being through personal locus of control. In: *European journal of social psychology* 46 (4), 492–505.
- Park, Crystal L. 2005: Religion as a Meaning-Making Framework in Coping with Life Stress. In: *Journal of Social Issues* 61, 707–729.
- Park, Hae-Seong/Bauer, Scott/Oescher, Jeffrey 2001: Religiousness as a predictor of alcohol use in high school students. In: *Journal of drug education* 31 (3), 289–303.
- Park, Jisung/Roh, Soonhee/Yeo, Younsook 2012: Religiosity, Social Support, and Life Satisfaction Among Elderly Korean Immigrants. In: *The Gerontologist* 52 (5), 641–649.
- Perkins, Elizabeth A./Small, Brent J./Balducci, Lodovico/Extermann, Martine/Robb, Claire/Haley, William E. 2007: Individual differences in well-being in older breast cancer survivors. In: *Critical reviews in oncology/hematology* 62 (1), 74–83.
- Poulin, Michael J./Silver, Roxane C./Gil-Rivas, Virginia/Holman, E. A./McIntosh, Daniel N. 2009: Finding Social Benefits After a Collective Trauma. Perceiving Societal Changes and Well-Being Following 9/11. In: *Journal of traumatic stress* 22 (2), 81–90.
- Prasad, Ravi 2003: Religious commitment and health locus of control as moderators of depression and life satisfaction in individuals who have experienced a traumatic health event. Dissertation. Texas Tech University.
- Reed, Tamilia D./Neville, Helen A. 2014: The Influence of Religiosity and Spirituality on Psychological Well-Being Among Black Women. In: *Journal of black psychology* 40 (4), 384–401.
- Robins, Alee/Fiske, Amy 2009: Explaining the Relation between Religiousness and Reduced Suicidal Behavior. Social Support Rather Than Specific Beliefs. In: *Suicide and Life-Threatening Behavior* 39, 386–395.
- Rogers, Darrin L./Skidmore, Susan T./Montgomery, Gary T./Reidhead, Mary A./van Reidhead, A. 2012: Spiritual integration predicts self-reported mental and physical health. In: *Journal of religion and health* 51 (4), 1188–1201.
- Roh, Soonhee/Lee, Yeon-Shim/Lee, Jae H./Martin, James I. 2014: Typology of religiosity/spirituality in relation to perceived health, depression, and life satisfaction among older Korean immigrants. In: *Aging & Mental Health* 18 (4), 444–453.
- Roscoe, Lori A./Corsentino, Elizabeth/Watkins, Shirley/McCall, Marcia/Sanchez-Ramos, Juan 2009: Well-Being of Family Caregivers of Persons with Late-Stage Huntington's Disease. Lessons in Stress and Coping. In: *Health communication* 24 (3), 239–248.
- Salsman, John M./Brown, Tamara L./Brechtling, Emily H./Carlson, Charles R. 2005: The link between religion and spirituality and psychological adjustment: the mediating role of optimism and social support. In: *Personality & social psychology bulletin* 31 (4), 522–535.

- Schuurmans-Stekhoven, James 2010: "Moved by the Spirit". Does Spirituality Moderate the Interrelationships Between Subjective Well-Being Subscales? In: *Journal of clinical psychology* 66 (7), 709–725.
- Smith, B. W./Tooley, E. M./Christopher, P. J./Kay, V. S. 2010: Resilience as the ability to bounce back from stress. A neglected personal resource? In: *The Journal of Positive Psychology* 5 (3), 166–176.
- Snowshoe, Angela/Crooks, Claire V./Tremblay, Paul F./Hinson, Riley E. 2017: Cultural Connectedness and Its Relation to Mental Wellness for First Nations Youth. In: *The Journal of Primary Prevention* 38 (1-2), 67–86.
- Soltani, Mohsen A. T./Karaminia, Reza/Hashemian, Sayedeh A. 2014: Prediction of war veteran's mental health based on spiritual well-being, social support and self-efficacy variables. The mediating role of life satisfaction. In: *Journal of education and health promotion* 3, 57. In: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4114000>.
- Steenwyk, Sherry A. M./Atkins, David C./Bedics, Jamie D./Whitley, Jr., Bernard E. 2010: Images of God as They Relate to Life Satisfaction and Hopelessness. In: *International Journal for the Psychology of Religion* 20 (2), 85–96.
- Steger, Michael F./Frazier, P. 2005: Meaning in life. One link in the chain from religiousness to well-being. In: *Journal of Counseling Psychology* 52 (4), 574–582.
- Steger, Michael F./Pickering, Natalie K./Adams, Erica/Burnett, Jennifer/Shin, Joo Y./Dik, Bryan J./Stauner, Nick 2010: The Quest for Meaning. Religious Affiliation Differences in the Correlates of Religious Quest and Search for Meaning in Life. In: *Psychology of religion and Spirituality* 2 (4), 206–226.
- Suhail, Kausar/Chaudhry, Haroon R. 2004: Predictors of Subjective Well-Being in an Eastern Muslim Culture. In: *Journal of Social and Clinical Psychology* 23 (3), 359–376.
- Tae, Young S./Heitkemper, Margaret/Kim, Mi Y. 2012: A path analysis: a model of depression in Korean women with breast cancer-mediating effects of self-esteem and hope. In: *Oncology nursing forum* 39 (1), E49-57.
- Tiliouine, Habib 2012: Subjective Wellbeing, Psychological needs, Meaning in life, Religious Practice and Income in the Population of Algeria. In: *Journal of Social Research & Policy* 3 (2). In: <https://doaj.org/article/c19b118b0ff2464c98cda55011d67867>.
- Tomas, J. M./Sancho, P./Galiana, L./Oliver, A. 2016: A Double Test on the Importance of Spirituality, the "Forgotten Factor", in Successful Aging. In: *Social Indicators Research* 127 (3), 1377–1389.
- Vitale, Rachel A. 2015: Spirituality, Resilience and Social Support as Predictors of Life Satisfaction in young Adults with a History of Childhood Trauma. Dissertation. Kent State University College.
- Wachholtz, Amy/Rogoff, MaiLan 2013: The relationship between spirituality and burnout among medical students. In: *Journal of contemporary medical education* 1 (2), 83–91.
- Wahlen, Carina 2013: Intrinsic religiosity, spirituality, life-satisfaction and worries, what is the relationship between them? Bachelorarbeit. DBS School of Arts. Dublin.
- Walsh, James M./Ciarrocchi, Joseph W./Piedmont, Ralph L./Haskins, Deborah 2007: Spiritual Transcendence And Religious Practices In Recovery From Pathological Gambling. Reducing Pain Or Enhancing Quality Of Life? In: *Research in the Social Scientific Study of Religion* 18, 155–176.
- Wang, Zhizhong/Koenig, Harold G./Ma, Hui/Al Shohaib, Saad 2016: Religion, Purpose in Life, Social Support, and Psychological Distress in Chinese University Students. In: *Journal of religion and health* 55 (3), 1055–1064.
- Warren, Peter/van Eck, Kathryn/Townley, Greg/Kloos, Bret 2015: Relationships Among Religious Coping, Optimism, and Outcomes for Persons With Psychiatric Disabilities. In: *Psychology of religion and Spirituality* 7 (2), 91–99.

- Weinert-Moreno, Inez 2012: The Influence of Religiosity on Psychological Well-Being and Life Satisfaction in an Elderly Population. Dissertation.
- White, Brian/Driver, Simon/Warren, Ann M. 2010: Resilience and Indicators of Adjustment During Rehabilitation From a Spinal Cord Injury. In: *Rehabilitation Psychology* 55 (1), 23–32.
- Wilkins, Teresa A./Piedmont, Ralph L./Magyar-Russell, Gina M. 2012: Spirituality Or Religiousness. Which Serves As The Better Predictor Of Elements Of Mental Health? In: *Research in the Social Scientific Study of Religion* 23, 53–73.
- Wilt, Joshua A./Grubbs, Joshua B./Exline, Julie J./Pargament, Kenneth I. 2016: Personality, Religious and Spiritual Struggles, and Well-Being. In: *Psychology of religion and Spirituality* 8 (4), 341–351.
- Wnuk, Marcin/Marcinkowski, Jerzy T./Fobair, Patricia 2012: The Relationship of Purpose in Life and Hope in Shaping Happiness Among Patients with Cancer in Poland. In: *Journal of Psychosocial Oncology* 30 (4), 461–483.
- Wyshak, Grace 2016: Income and Subjective Well-Being. New Insights from Relatively Healthy American Women, Ages 49-79. In: *Plos one* 11 (2), 1-16.
- Yoon, Dong P./Lee, Eun-Kyoung O. 2006: The Impact of Religiousness, Spirituality, and Social Support on Psychological Well-Being Among Older Adults in Rural Areas. In: *Journal of Gerontological Social Work* 48 (3-4), 281–298.
- Zullig, Keith J./Ward, Rose M./Horn, Thelma 2006: The Association Between Perceived Spirituality, Religiosity, and Life Satisfaction. The Mediating Role of Self-Rated Health. In: *Social Indicators Research* 79 (2), 255–274.

## Anhang D: R-Session-Info

R version 3.5.1 (2018-07-02)  
Platform: x86\_64-w64-mingw32/x64 (64-bit)  
Running under: windows >= 8 x64 (build 9200)

Matrix products: default

locale:

[1] LC\_COLLATE=German\_Germany.1252 LC\_CTYPE=German\_Germany.1252 LC\_MONETARY=German\_Germany.1252  
[4] LC\_NUMERIC=C LC\_TIME=German\_Germany.1252

attached base packages:

[1] stats graphics grDevices utils datasets methods base

other attached packages:

[1] metafor\_2.0-0 Matrix\_1.2-14 texreg\_1.36.23 stargazer\_5.2.2 xtable\_1.8-3 semPlot\_1.1 foreign\_0.8-70  
[8] metaSEM\_1.2.0 openMX\_2.11.5

Loaded via a namespace (and not attached):

[1] nlme_3.1-137	RColorBrewer_1.1-2	mi_1.0	tools_3.5.1	backports_1.1.2
[6] R6_2.3.0	rpart_4.1-13	d3Network_0.5.2.1	Hmisc_4.1-1	lazyeval_0.2.1
[11] colorspace_1.3-2	nnet_7.3-12	tidyselect_0.2.5	gridExtra_2.3	mnormt_1.5-5
[16] compiler_3.5.1	qgraph_1.5	fdrtool_1.2.15	htmlTable_1.12	network_1.13.0.1
[21] scales_1.0.0	checkmate_1.8.5	mvtnorm_1.0-8	psych_1.8.10	pbapply_1.3-4
[26] sem_3.1-9	stringr_1.3.1	digest_0.6.18	pbivnorm_0.6.0	minqa_1.2.4
[31] base64enc_0.1-3	jpeg_0.1-8	pkgconfig_2.0.2	htmltools_0.3.6	lme4_1.1-19
[36] lisaRelTOR_0.1.4	htmlwidgets_1.3	rlang_0.3.0.1	rstudioapi_0.8	huge_1.2.7
[41] bindr_0.1.1	gttools_3.8.1	statnet.common_4.1.4	acepack_1.4.1	dpIyr_0.7.8
[46] zip_1.0.0	magrittr_1.5	Formula_1.2-3	Rcpp_1.0.0	munseIl_0.5.0
[51] prediction_0.3.6	abind_1.4-5	rockchalk_1.8.129	whisker_0.3-2	stringi_1.2.4
[56] yaml_2.2.0	carData_3.0-2	MASS_7.3-51.1	plyr_1.8.4	matrixcalc_1.0-3
[61] lavaan_0.6-3	grid_3.5.1	parallel_3.5.1	forcats_0.3.0	crayon_1.3.4
[66] lattice_0.20-38	kutils_1.45	splines_3.5.1	sna_2.4	knitr_1.20
[71] pillar_1.3.0	igraph_1.2.2	rjson_0.2.20	boot_1.3-20	corpcor_1.6.9
[76] BDgraph_2.53	reshape2_1.4.3	stats4_3.5.1	XML_3.98-1.16	glue_1.3.0
[81] latticeExtra_0.6-28	data.table_1.11.8	png_0.1-7	nloptr_1.2.1	tidyr_0.8.2
[86] gtable_0.2.0	purrr_0.2.5	assertthat_0.2.0	ggplot2_3.1.0	openxlsx_4.1.0
[91] broom_0.5.0	semTools_0.5-1	coda_0.19-2	survival_2.43-3	glasso_1.10
[96] tibble_1.4.2	arm_1.10-1	ggm_2.3	ellipse_0.4.1	bindrcpp_0.2.2
[101] cluster_2.0.7-1				

# Anhang E: Forest-Plots

Abbildung 46: Forest-Plot Zelle 1

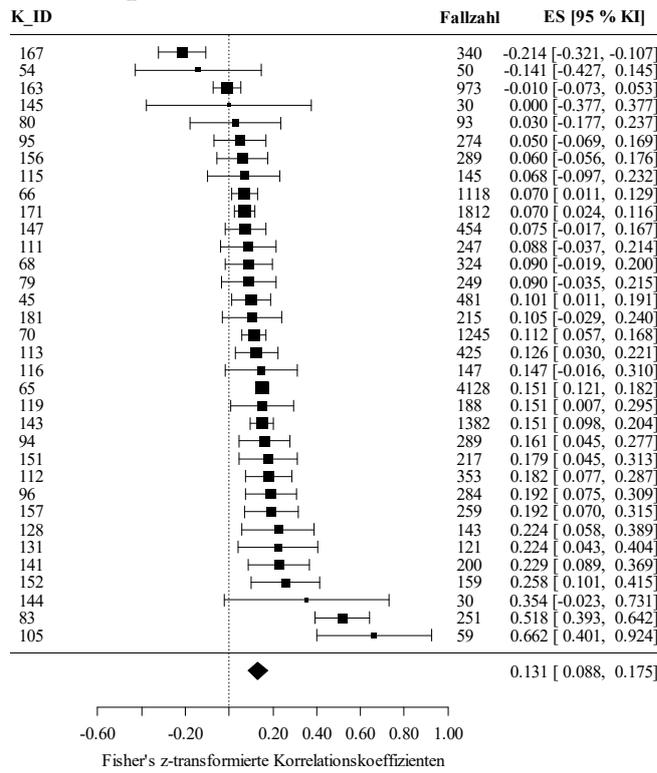


Abbildung 47: Forest-Plot Zelle 2

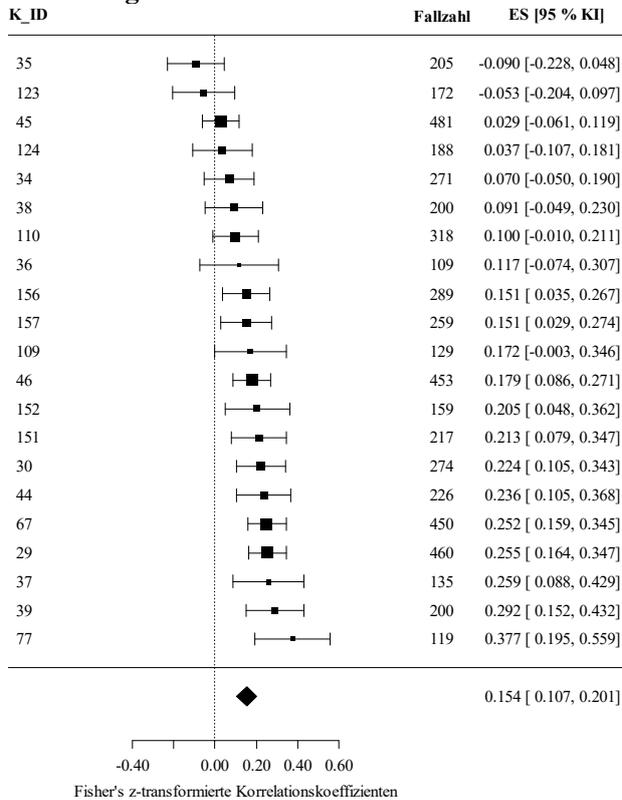


Abbildung 48: Forest-Plot Zelle 3

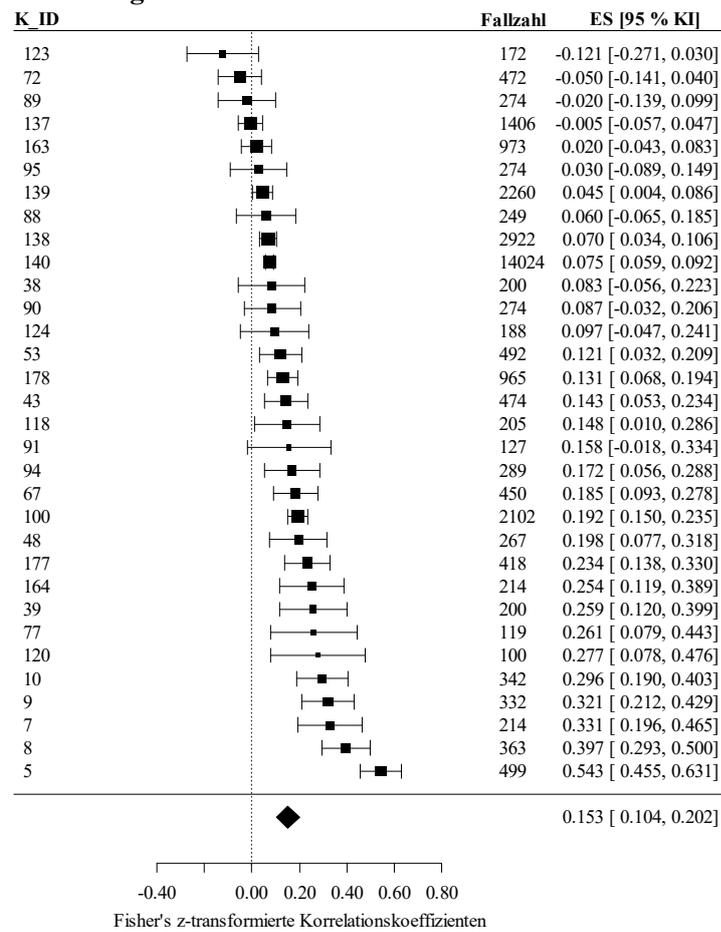
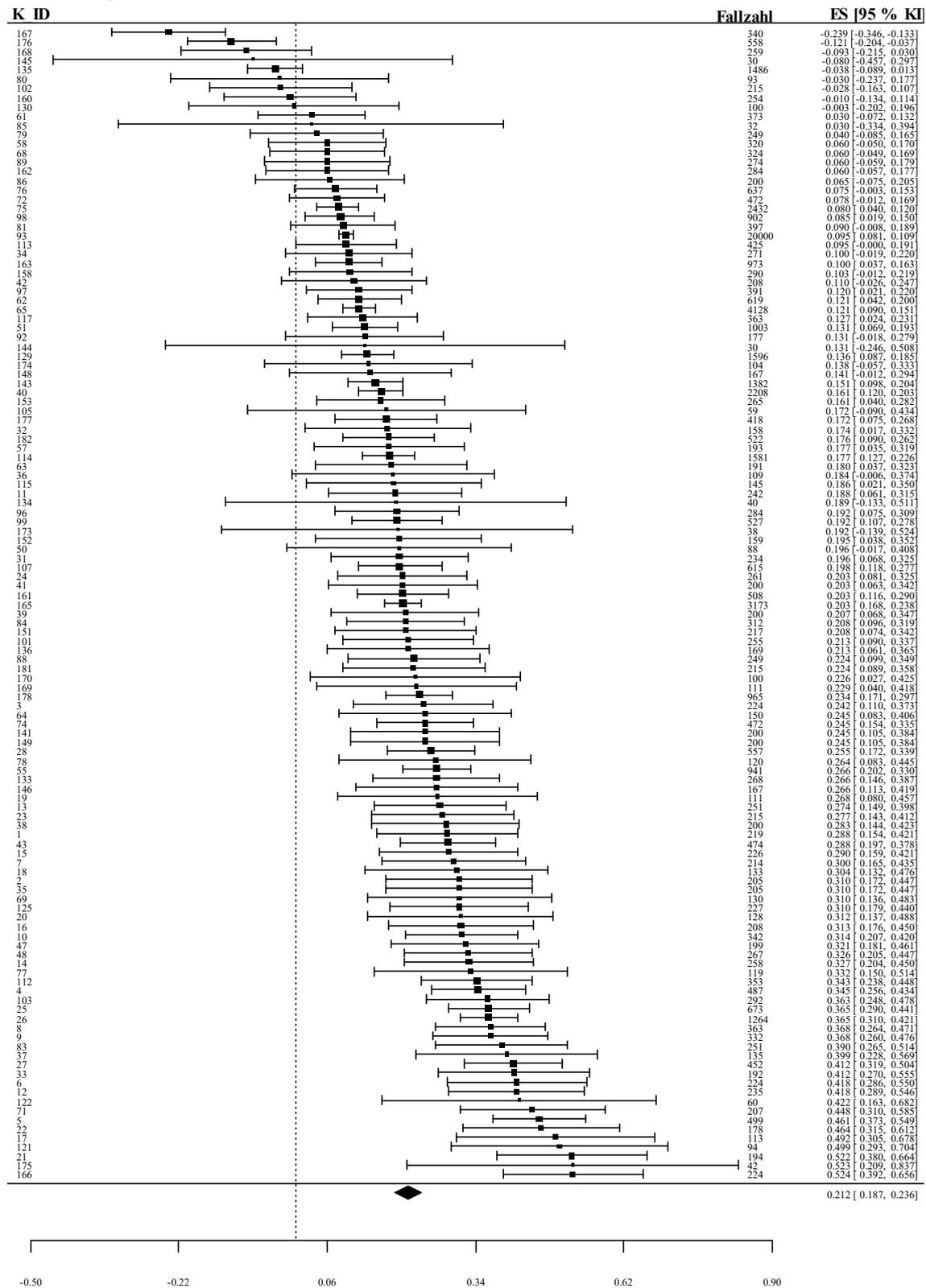
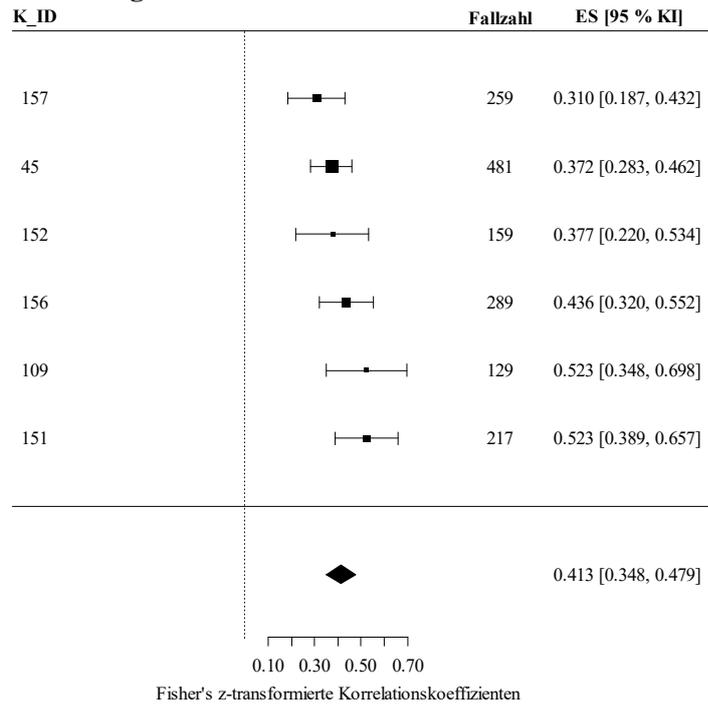


Abbildung 49: Forest-Plot Zelle 4



**Abbildung 50: Forest-Plot Zelle 5**



**Abbildung 51: Forest-Plot Zelle 6**

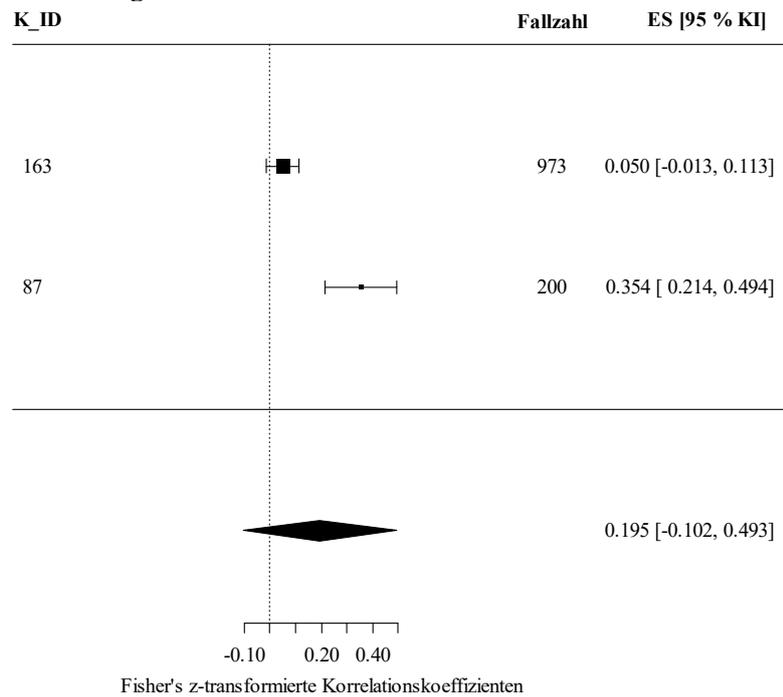


Abbildung 52: Forest-Plot Zelle 7

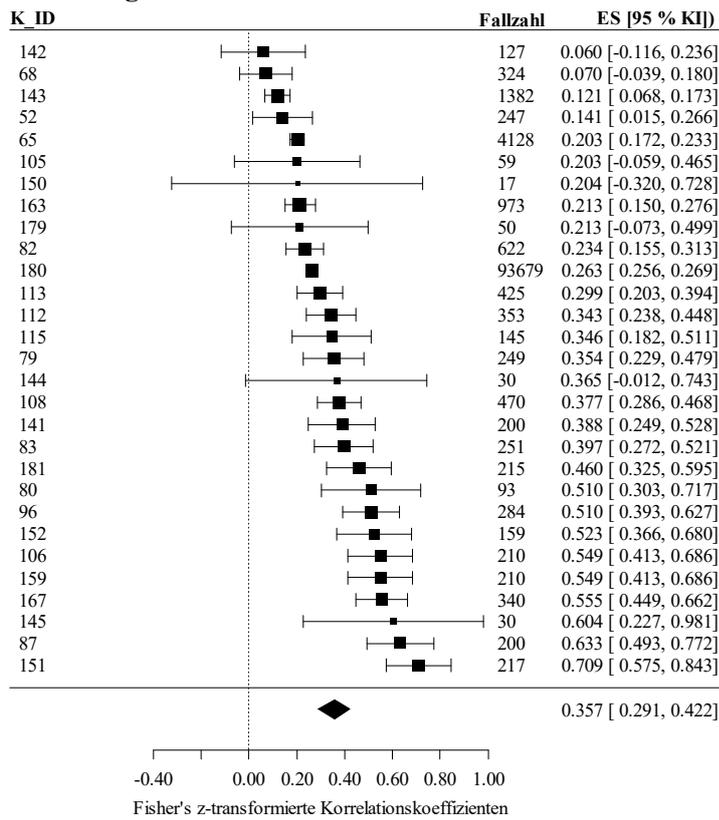


Abbildung 53: Forest-Plot Zelle 8

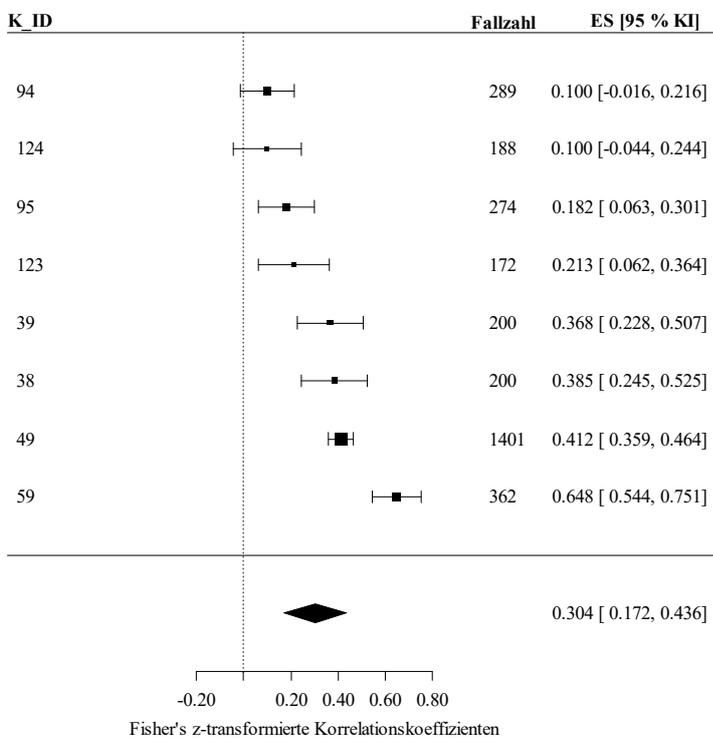


Abbildung 54: Forest-Plot Zelle 9

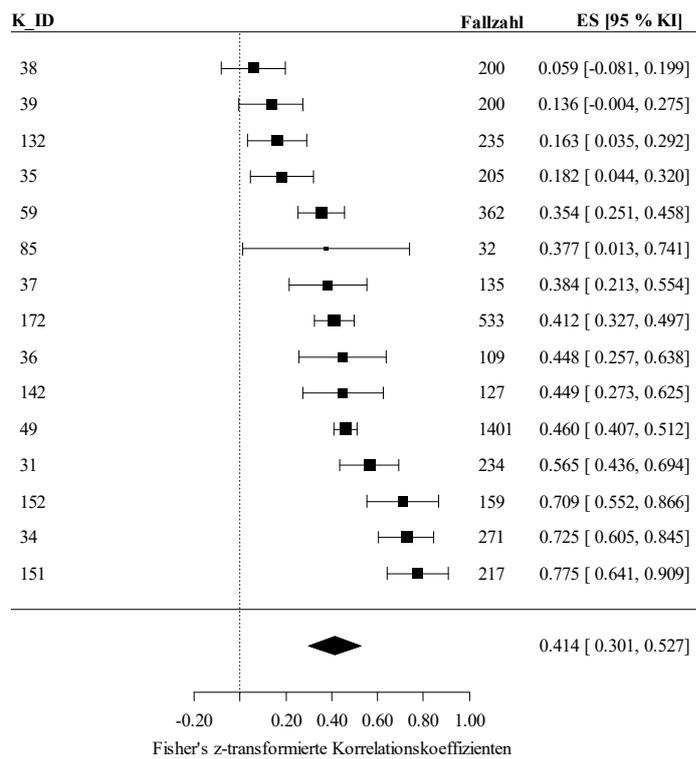
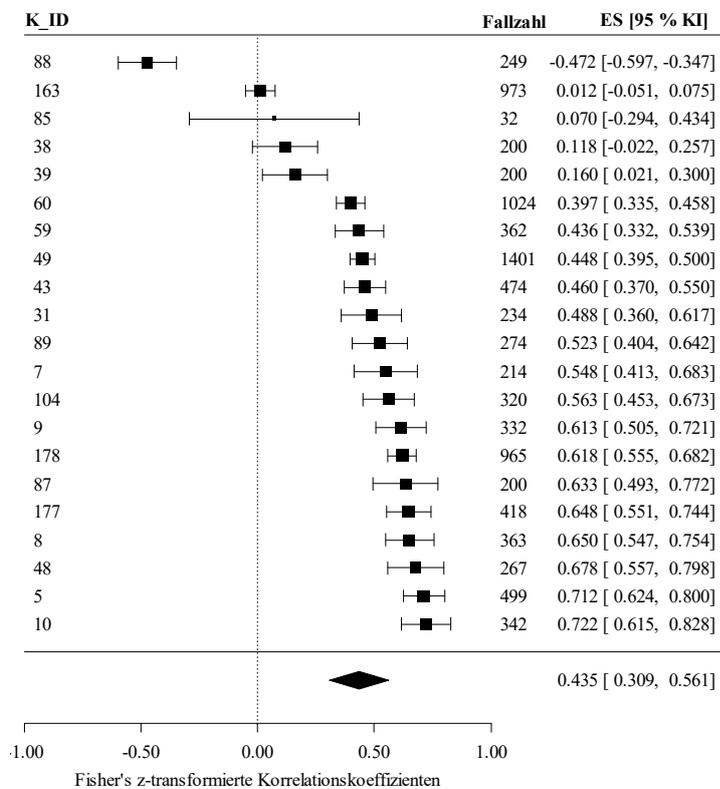


Abbildung 55: Forest-Plot Zelle 10



## Literaturverzeichnis

- Abdala, Gina Andrade/Kimura, Miako/Koenig, Harold G./Reinert, Katia G./Horton, Khim 2015: Religiosity and quality of life in older adults. Literature review. In: *Life Style* 2 (2), 25–51.
- Abdel-Khalek, Ahmed M./Lester, David 2017: The association between religiosity, generalized self-efficacy, mental health, and happiness in Arab college students. In: *Personality and Individual Differences* 109, 12–16.
- Abu-Raiya, Hisham/Agbaria, Qutaiba 2016: Religiousness and Subjective Well-Being Among Israeli-Palestinian College Students. Direct or Mediated Links? In: *Social indicators research* 126 (2), 829–844.
- Aghababaei, Naser 2013: God, the good life, and HEXACO. The relations among religion, subjective well-being and personality. In: *Mental Health, Religion & Culture* 17 (3), 284–290.
- Alamolhoda, Marzieh/Ayatollahi, Seyyed Mohammad Taghi/Bagheri, Zahra 2017: A comparative study of the impacts of unbalanced sample sizes on the four synthesized methods of meta-analytic structural equation modeling. In: *BMC research notes* 10 (1), 446.
- Albani, Cornelia/Bailer, Harald/Blaser, Gerd/Geyer, Michael/Brähler, Elmar/Grulke, Norbert 2002: Erfassung religiöser und spiritueller Einstellungen. Psychometrische Überprüfung der deutschen Version des "System of Belief Inventory" (SBI-15R-D) von Holland et al. in einer repräsentativen Bevölkerungsstichprobe. In: *PPMP - Psychotherapie - Psychosomatik - Medizinische Psychologie* (Januar 2002), 306–313.
- American Psychological Association 2008: Reporting standards for research in psychology. Why do we need them? What might they be? In: *The American psychologist* 63 (9), 839–851.
- Andrews, Frank M./Withey, Stephen B. 1976: *Social Indicators of Well-Being*. Boston, MA: Springer US.
- Angel, Hans-Ferdinand 2006: Die Entwicklung von Religiosität. Psychologische Perspektiven. In: Angel, Hans-Ferdinand (Hrsg.): *Religiosität. Anthropologische, theologische und sozialwissenschaftliche Klärungen. Praktische Theologie, Religionspädagogik, Diakonie*. Stuttgart: W. Kohlhammer, 133–159.
- Ano, Gene G./Vasconcelles, Erin B. 2005: Religious coping and psychological adjustment to stress: a meta-analysis. In: *Journal of clinical psychology* 61 (4), 461–480.
- Antonakis, John/Bendahan, Samuel/Jacquart, Philippe/Lalive, Rafael 2010: On making causal claims. A review and recommendations. In: *The Leadership Quarterly* 21 (6), 1086–1120.
- Anzures-Cabrera, Judith/Higgins, Julian P. T. 2010: Graphical displays for meta-analysis. An overview with suggestions for practice. In: *Research synthesis methods* 1 (1), 66–80.
- Arends, Lidia R. 2006: *Multivariate Meta-Analysis: modelling the Heterogeneity*. Dissertation. Erasmus University Rotterdam. Rotterdam.
- Argyle, Michael 2002: *The psychology of happiness*. 2. unveränderte Auflage. London: Routledge.
- Argyle, Michael 2003: Causes and Correlates of Happiness. In: Kahneman, Daniel (Hrsg.): *Well-Being. Foundations of Hedonic Psychology*. New York: Russell Sage Foundation.
- Bai, Mei/Lazenby, Mark 2015: A systematic review of associations between spiritual well-being and quality of life at the scale and factor levels in studies among patients with cancer. In: *Journal of palliative medicine* 18 (3), 286–298.
- Baier, Colin/Wright, Bradley 2001: "If You Love Me, Keep My Commandments". A Meta-Analysis of the Effect of Religion on Crime. In: *Journal of Research in Crime and Delinquency* 38 (1), 3–21.

- Baier, Karl 2006: Spiritualitätsforschung heute. In: Baier, Karl (Hrsg.): Handbuch Spiritualität. Zugänge. Traditionen. Interreligiöse Prozesse. Darmstadt: WBG - Wissenschaftliche Buchgesellschaft, 4–46.
- Bamberg, Sebastian/Möser, Guido 2007: Twenty years after Hines, Hungerford, and Tomera. A new meta-analysis of psycho-social determinants of pro-environmental behaviour. In: *Journal of Environmental Psychology* 27 (1), 14–25.
- Bartels, Meike 2015: Genetics of wellbeing and its components satisfaction with life, happiness, and quality of life: a review and meta-analysis of heritability studies. In: *Behavior genetics* 45 (2), 137–156.
- Barton, Yakov A./Miller, Lisa 2015: Spirituality and positive psychology go hand in hand: an investigation of multiple empirically derived profiles and related protective benefits. In: *Journal of religion and health* 54 (3), 829–843.
- Bauer, Daniel J. 2017: A more general model for testing measurement invariance and differential item functioning. In: *Psychological methods* 22 (3), 507–526.
- Becker, Betsy J. 1992: Using Results From Replicated Studies to Estimate Linear Models. In: *Journal of Educational and Behavioral Statistics* 17 (4), 341–362.
- Becker, Betsy J. 1995: Corrections to "Using Results From Replicated Studies to Estimate Linear Models". In: *Journal of Educational and Behavioral Statistics* 20 (1), 100–102.
- Becker, Betsy J. 2000: Multivariate Meta-analysis. In: Tinsley, Howard E./Brown, Steven D. (Hrsg.): *Handbook of applied multivariate statistics and mathematical modeling*. San Diego, Calif.: Academic Press, 499–526.
- Becker, Betsy J. 2001: Examining theoretical models through research synthesis. The benefits of model-driven meta-analysis. In: *Evaluation & the health professions* 24 (2), 190–217.
- Becker, Betsy J. 2009: Model-Based Meta-Analysis. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 377–395.
- Becker, Betsy J./Fahrbach, Kyle R. (1994), A comparison of approaches to the synthesis of correlation matrices., New Orleans. In:
- Becker, Betsy J./Schram, Christine M. 1994: Examining explanatory models through research synthesis. In: Cooper, Harris M. (Hrsg.): *The handbook of research synthesis*. New York, NY: Russell Sage Foundation, 357–381.
- Becker, Betsy J./Wu, Meng-Jia 2007: The Synthesis of Regression Slopes in Meta-Analysis. In: *Statistical Science* 22 (3), 414–429.
- Becker, Peter 1991: Theoretische Grundlagen. In: Abele, Andrea/Becker, Peter (Hrsg.): *Wohlbefinden. Theorie - Empirie - Diagnostik. Juventa-Materialien*, 2. Auflage. Weinheim: Juventa-Verlag, 13–49.
- Bergan, Anne/McConatha, Jasmin Tahmeseb 2001: Religiosity and Life Satisfaction. In: *Activities, Adaptation & Aging* 24 (3), 23–34.
- Bergh, Donald D./Aguinis, Herman/Heavey, Ciaran/Ketchen, David J./Boyd, Brian K./Su, Peiran/Lau, Cubie L. L./Joo, Harry 2014: Using meta-analytic structural equation modeling to advance strategic management research. Guidelines and an empirical illustration via the strategic leadership-performance relationship. In: *Strategic Management Journal* 37 (3), 477–497.
- Bergin, Allen E. 1983: Religiosity and mental health. A critical reevaluation and meta-analysis. In: *Professional Psychology: Research and Practice* 14 (2), 170–184.
- Bergin, Allen E. 1991: Values and religious issues in psychotherapy and mental health. In: *American Psychologist* 46 (4), 394–403.

- Berth, Hendrik/Klein, Constantin/Balck, Friedrich 2011: Begriffsbestimmungen. Gesundheit, Krankheit, Wohlbefinden und Lebensqualität. In: Klein, Constantin/Berth, Hendrik/Balck, Friedrich (Hrsg.): *Gesundheit - Religion - Spiritualität. Konzepte, Befunde und Erklärungsansätze. Gesundheitsforschung*. Weinheim: Juventa-Verlag, 47–64.
- Bickel, Robert (Hrsg.), 2007: *Multilevel analysis for applied research*. New York: Guilford Press.
- Boer, Diana/Deinert, Anika/Homan, Astrid C./Voelpel, Sven C. 2016: Revisiting the mediating role of leader–member exchange in transformational leadership. The differential impact model. In: *European Journal of Work and Organizational Psychology* 25 (6), 883–899.
- Borenstein, Michael/Hedges, Larry V./Higgins, Julian P. T./Rothstein, Hannah R. 2011: *Introduction to meta-analysis*. Reprinted. Chichester: Wiley.
- Borenstein, Michael/Higgins, Julian P. T./Hedges, Larry V./Rothstein, Hannah R. 2017: Basics of meta-analysis. I<sup>2</sup> is not an absolute measure of heterogeneity. In: *Research synthesis methods* 8 (1), 5–18.
- Borman, Geoffrey D./Grigg, Jeffrey A. 2009: Visual and Narrative Interpretation. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 497–520.
- Bosnjak, Michael/Viechtbauer, Wolfgang 2009: Die Methode der Meta-Analyse zur Evidenz-basierung von Gesundheitsrisiken. Beiträge der Sozial-, Verhaltens- und Wirtschaftswissenschaften. In: *Zentralblatt für Arbeitsmedizin, Arbeitsschutz und Ergonomie* 59 (11), 322–333.
- Brown, Steven P./Peterson, Robert A. 1993: Antecedents and Consequences of Salesperson Job Satisfaction. Meta-Analysis and Assessment of Causal Effects. In: *Journal of Marketing Research* 30 (1), 63.
- Browne, Michael W. 1984: Asymptotically distribution-free methods for the analysis of covariance structures. In: *The British journal of mathematical and statistical psychology* 37 (Pt 1), 62–83.
- Brüderl, Josef 2004: Meta-Analyse in der Soziologie. Bilanz der deutschen Scheidungsforschung oder „statistischer Fruchtsalat“? / Meta-Analysis in Sociology: Review of German Research on Divorce or “Statistical Fruit Salad”? In: *Zeitschrift für Soziologie* 33 (1).
- Bruni, Luigino/Porta, Pier Luigi 2014: Happiness and Quality of Life Reconciled. In: Bruni, Luigino/Porta, Pier L. (Hrsg.): *Handbook of Research Methods and Applications in Happiness and Quality of Life*. Cheltenham, Gloucestershire: Edward Elgar Publishing.
- Bucher, Anton A. 2014: *Psychologie der Spiritualität*. 2., vollständig überarbeitete Auflage. Weinheim: Beltz-Juventa.
- Bushman, Brad J./Wang, Morgan C. 2009: Vote-Counting Procedures in Meta-Analysis. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 208–220.
- Byrne, Barbara M. 2006: *Structural equation modeling with EQS*. 2nd ed. Mahwah, N.J: Lawrence Erlbaum Associates.
- Capaldi, Colin A./Dopko, Raelyne L./Zelenski, John M. 2014: The relationship between nature connectedness and happiness: a meta-analysis. In: *Frontiers in psychology* 5, 976.
- Card, Noel A. 2012: *Applied meta-analysis for social science research*. New York, NY: Guilford Press.
- Card, Noel A. 2017: Advances in meta-analysis methodologies contribute to advances in research accumulation. Comments on Cheung & Hong and Johnson et al. In: *Health psychology review* 11 (3), 302–305.
- Casas, Ferran/Sarriera, Jorge C./Alfaro, Jaime/González, Mònica/Bedin, Lívia/Abs, Daniel/Figuer, Cristina/Valdenegro, Boris 2015: Reconsidering Life Domains that Contribute to

- Subjective Well-Being Among Adolescents with Data from Three Countries. In: *Journal of Happiness Studies* 16 (2), 491–513.
- Charlemagne-Badal, Sherma J./Lee, Jerry W./Butler, Terry L./Fraser, Gary E. 2015: Conceptual Domains Included in Wellbeing and Life Satisfaction Instruments. A Review. In: *Applied Research in Quality of Life* 10 (2), 305–328.
- Cheung, Mike W.-L. 2008: A model for integrating fixed-, random-, and mixed-effects meta-analyses into structural equation modeling. In: *Psychological methods* 13 (3), 182–202.
- Cheung, Mike W.-L. 2009: Constructing Approximate Confidence Intervals for Parameters With Structural Equation Models. In: *Structural Equation Modeling: A Multidisciplinary Journal* 16 (2), 267–294.
- Cheung, Mike W.-L. 2014a: Fixed- and random-effects meta-analytic structural equation modeling: examples and analyses in R. In: *Behavior research methods* 46 (1), 29–40.
- Cheung, Mike W.-L. 2014b: metaSEM: an R package for meta-analysis using structural equation modeling. In: *Frontiers in psychology* 5, 1–7.
- Cheung, Mike W.-L. 2015: *Meta-analysis*. Chichester, West Sussex: Wiley.
- Cheung, Mike W.-L. 2017a: Issues in Solving the Problem of Effect Size Heterogeneity in Meta-Analytic Structural Equation Modeling. In:
- Cheung, Mike W.-L. 2017b: An R Package for Meta-Analysis using Structural Equation Modeling. Department of Psychology, National University of Singapore. In:
- Cheung, Mike W.-L. 2018: Some reflections on combining meta-analysis and structural equation modeling. In: *Research synthesis methods*, 1–8.
- Cheung, Mike W.-L./Chan, Wai 2005a: Classifying Correlation Matrices Into Relatively Homogeneous Subgroups. A Cluster Analytic Approach. In: *Educational and Psychological Measurement* 65 (6), 954–979.
- Cheung, Mike W.-L./Chan, Wai 2005b: Meta-analytic structural equation modeling: a two-stage approach. In: *Psychological methods* 10 (1), 40–64.
- Cheung, Mike W.-L./Chan, Wai 2009: A Two-Stage Approach to Synthesizing Covariance Matrices in Meta-Analytic Structural Equation Modeling. In: *Structural Equation Modeling: A Multidisciplinary Journal* 16 (1), 28–53.
- Cheung, Mike W.-L./Cheung, Shu-Fai 2016: Random-effects models for meta-analytic structural equation modeling: review, issues, and illustrations. In: *Research synthesis methods* 7 (2), 140–155.
- Cheung, Mike W.-L./Hafdahl, Adam R. 2016: Special issue on meta-analytic structural equation modeling: introduction from the guest editors. In: *Research synthesis methods* 7 (2), 112–120.
- Cheung, Mike W.-L./Ho, Roger C. M./Lim, Yonghao/Mak, Anselm 2012: Conducting a meta-analysis: basics and good practices. In: *International journal of rheumatic diseases* 15 (2), 129–135.
- Cheung, Mike W.-L./Hong, Ryan Y. 2017: Applications of meta-analytic structural equation modelling in health psychology. Examples, issues, and recommendations. In: *Health psychology review* 11 (3), 265–279.
- Cheung, Mike W.-L./Vijayakumar, Ranjith 2016: A Guide to Conducting a Meta-Analysis. In: *Neuropsychology review* 26 (2), 121–128.
- Cheung, Shu-Fai 2001: Examining solutions to two practical issues in meta-analysis: Dependent correlations and missing data in correlation matrices. Dissertation. Chinese University. Hong Kong.
- Cheung, Shu-Fai/Sun, Rong Wei/Chan, Darius K.-S. 2018: Correlation-Based Meta-Analytic Structural Equation Modeling. In: *Organizational Research Methods* 20, 1-25.
- Chida, Yoichi/Step toe, Andrew/Powell, Lynda H. 2009: Religiosity/spirituality and mortality. A systematic quantitative review. In: *Psychotherapy and psychosomatics* 78 (2), 81–90.

- Chiu, Lyren/Emblen, Julia D./van Hofwegen, Lynn/Sawatzky, Richard/Meyerhoff, Heather 2004: An integrative review of the concept of spirituality in the health sciences. In: *Western journal of nursing research* 26 (4), 405–428.
- Cho, Kyunghwa 2015: *Meta-Analysis of Factor-Analysis*. Dissertation. Florida State University. Tallahassee.
- Clarke, Mike 2009: Reporting Format. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 521–536.
- Cohen, Adam B. 2002: The Importance of Spirituality in Well-Being for Jews and Christians. In: *Journal of Happiness Studies* 3 (3), 287–310.
- Cohen, Adam B./Johnson, Kathryn A. 2011: Religion and Well-being: A Psychological Perspective. In:
- Cohen, Adam B./Johnson, Kathryn A. 2017: The Relation between Religion and Well-Being. In: *Applied Research in Quality of Life* 12 (3), 533–547.
- Cohen, Adam B./Wu, M. S./Miller, J. 2016: Religion and Culture. Individualism and Collectivism in the East and West. In: *Journal of Cross-Cultural Psychology* 47 (9), 1236–1249.
- Cohen, Carl I./Vengassery, Aninditha/Garcia Aracena, Elena F. 2017: A Longitudinal Analysis of Quality of Life and Associated Factors in Older Adults with Schizophrenia Spectrum Disorder. In: *The American journal of geriatric psychiatry : official journal of the American Association for Geriatric Psychiatry* 25 (7), 755–765.
- Compton, W. C. 2001: Toward a tripartite factor structure of mental health: subjective well-being, personal growth, and religiosity. In: *The Journal of psychology* 135 (5), 486–500.
- Cooper, Harris M. 2009: Hypothesis and Problems in Research Synthesis. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 19–35.
- Cooper, Harris M. 2010: *Research synthesis and meta-analysis*. 4. ed. Los Angeles, Calif.: Sage.
- Cooper, Harris M./Hedges, Larry V. 2009a: Potentials and Limitations. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 561–572.
- Cooper, Harris M./Hedges, Larry V. 2009b: Research Synthesis as a scientific process. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 3–16.
- Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.), 2009: *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation.
- Crespo, Ricardo F./Mesurado, Belén 2015: Happiness Economics, Eudaimonia and Positive Psychology. From Happiness Economics to Flourishing Economics. In: *Journal of Happiness Studies* 16 (4), 931–946.
- Crivelli, Luca/Della Nella, Sara/Lucchini, Mario 2014: Happiness and health. In: Bruni, Luigino/Porta, Pier L. (Hrsg.): *Handbook of Research Methods and Applications in Happiness and Quality of Life*. Cheltenham, Gloucestershire: Edward Elgar Publishing.
- Cummins, Robert A. 1996: The domains of life satisfaction. An attempt to order chaos. In: *Social indicators research* 38 (3), 303–328.
- Curtiss, Joshua/Andrews, Leigh/Davis, Michelle/Smits, Jasper/Hofmann, Stefan G. 2017: A meta-analysis of pharmacotherapy for social anxiety disorder. An examination of efficacy, moderators, and mediators. In: *Expert opinion on pharmacotherapy* 18 (3), 243–251.
- Daniel Lüdecke 2018: *sjPlot - Data Visualization for Statistics in Social Science [Computer software]*: Zenodo. In:
- Dehaan, Laura G./Yonker, Julie E./Affholter, Carolyn 2011: More than enjoying the sunset: conceptualization and measurement of religiosity for adolescents and emerging adults

- and its implications for developmental inquiry. In: *Journal of Psychology and Christianity* 30 (3).
- DeNeve, K. M./Cooper, Harris M. 1998: The happy personality: a meta-analysis of 137 personality traits and subjective well-being. In: *Psychological bulletin* 124 (2), 197–229.
- Detle, E. Dorothea 2005: *Berufserfolg und Lebenszufriedenheit - Eine längsschnittliche Analyse der Zusammenhänge*. Dissertation. Friedrich-Alexander-Universität. Erlangen-Nürnberg.
- Dezutter, Jessie/Soenens, Bart/Hutsebaut, Dirk 2006: Religiosity and mental health. A further exploration of the relative importance of religious behaviors vs. religious attitudes. In: *Personality and Individual Differences* 40 (4), 807–818.
- Diener, Ed 1984: Subjective well-being. In: *Psychological bulletin* 95 (3), 542–575.
- Diener, Ed 2006: Guidelines for National Indicators of Subjective Well-Being and Ill-Being. In: *Journal of Happiness Studies* 7 (4), 397–404.
- Diener, Ed 2008: Myths in the Science of Happiness, and Directions for Future Research. In: Eid, Michael (Hrsg.): *The science of subjective well-being*. New York NY: The Guilford Press, 493–515.
- Diener, Ed/Clifton, Don 2002: Life Satisfaction and Religiosity in Broad Probability Samples. In: *Psychology Inquiry* 13 (3), 206–209.
- Diener, Ed/Heintzelman, Samantha J./Kushlev, Kostadin/Tay, Louis/Wirtz, Derrick/Lutes, Lesley D./Oishi, Shigehiro 2017: Findings all psychologists should know from the new science on subjective well-being. In: *Canadian Psychology/Psychologie canadienne* 58 (2), 87–104.
- Diener, Ed/Lucas, Richard E. 2003: Personality and Subjective Well-Being. In: Kahneman, Daniel (Hrsg.): *Well-Being. Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, 213–229.
- Diener, Ed/Lucas, Richard E./Schimmack, Ulrich/Helliwell, John 2009: *Well-Being for Public Policy*. Oxford: Oxford University Press.
- Diener, Ed/Oishi, Shigehiro/Lucas, Richard E. 2011: Subjective Well-Being: The Science of Happiness and Life Satisfaction. In: Lopez, Shane J./Snyder, C. R. (Hrsg.): *Handbook of positive psychology*. Oxford library of psychology, 2. Auflage. Oxford: Oxford University Press, 187–194.
- Diener, Ed/Ryan, Katherine 2009: Subjective well-being. A General Overview. In: *South African Journal of Psychology* 39 (4), 391–406.
- Diener, Ed/Seligman, Martin E. P. 2004: Beyond Money: Toward an Economy of Well-Being. In: *Psychological science in the public interest : a journal of the American Psychological Society* 5 (1), 1–31.
- Diener, Ed/Suh, Eunkook M./Lucas, Richard E./Smith, Heidi L. 1999: Subjective well-being. Three decades of progress. In: *Psychological bulletin* 125 (2), 276–302.
- Diener, Ed/Tay, Louis 2014: New Frontiers: Societal Measures of Subjective Well-Being for Input to Policy. In: Bruni, Luigino/Porta, Pier L. (Hrsg.): *Handbook of Research Methods and Applications in Happiness and Quality of Life*. Cheltenham, Gloucestershire: Edward Elgar Publishing, 35–52.
- Diener, Ed/Tay, Louis/Myers, David G. 2011: The religion paradox: if religion makes people happy, why are so many dropping out? In: *Journal of personality and social psychology* 101 (6), 1278–1290.
- Doane, Michael J. 2013: The association between religiosity and subjective well-being. The unique contribution of religious service attendance and the mediating role of perceived religious social support. In: *The Irish Journal of Psychology* 34 (1), 49–66.

- Doane, Michael J./Elliott, Marta 2016: Religiosity and Self-Rated Health: A Longitudinal Examination of Their Reciprocal Effects. In: *Journal of religion and health* 55 (3), 844–855.
- Dolan, Paul/Peasgood, Tessa/White, Mathew 2008: Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. In: *Journal of Economic Psychology* 29 (1), 94–122.
- Donahue, Michael J. 1985: Intrinsic and extrinsic religiousness. Review and meta-analysis. In: *Journal of personality and social psychology* 48 (2), 400–419.
- Donahue, Michael J./Benson, Peter L. 1995: Religion and the Well-Being of Adolescents. In: *Journal of Social Issues* 51 (2), 145–160.
- Dörr, Anette 2001: *Religiosität und psychische Gesundheit*. Hamburg: Kováč.
- Duval, Sue 2005: The Trim and Fill Method. In: Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael (Hrsg.): *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester, West Sussex: Wiley, 127–144.
- Eger, Robert J./Maridal, J. Haavard 2015: A statistical meta-analysis of the wellbeing literature. In: *International Journal of Wellbeing* 5 (2), 45–74.
- Eichhorn, J. 2012: Happiness for Believers? Contextualizing the Effects of Religiosity on Life-Satisfaction. In: *European Sociological Review* 28 (5), 583–593.
- Eisend, Martin 2014: *Metaanalyse*. 1. Auflage. München: Rainer Hampp Verlag.
- Eisend, Martin 2017: *Meta-Analysis in Advertising Research*. In: *Journal of Advertising*, 1–15.
- Eller, Tönnes Freerk 2015: *Hält die Service-Profit Chain was sie verspricht? Eine metaanalytische Betrachtung*. Dissertation. Katholischen Universität Eichstätt-Ingolstadt. Eichstätt-Ingolstadt.
- Elliott, Marta/Hayward, David 2009: Religion and Life Satisfaction Worldwide. The Role of Government Regulation. In: *Sociology of Religion* 70 (3), 285–310.
- Ellison, Christopher G. 1991: Religious involvement and subjective well-being. In: *Journal of health and social behavior* 32 (1), 80–99.
- Ellison, Christopher G./Gay, David A./Glass, Thomas A. 1989: Does Religious Commitment Contribute to Individual Life Satisfaction? In: *Social Forces* 68 (1), 100–123.
- Ellison, Christopher G./Levin, Jeffrey S. 1998: The religion-health connection: evidence, theory, and future directions. In: *Health education & behavior : the official publication of the Society for Public Health Education* 25 (6), 700–720.
- Emmons, Robert A./Cheung, Chi/Tehrani, Keivan 1998: Assessing Spirituality Through Personal Goals: Implications for Research on Religion and Subjective Well-Being. In: *Social indicators research* 45 (1/3), 391–422.
- Enders, Craig K. 2010: *Applied missing data analysis*. New York: Guilford Press.
- Epskamp, Sacha 2015: semPlot: Unified Visualizations of Structural Equation Models. In: *Structural Equation Modeling: A Multidisciplinary Journal* 22 (3), 474–483.
- Fernández, Belén/Jamshidi, Laleh/Declercq, Lies/Beretvas, S. Natasha/Onghena, Patrick/van den Noortgate, Wim 2019: The Application of Meta-Analytic Models with Multiple Random Effects: A Systematic Review. In: *PsyArXiv Preprint*.
- Ferriss, Abbott L. 2002: Religion and the Quality of Life. In: *Journal of Happiness Studies* 3 (3), 199–215.
- Fiebig, Joachim/Rüdebusch, Gerrit/Urban, Dieter 2012: Meta-Analyse und "publication bias". In: *SISS: Schriftenreihe des Instituts für Sozialwissenschaften an der Universität Stuttgart* (Nr. 1).
- Fiebig, Joachim/Urban, Dieter 2014: Meta-Analyse mit Mehrebenenmodellen unter Verwendung von HLM. In: *SISS: Schriftenreihe des Instituts für Sozialwissenschaften an der Universität Stuttgart* (Nr. 1).

- Fischer, Ronald/Boer, Diana 2011: What is more important for national well-being: money or autonomy? A meta-analysis of well-being, burnout, and anxiety across 63 societies. In: *Journal of personality and social psychology* 101 (1), 164–184.
- Francis, Leslie J./Katz, Yaacov J./Yablon, Yaacov/Robbins, Mandy 2004: Religiosity, Personality, and Happiness. A Study among Israeli Male Undergraduates. In: *Journal of Happiness Studies* 5 (4), 315–333.
- Frederick, Shane/Loewenstein, George 2003: Hedonic Adaptation. In: Kahneman, Daniel (Hrsg.): *Well-Being. Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, 302–329.
- Fredrickson, Barbara L. 2004: The broaden-and-build theory of positive emotions. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 359 (1449), 1367–1378.
- Frey, Bruce B./Daaleman, Timothy P./Peyton, Vicki 2005: Measuring a Dimension of Spirituality for Health Research. Validity of the Spirituality Index of Well-Being. In: *Research on Aging* 27 (5), 556–577.
- Frick, Eckhard/Büssing, Arndt/Baumann, Klaus/Weig, Wolfgang/Jacobs, Christoph 2016: Do Self-efficacy Expectation and Spirituality Provide a Buffer Against Stress-Associated Impairment of Health? A Comprehensive Analysis of the German Pastoral Ministry Study. In: *Journal of religion and health* 55 (2), 448–468.
- Furlow, Carolyn F./Beretvas, Natasha S. 2005: Meta-analytic methods of pooling correlation matrices for structural equation modeling under different patterns of missing data. In: *Psychological methods* 10 (2), 227–254.
- Furlow, Carolyn F./Beretvas, Natasha S. 2010: An Evaluation of Multiple Imputation for Meta-Analytic Structural Equation Modelling. In: *Journal of Modern Applied Statistical Methods* 9 (1), 129–143.
- Gabriel, Karl/Reuter, Hans-Richard/Große Kracht, Hermann-Josef 2010: *Religion und Gesellschaft*. 2. durchgesehene Auflage. Paderborn: Schöningh.
- Gabriel, Oscar W. 2012: Zufriedenheit mit den individuellen und kollektiven Lebensbedingungen im vereinigten Deutschland. In: Keil, Silke I./van Deth, Jan W. (Hrsg.): *Deutschlands Metamorphosen. Ergebnisse des European Social Survey 2002 bis 2008. Studien zur Wahl- und Einstellungsforschung*, 1. Auflage. Baden-Baden: Nomos, 99–138.
- Garssen, Bert/Uwland-Sikkema, Noline F./Visser, Anja 2015: How Spirituality Helps Cancer Patients with the Adjustment to their Disease. In: *Journal of religion and health* 54 (4), 1249–1265.
- Garssen, Bert/Visser, Anja/Jager Meezenbroek, Eltica de 2016: Examining whether spirituality predicts subjective well-being. How to avoid tautology. 8 (2), 141–148.
- Gartner, John/Larson, David B./Allen, George D. 1991: Religious commitment and mental health: A review of the empirical literature. In: *Journal of Psychology and Theology* 19 (1), 6–25.
- George, Linda K. 2010: Still happy after all these years: research frontiers on subjective well-being in later life. In: *The journals of gerontology. Series B, Psychological sciences and social sciences* 65B (3), 331–339.
- George, Linda K./Ellison, Christopher G./Larson, David B. 2002: Explaining the Relationships Between Religious Involvement and Health. In: *Psychological Inquiry* 13 (3), 190–200.
- George, Linda K./Larson, David B./Koenig, Harold G./McCullough, Michael E. 2000: Spirituality and Health. What We Know, What We Need to Know. In: *Journal of Social and Clinical Psychology* 19 (1), 102–116.
- Gignac, Gilles E./Szodorai, Eva T. 2016: Effect size guidelines for individual differences researchers. In: *Personality and Individual Differences* 102, 74–78.

- Gilboa, Itzhak/Schmeidler, David 2001: A cognitive model of individual well-being. In: *Social Choice and Welfare* 18 (2), 269–288.
- Glaesmer, Heide/Grande, Gesine/Braehler, Elmar/Roth, Marcus 2011: The German Version of the Satisfaction With Life Scale (SWLS). In: *European Journal of Psychological Assessment* 27 (2), 127–132.
- Glass, Gene V. 1976: Primary, Secondary, and Meta-Analysis of Research. In: *Educational Researcher* 5 (10), 3–8.
- Gleser, Leon/Olkin, Ingram 2009: Stochastically Dependent Effect Sizes. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 358–376.
- González, Mònica/Coenders, Germà/Saez, Marc/Casas, Ferran 2010: Non-linearity, Complexity and Limited Measurement in the Relationship Between Satisfaction with Specific Life Domains and Satisfaction with Life as a Whole. In: *Journal of Happiness Studies* 11 (3), 335–352.
- Goyal, Madhav/Singh, Sonal/Sibinga, Erica M. S./Gould, Neda F./Rowland-Seymour, Anastasia/Sharma, Ritu/Berger, Zackary/Sleicher, Dana/Maron, David D./Shihab, Hasan M./Ranasinghe, Padmini D./Linn, Shauna/Saha, Shonali/Bass, Eric B./Haythornthwaite, Jennifer A. 2014: Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. In: *JAMA internal medicine* 174 (3), 357–368.
- Green, Morgan/Elliott, Marta 2010: Religion, health, and psychological well-being. In: *Journal of religion and health* 49 (2), 149–163.
- Greenhouse, Joel B./Iyengar, Satish 2009: Sensitivity Analysis and Diagnostics. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 417–433.
- Greiffenhagen, Martin/Greiffenhagen, Sylvia 1988: *Das Glück*. München: Piper.
- Grinde, Bjørn 2016: Why Negative Feelings are Important when Assessing Well-Being. In: *Journal of Happiness Studies* 17 (4), 1741–1752.
- Gu, Jenny/Strauss, Clara/Bond, Rod/Cavanagh, Kate 2015: How do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and well-being? A systematic review and meta-analysis of mediation studies. In: *Clinical psychology review* 37, 1–12.
- Guerini, Michaela/Nuvolati, Giampaolo 2014: Happiness, Subjective and Objective Indicators. In: Bruni, Luigino/Porta, Pier L. (Hrsg.): *Handbook of Research Methods and Applications in Happiness and Quality of Life*. Cheltenham, Gloucestershire: Edward Elgar Publishing, 400–418.
- Hackett, Conrad/Marshall, Joey/Schiller, Anna 2019: Religion's Relationship to Happiness, Civic Engagement and Health Around the World.
- Hackney, Charles H./Sanders, Glenn S. 2003: Religiosity and Mental Health. A Meta-Analysis of Recent Studies. In: *Journal for the Scientific Study of Religion* 42 (1), 43–55.
- Hagger, Martin S./Polet, Juho/Lintunen, Taru 2018: The reasoned action approach applied to health behavior: Role of past behavior and tests of some key moderators using meta-analytic structural equation modeling. In: *Social Science & Medicine* 213, 85–94.
- Hammerstrøm, Karianne/Wade, Anne/Klint Jørgensen/Anne-Marie 2016: Searching for studies: Guidelines on information retrieval for Campbell Systematic Reviews. In: *Campbell Methods Guides* (Nr. 1). In:
- Handal, Paul J./Fenzel, Mickey 2000: Personal spirituality as a moderator between stressors and subjective wellbeing. In: *Journal of Psychology and Theology* 28 (3), 221–228.

- Haybron, Daniel 2008: Philosophy and the Science of Subjective Well-Being. In: Eid, Michael (Hrsg.): The science of subjective well-being. New York NY: The Guilford Press, 17–43.
- Hayward, David/Krause, Neal 2014: Religion, mental health, and well-being: Social aspects. In: Saroglou, Vassilis (Hrsg.): Religion, personality, and social behavior. New York: Psychology Press, 255–280.
- Headey, Bruce/Hoehne, Gerhard/Wagner, Gert G. 2014: Does Religion Make You Healthier and Longer Lived? Evidence for Germany. In: Social indicators research 119 (3), 1335–1361.
- Headey, Bruce/Schupp, Juergen/Tucci, Ingrid/Wagner, Gert G. 2010: Authentic happiness theory supported by impact of religion on life satisfaction. A longitudinal analysis with data for Germany. In: The Journal of Positive Psychology 5 (1), 73–82.
- Hedges, Larry V. 2009: Statistical Considerations. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): The Handbook of Research Synthesis and Meta-Analysis. New York: Russell Sage Foundation, 37–47.
- Hedges, Larry V. 2016: Applying meta-analysis to structural equation modeling. In: Research synthesis methods 7 (2), 209–214.
- Hedges, Larry V./Olkin, Ingram 1985: Statistical Method for Meta-Analysis. Orlando: Academic Press.
- Hedges, Larry V./Tipton, Elizabeth/Johnson, Matthew C. 2010: Robust variance estimation in meta-regression with dependent effect size estimates. In: Research synthesis methods 1 (1), 39–65.
- Hedges, Larry V./Vevea, Jack L. 1998: Fixed- and random-effects models in meta-analysis. In: Psychological methods 3 (4), 486–504.
- Heidl, Christian/Landenberger, Margarete/Jahn/Patrick 2012: Lebenszufriedenheit in Westdeutschland - Eine Querschnittsanalyse mit den Daten des Sozio-oekonomischen Panels. In: SOEppapers (Nr. 521). In:
- Heil, Peter 2010: Was ist „Religion“? Eine Einführung in unser wissenschaftliches Reden über Religion. In: Internationale Zeitschrift für Philosophie und Psychosomatik (1), 1–29.
- Heinemann, Barbara/Wörmann, Viktorya 2010: Psychische Beanspruchung, Lebenszufriedenheit und Religiosität: Beeinflusst Religiosität den Zusammenhang zwischen psychischer Beanspruchung und Lebenszufriedenheit bei Patienten mit psychischen Erkrankungen? In: Internationale Zeitschrift für Philosophie und Psychosomatik 2010 (1).
- Hildebrand, Anja (Hrsg.), 2015: Vertriebsmitarbeiter und Innovationen. Wiesbaden: Springer Fachmedien.
- Hill, Peter C. 2013: Measurement Assessment and Issues in the Psychology of Religion and Spirituality. In: Paloutzian, Raymond F./Park, Crystal L. (Hrsg.): Handbook of the psychology of religion and spirituality, Second edition. New York: Guilford Press, 48–74.
- Hill, Peter C./Pargament, Kenneth I. 2003: Advances in the conceptualization and measurement of religion and spirituality. Implications for physical and mental health research. In: American Psychologist 58 (1), 64–74.
- Hill, Peter C./Pargament, Kenneth I./Hood, Ralph W./McCullough, Michael E./Swyers, James P./Larson, David B./Zinnbauer, Brian J. 2000: Conceptualizing Religion and Spirituality. Points of Commonality, Points of Departure. In: Journal for the Theory of Social Behaviour 30 (1), 51–77.
- Hofmann, Liane 2009: Spiritualität und Religiosität in der psychotherapeutischen Praxis. Dissertation. Carl von Ossietzky Universität. Oldenburg.
- Hong, Ryan Y./Cheung, Mike W.-L. 2015: The Structure of Cognitive Vulnerabilities to Depression and Anxiety. In: Clinical Psychological Science 3 (6), 892–912.

- Hopewell, Sally/Clarke, Mike/Mallett, Sue 2005: Grey Literature and Systematic Reviews. In: Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael (Hrsg.): *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester, West Sussex: Wiley, 49-72.
- Hox, Joop J. 2010: *Multilevel analysis*. 2. Auflage. New York, NY: Routledge.
- Huber, Stefan 2003: *Zentralität und Inhalt*. Wiesbaden: Springer VS.
- Hunter, John E./Schmidt, Frank L./Jackson, Gregg B. 1982: *Meta-analysis*. Beverly Hills: Sage.
- Ivtzan, Itai/Chan, Christine P. L./Gardner, Hannah E./Prashar, Kiran 2013: Linking religion and spirituality with psychological well-being: examining self-actualisation, meaning in life, and personal growth initiative. In: *Journal of religion and health* 52 (3), 915–929.
- Jackson, Brenda R./Bergeman, C. S. 2011: How Does Religiosity Enhance Well-Being? The Role of Perceived Control. In: *Psychology of religion and spirituality* 3 (2), 149–161.
- Jager Meezenbroek, Eltica de/Garssen, Bert/van den Berg, Machteld/van Dierendonck, Dirk/Visser, Adriaan/Schaufeli, Wilmar B. 2012: Measuring spirituality as a universal human experience: a review of spirituality questionnaires. In: *Journal of religion and health* 51 (2), 336–354.
- Jagodzinski, Wolfgang 2010: Economic, Social, and Cultural Determinants of Life Satisfaction. Are there Differences Between Asia and Europe? In: *Social indicators research* 97 (1), 85–104.
- Jak, Suzanne 2015: *Meta-Analytic Structural Equation Modelling*. Cham: Springer.
- Jak, Suzanne/Cheung, Mike W.-L. 2017: Accounting for Missing Correlation Coefficients in Fixed-Effects MASEM. In: *Multivariate behavioral research* 53 (1), 1–14.
- Jak, Suzanne/Cheung, Mike W.-L. 2018a: Meta-Analytic Structural Equation Modeling with Moderating Effects on SEM Parameters.
- Jak, Suzanne/Cheung, Mike W.-L. 2018b: Testing moderator hypotheses in meta-analytic structural equation modeling using subgroup analysis. In: *Behavior research methods*.
- Jak, Suzanne/Roorda, Debora L./Koomen, Helma M. Y. 2013: Meta-analytic structural equation modelling with missing correlations. In: *Netherlands Journal of Psychology* 67, 132–139.
- Jakobs, Monika 2006: Religiosität als biographische Verarbeitung von Religion. Religionssoziologische Perspektiven. In: Angel, Hans-Ferdinand (Hrsg.): *Religiosität. Anthropologische, theologische und sozialwissenschaftliche Klärungen. Praktische Theologie, Religionspädagogik, Diakonie*. Stuttgart: W. Kohlhammer, 116–132.
- Jim, Heather S. L./Pustejovsky, James E./Park, Crystal L./Danhauer, Suzanne C./Sherman, Allen C./Fitchett, George/Merluzzi, Thomas V./Munoz, Alexis R./George, Login/Snyder, Mallory A./Salsman, John M. 2015: Religion, spirituality, and physical health in cancer patients: A meta-analysis. In: *Cancer* 121 (21), 3760–3768.
- Joshanloo, Mohsen/Afshari, Samaneh 2011: Big Five Personality Traits and Self-Esteem as Predictors of Life Satisfaction in Iranian Muslim University Students. In: *Journal of Happiness Studies* 12 (1), 105–113.
- Judd, Daniel 1986: Religious Affiliation and mental health. In: *AMCAP Journal* 12 (2), 71–108.
- Kahneman, Daniel 2003: Objective Happiness. In: Kahneman, Daniel (Hrsg.): *Well-Being. Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, 3–25.
- Kahneman, Daniel/Krueger, Alan B. 2006: Developments in the Measurement of Subjective Well-Being. In: *Journal of Economic Perspectives* 20 (1), 3–24.
- Kashdan, Todd B./Nezlek, John B. 2012: Whether, when, and how is spirituality related to well-being? Moving beyond single occasion questionnaires to understanding daily process. In: *Personality & social psychology bulletin* 38 (11), 1523–1535.

- Ke, Zijun/Zhang, Qian/Tong, Xin 2018: Bayesian Meta-Analytic SEM: A One-Stage Approach to Modeling Between-Studies Heterogeneity in Structural Parameters. In: *Structural Equation Modeling: A Multidisciplinary Journal* 6 (4), 1–23.
- Kelley-Gillespie, Nancy 2009: An Integrated Conceptual Model of Quality of Life for Older Adults Based on a Synthesis of the Literature. In: *Applied Research in Quality of Life* 4 (3), 259–282.
- Kim-Prieto, Chu/Diener, Ed 2009: Religion as a source of variation in the experience of positive and negative emotions. In: *The Journal of Positive Psychology* 4 (6), 447–460.
- Klein, Constantin/Albani, Cornelia 2007: Religiosität und psychische Gesundheit. Eine Übersicht über Befunde, Erklärungsansätze und Konsequenzen für die klinische Praxis. In: *Psychiatrische Praxis* 34 (2), 58–65.
- Klein, Constantin/Albani, Cornelia 2011a: Die Bedeutung von Religion für die psychische Befindlichkeit: Mögliche Erklärungsansätze und allgemeines Wirkmodell. In: *German Journal for young Researcher* 3 (1), 8–58.
- Klein, Constantin/Albani, Cornelia 2011b: Religiosität und psychische Gesundheit - empirische Befunde und Erklärungsansätze. In: Klein, Constantin/Berth, Hendrik/Balck, Friedrich (Hrsg.): *Gesundheit - Religion - Spiritualität. Konzepte, Befunde und Erklärungsansätze. Gesundheitsforschung*. Weinheim: Juventa-Verlag, 215–246.
- Klein, Constantin/Berth, Hendrik/Balck, Friedrich 2011: Zur Einführung. Die Relevanz von Religiosität und Spiritualität für die Gesundheitsforschung. In: Klein, Constantin/Berth, Hendrik/Balck, Friedrich (Hrsg.): *Gesundheit - Religion - Spiritualität. Konzepte, Befunde und Erklärungsansätze. Gesundheitsforschung*. Weinheim: Juventa-Verlag, 11–24.
- Klein, Thomas/Kopp, Johannes/Rapp, Ingmar 2013: Metaanalyse mit Originaldaten / Meta-Analysis of Original Data. In: *Zeitschrift für Soziologie* 42 (3).
- Knoblauch, Hubert 2005: Einleitung. Soziologie der Spiritualität. In: *Zeitschrift für Religionswissenschaft* 13 (2).
- Knoblauch, Hubert 2008: Die populäre Religion. In: *tv diskurs* 12 (44), 42–47.
- Koenig, Harold G. 2008: Concerns about measuring "spirituality" in research. In: *The Journal of nervous and mental disease* 196 (5), 349–355.
- Koenig, Harold G. 2012: Religion, spirituality, and health: the research and clinical implications. In: *ISRN psychiatry* 2012, 278730.
- Koenig, Harold G./King, Dana E./Carson, Verna B. 2012: *Handbook of Religion and Health*. 2. Auflage: Oxford University Press.
- Koenig, Harold G./Larson, David B. 2009: Religion and mental health. Evidence for an association. In: *International Review of Psychiatry* 13 (2), 67–78.
- Koenig, Matthias/Wolf, Christof 2013: Religion und Gesellschaft – Aktuelle Perspektiven. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 65 (S1), 1–23.
- Kontantopoulos, Spyros/Hedges, Larry V. 2009: Analyzing Effect Sizes: Fixed-Effects Models. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 279–293.
- Kopp, Johannes/Lois, Daniel 2014: Ein Ausblick auf fortgeschrittene statistische Analyseverfahren. In: Kopp, Johannes/Lois, Daniel (Hrsg.): *Sozialwissenschaftliche Datenanalyse. Eine Einführung. Lehrbuch, 2., überarbeitete und aktualisierte Auflage*. Wiesbaden: Springer VS, 187–214.
- Körtner, Ulrich H. J. 2009: Spiritualität, Religion und Kultur - eine begriffliche Annäherung. In: Körtner, Ulrich H./Inthorn, Julia/Kletečka-Pulker, Maria et al. (Hrsg.): *Spiritualität, Religion und Kultur am Krankenbett. Schriftenreihe Ethik und Recht in der Medizin, Bd. 3*. Vienna: Springer Vienna.

- Kortt, Michael A./Dollery, Brian/Grant, Bligh 2015: Religion and Life Satisfaction Down Under. In: *Journal of Happiness Studies* 16 (2), 277–293.
- Kraft, Priscilla S./Bausch, Andreas 2016: How Do Transformational Leaders Promote Exploratory and Exploitative Innovation? Examining the Black Box through MASEM. In: *Journal of Product Innovation Management* 33 (6), 687–707.
- Krok, Dariusz 2014: Religiousness and perceived social support as predictive factors for mental health outcomes. In: *Archives of Psychiatry and Psychotherapy* 16, 65–76.
- Kusche, Elke 2013: Der Einfluss von Religiosität/Spiritualität auf den prä- und postoperativen Gesundheitszustand von Patienten mit Bypass-OP. Dissertation. Friedrich-Schiller-Universität. Jena.
- Lamers, Sanne M. A./Bolier, Linda/Westerhof, Gerben J./Smit, Filip/Bohlmeijer, Ernst T. 2012: The impact of emotional well-being on long-term recovery and survival in physical illness: a meta-analysis. In: *Journal of behavioral medicine* 35 (5), 538–547.
- Landis, Ronald S. 2013: Successfully Combining Meta-analysis and Structural Equation Modeling. Recommendations and Strategies. In: *Journal of Business and Psychology* 28 (3), 251–261.
- Larsen, Randy/Eid, Michael 2008: Ed Diener and the Science of Subjective Well-Being. In: Eid, Michael (Hrsg.): *The science of subjective well-being*. New York NY: The Guilford Press, 1–13.
- Lau, Esther Yuet Ying/Cheung, Sing-Hang/Lam, Jasmine/Hui, C. Harry/Cheung, Shu-Fai/Mok, Doris Shui Ying 2015: Purpose-Driven Life. Life Goals as a Predictor of Quality of Life and Psychological Health. In: *Journal of Happiness Studies* 16 (5), 1163–1184.
- Leifeld, Philip 2013: texreg: Conversion of Statistical Model Output in R to LATEX and HTML Tables. In: *Journal of Statistical Software* 55 (8).
- Leondari, Angeliki/Gialamas, Vasilios 2009: Religiosity and psychological well-being. In: *International journal of psychology : Journal international de psychologie* 44 (4), 241–248.
- Lesener, Tino/Gusy, Burkhard/Wolter, Christine 2019: The job demands-resources model: A meta-analytic review of longitudinal studies. In: *Work & Stress* 33 (1), 76–103.
- Lewis, Christopher Alan/Cruise, Sharon Mary 2006: Religion and happiness. Consensus, contradictions, comments and concerns. In: *Mental Health, Religion & Culture* 9 (3), 213–225.
- Liberati, Alessandro/Altman, Douglas G./Tetzlaff, Jennifer/Mulrow, Cynthia/Gøtzsche, Peter C./Ioannidis, John P. A./Clarke, Mike/Devereaux, P. J./Kleijnen, Jos/Moher, David 2009: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. In: *PLoS medicine* 6 (7).
- Lim, Chaeyoon 2015: Religion and Subjective Well-Being Across Religious Traditions. Evidence from 1.3 Million Americans. In: *Journal for the Scientific Study of Religion* 54 (4), 684–701.
- Lim, Chaeyoon/Putnam, Robert D. 2010: Religion, Social Networks, and Life Satisfaction. In: *American Sociological Review* 75 (6), 914–933.
- Lipsey, Mark W. 1997: Using linked meta-analysis to build policy models. In: *NIDA research monograph* 170, 216–233.
- Lipsey, Mark W./Wilson, David B. 2010: *Practical meta-analysis*. Thousand Oaks, Calif: Sage.
- Little, Todd D./Kline, Rex B. (Hrsg.), 2016: *Principles and practice of structural equation modeling*. 4th edition. New York: The Guilford Press.

- López-López, José A./Page, Matthew J./Lipsey, Mark W./Higgins, Julian P. T. 2018: Dealing with effect size multiplicity in systematic reviews and meta-analyses. In: *Research synthesis methods*.
- Lucas, Richard E. 2008: Personality and Subjective Well-Being. In: Eid, Michael (Hrsg.): *The science of subjective well-being*. New York NY: The Guilford Press, 171–194.
- Lucchini, Mario/Della Nella, Sara/Crivelli, Luca 2003: Genetic and Environmental Contributions to Life Satisfaction. In: Kahneman, Daniel (Hrsg.): *Well-Being. Foundations of Hedonic Psychology*. New York: Russell Sage Foundation, 176–195.
- Luhmann, Maike/Hofmann, Wilhelm/Eid, Michael/Lucas, Richard E. 2012: Subjective well-being and adaptation to life events: a meta-analysis. In: *Journal of personality and social psychology* 102 (3), 592–615.
- Luhmann, Niklas 2002: *Die Religion der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Lykken, David/Tellegen, Auke 1996: Happiness is a stochastic phenomenon. In: *Psychological Science* 7 (3), 186–189.
- MacLeod, Andrew K. 2015: Well-Being. Objectivism, Subjectivism or Subjectivism? In: *Journal of Happiness Studies* 16 (4), 1073–1089.
- Maggino, Filomena 2014: Indicators of Happiness vs Quality of Happiness: Methodology and Theory. In: Bruni, Luigino/Porta, Pier L. (Hrsg.): *Handbook of Research Methods and Applications in Happiness and Quality of Life*. Cheltenham, Gloucestershire: Edward Elgar Publishing, 196–228.
- Mahoney, Annette/Pargament, Kenneth 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. In: *Journal of Family Psychology* 15 (4), 559–596.
- Masters, Kevin S./Hooker, Stephanie A. 2013: Religiousness/spirituality, cardiovascular disease, and cancer: cultural integration for health research and intervention. In: *Journal of consulting and clinical psychology* 81 (2), 206–216.
- Masters, Kevin S./Spielmans, Glen I. 2007: Prayer and health: review, meta-analysis, and research agenda. In: *Journal of behavioral medicine* 30 (4), 329–338.
- Matt, George E./Cook, Thomas D. 2009: Threats to the Validity of Generalized Inferences. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 537–560.
- Mavridis, Dimitris/Salanti, Georgia 2013: A practical introduction to multivariate meta-analysis. In: *Statistical methods in medical research* 22 (2), 133–158.
- Mayerl, Jochen 2001: Ist Glück ein affektiver Indikator des subjektiven Wohlbefindens? In: *SISS: Schriftenreihe des Instituts für Sozialwissenschaften an der Universität Stuttgart* (Nr. 4).
- Mayring, Philipp 1991: Arbeit und Wohlbefinden. In: Abele, Andrea/Becker, Peter (Hrsg.): *Wohlbefinden. Theorie - Empirie - Diagnostik*. Juventa-Materialien, 2. Auflage. Weinheim: Juventa-Verlag, 51–70.
- McArdle, J. Jack/McDonald, Roderick P. 1984: Some algebraic properties of the Reticular Action Model for moment structures. In: *The British journal of mathematical and statistical psychology* 37 (2), 234–251.
- McClain, Colleen S./Rosenfeld, Barry/Breitbart, William 2003: Effect of spiritual well-being on end-of-life despair in terminally-ill cancer patients. In: *The Lancet* 361 (361), 1603–1607.
- McCullough, Michael E./Hoyt, W. T./Larson, David B./Koenig, Harold G./Thoresen, Carl E. 2000: Religious involvement and mortality: a meta-analytic review. In: *Health psychology* 19 (3), 211–222.

- McCullough, Michael E./Willoughby, Brian L. B. 2009: Religion, self-regulation, and self-control: Associations, explanations, and implications. In: *Psychological bulletin* 135 (1), 69–93.
- McGillivray, Mark 1991: The human development index. Yet another redundant composite development indicator? In: *World Development* 19 (10), 1461–1468.
- Mehnert, Anja 2006: Sinnfindung und Spiritualität bei Patienten mit chronischen körperlichen Erkrankungen. In: *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 49 (8), 780–787.
- Mehnert, Anja/Höcker, Anja 2011: Religion und körperliche Gesundheit - empirische Befunde und Erklärungsansätze. In: Klein, Constantin/Berth, Hendrik/Balck, Friedrich (Hrsg.): *Gesundheit - Religion - Spiritualität. Konzepte, Befunde und Erklärungsansätze. Gesundheitsforschung*. Weinheim: Juventa-Verlag, 247–258.
- Meulemann, Heiner 2007: Lebenszufriedenheit, Lebensbereiche und Religiosität. In: Nollmann, Gerd (Hrsg.): *Sozialstruktur und Gesellschaftsanalyse, 1. Auflage*. Wiesbaden: Springer VS, 261–277.
- Michel, Jesse S./Viswesvaran, Chockalingam/Thomas, Jeffrey 2011: Conclusions from meta-analytic structural equation models generally do not change due to corrections for study artifacts. In: *Research synthesis methods* 2 (3), 174–187.
- Mochon, Daniel/Norton, Michael I./Ariely, Dan 2011: Who Benefits from Religion? In: *Social indicators research* 101 (1), 1–15.
- Monod, Stéphanie/Brennan, Mark/Rochat, Etienne/Martin, Estelle/Rochat, Stéphane/Büla, Christophe J. 2011: Instruments measuring spirituality in clinical research. A systematic review. In: *Journal of general internal medicine* 26 (11), 1345–1357.
- Montazemi, Ali Reza/Qahri-Saremi, Hamed 2015: Factors affecting adoption of online banking. A meta-analytic structural equation modeling study. In: *Information & Management* 52 (2), 210–226.
- Moons, Philip/Budts, Werner/Geest, Sabina de 2006: Critique on the conceptualisation of quality of life: a review and evaluation of different conceptual approaches. In: *International journal of nursing studies* 43 (7), 891–901.
- Moreira-Almeida, Alexander/Lotufo Neto, Francisco/Koenig, Harold G. 2006: Religiousness and mental health. A review. In: *Revista Brasileira de Psiquiatria* 28 (3), 242–250.
- Morton, Kelly R./Lee, Jerry W./Martin, Leslie R. 2017: Pathways from Religion to Health. Mediation by Psychosocial and Lifestyle Mechanisms. In: *Psychology of religion and spirituality* 9 (1), 106–117.
- Möser, Guido/Schmidt, Peter 2016: Meta-analysis: An alternative to narrative reviews for synthesising social science research?
- Murken, Sebastian 1997: Ungesunde Religiosität - Entscheidungen der Psychologie? In: Klinkhammer, Gritt M. (Hrsg.): *Kritik an Religionen. Religionswissenschaft und der kritische Umgang mit Religionen*. Marburg: Diagonal-Verlag, 157–172.
- Murken, Sebastian 1998: *Gottesbeziehung und psychische Gesundheit*. Münster: Waxmann.
- Muthén, Linda/Muthén, Bengt 2017: *Mplus User's Guide*. 8. Edition. Los Angeles.
- Myers, David G. 2000: The funds, friends, and faith of happy people. In: *American Psychologist* 55 (1), 56–67.
- Myers, David G. 2008: Religion and Human Flourishing. In: Eid, Michael (Hrsg.): *The science of subjective well-being*. New York NY: The Guilford Press, 323–343.
- Myers, David G./Diener, Ed 1995: Who is happy? In: *Psychological Science* 6 (1), 10–19.
- Mytko, Johanna/Knight, Sara 1999: Body, mind and spirit: towards the integration of religiosity and spirituality in cancer quality of life research. In: *Psycho-Oncology* (8), 439–450.

- Neale, Michael C./Hunter, Michael D./Pritikin, Joshua N./Zahery, Mahsa/Brick, Timothy R./Kirkpatrick, Robert M./Estabrook, Ryne/Bates, Timothy C./Maes, Hermine H./Boker, Steven M. 2016: OpenMx 2.0: Extended Structural Equation and Statistical Modeling. In: *Psychometrika* 81 (2), 535–549.
- Neale, Michael C./Miller, M. B. 1997: The use of likelihood-based confidence intervals in genetic models. In: *Behavior genetics* 27 (2), 113–120.
- Neville, S./Alpass, F. 1999: Factors influencing health and well-being in the older adult. In: *Nursing praxis in New Zealand inc* 14 (3), 361–365.
- Niemand, Thomas (Hrsg.), 2014: *Produktpiraterie*. Wiesbaden: Springer Fachmedien.
- Norris, Pippa/Inglehart, Ronald 2011: *Sacred and Secular*. Cambridge: Cambridge University Press.
- Okulicz-Kozaryn, Adam 2009: *Religiosity and Life Satisfaction*.
- Okulicz-Kozaryn, Adam 2010: Religiosity and life satisfaction across nations. In: *Mental Health, Religion & Culture* 13 (2), 155–169.
- Okun, Morris A./Olding, Robert W./Cohn, Catherine M. 1990: A meta-analysis of subjective well-being interventions among elders. In: *Psychological bulletin* 108 (2), 257–266.
- Okun, Morris A./Stock, William A. 1987: Correlates and Components of Subjective Well-Being Among the Elderly. In: *Journal of Applied Gerontology* 6 (1), 95–112.
- Oman, Doug 2013: *Defining Religion and Spirituality*. In: Paloutzian, Raymond F./Park, Crystal L. (Hrsg.): *Handbook of the psychology of religion and spirituality*, Second edition. New York: Guilford Press, 23–47.
- Oort, Frans J./Jak, Suzanne 2016: Maximum likelihood estimation in meta-analytic structural equation modeling. In: *Research synthesis methods* 7 (2), 156–167.
- Ormel, Johan/Lindenberg, Siegwart/Steeverink, Nardi/Verbrugge, Lois M. 1999: Subjective Well-Being and Social Production Functions. In: *Social indicators research* 46 (1), 61–90.
- Osborne, Danny/Milojev, Petar/Sibley, Chris G. 2016: Examining the indirect effects of religious orientations on well-being through personal locus of control. In: *European Journal of Social Psychology* 46 (4), 492–505.
- Ostermann, Thomas/Büssing, Arndt/Matthiessen, Peter F. 2004: Pilotstudie zur Entwicklung eines Fragebogens zur Erfassung der spirituellen und religiösen Einstellung und des Umgangs mit Krankheit (SpREUK). In: *Forschende Komplementärmedizin und klassische Naturheilkunde* 11 (6), 346–353.
- Pargament, Kenneth I. 2002: Authors' responses. Is Religion Nothing But. .? Explaining Religion Versus Explaining Religion Away. In: *Psychological Inquiry* 13 (3), 239–244.
- Park, Crystal L. 2007: Religiousness/spirituality and health: a meaning systems perspective. In: *Journal of behavioral medicine* 30 (4), 319–328.
- Park, Crystal L./Sherman, Allen C./Jim, Heather S. L./Salsman, John M. 2015: Religion/spirituality and health in the context of cancer: Cross-domain integration, unresolved issues, and future directions. In: *Cancer* 121 (21), 3789–3794.
- Park, Crystal L./Slattery, Jeanne M. 2013: Religion, Spirituality and Mental Health. In: Paloutzian, Raymond F./Park, Crystal L. (Hrsg.): *Handbook of the psychology of religion and spirituality*, Second edition. New York: Guilford Press, 540–559.
- Park, Jisung/Roh, Soonhee/Yeo, Younsook 2012: Religiosity, social support, and life satisfaction among elderly Korean immigrants. In: *The Gerontologist* 52 (5), 641–649.
- Park, Sunyoung/Beretvas, S. Natasha 2018: Synthesizing effects for multiple outcomes per study using robust variance estimation versus the three-level model. In: *Behavior research methods*.
- Pavot, William/Diener, Ed 1993: Review of the Satisfaction With Life Scale. In: *Psychological Assessment* 5 (2), 164–172.

- Pavot, William/Diener, Ed 2008: The Satisfaction With Life Scale and the emerging construct of life satisfaction. In: *The Journal of Positive Psychology* 3 (2), 137–152.
- Payne, I. Reed/Bergin, Allen E./Bielema, Kimberly A./Jenkins, Paul H. 1991: Review of Religion and Mental Health. In: *Prevention in Human Services* 9 (2), 11–40.
- Peters, Lawrence H./Hartke, Darrell D./Pohmann, John T. 1985: Fiedler's Contingency Theory of Leadership. An application of the meta-analysis procedures of Schmidt and Hunter. In: *Psychological bulletin* 97 (2), 274–285.
- Peterson, Mark/Webb, Dave 2006: Religion and Spirituality in Quality of Life Studies. In: *Applied Research in Quality of Life* 1 (1), 107–116.
- Pickel, Gert 2011: *Religionssoziologie*. 1. Auflage. Wiesbaden: Springer VS.
- Pickel, Gert/Sammet, Kornelia 2014: *Einführung in die Methoden der sozialwissenschaftlichen Religionsforschung*. Wiesbaden: Springer VS.
- Pigott, Therese D. 2009: Handling Missing Data. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 399–416.
- Pigott, Therese D. 2012: *Advances in Meta-Analysis*. Boston: Springer US.
- Pinquart, Martin 2002: Creating and maintaining purpose in life in old age. A meta-analysis. In: *Ageing International* 27 (2), 90–114.
- Pinquart, Martin/Sörensen, Silvia 2000: Influences of socioeconomic status, social network, and competence on subjective well-being in later life. A meta-analysis. In: *Psychology and aging* 15 (2), 187–224.
- Pollack, Detlef 2014: *Religionssoziologie in Deutschland seit 1945*. In: *Preprints and Working Papers of the Center for Religion and Modernity* (Nr. 6).
- Pollack, Detlef 2015: *Religionssoziologie in Deutschland seit 1945. Tendenzen – Kontroversen – Konsequenzen*. In: *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie* 67 (3), 433–474.
- Poloma, Margaret M./Pendleton, Brian F. 1990: Religious domains and general well-being. In: *Social indicators research* 22 (3), 255–276.
- Prati, Gabriele/Pietrantonio, Luca 2009: Optimism, Social Support, and Coping Strategies As Factors Contributing to Posttraumatic Growth. A Meta-Analysis. In: *Journal of Loss and Trauma* 14 (5), 364–388.
- Pressman, Sarah D./Cohen, Sheldon 2005: Does positive affect influence health? In: *Psychological bulletin* 131 (6), 925–971.
- Prevost, A. Toby/Mason, Dan/Griffin, Simon/Kinmonth, Ann-Louise/Sutton, Stephen/Spiegelhalter, David 2007: Allowing for correlations between correlations in random-effects meta-analysis of correlation matrices. In: *Psychological methods* 12 (4), 434–450.
- Quinn, Jamie M. 2016: *Predictors of Reading Comprehension*. Dissertation. Florida State University. Tallahassee.
- Quinn, Jamie M./Wagner, Richard K. 2018: Using Meta-analytic Structural Equation Modeling to Study Developmental Change in Relations Between Language and Literacy. In: *Child development* 89 (6), 1956–1969.
- R Core Team 2018: *R: A language and environment for statistical computing*. [Computer software]. Wien: R Foundation for Statistical Computing. In: <https://www.r-project.org/>.
- Raudenbush, Stephen W. 2009: Analyzing effect sizes: Random effects models. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 296–315.
- Raudenbush, Stephen W./Becker, Betsy J./Kalaian, Hripsime 1988: Modeling multivariate effect sizes. In: *Psychological bulletin* 103 (1), 111–120.
- Raudenbush, Stephen W./Bryk, Anthony S. 1985: Empirical Bayes Meta-Analysis. In: *Journal of Educational Statistics* 10 (2), 75.

- Reed, Reffrey/Baxter, Pam 2009: Using Reference Databases. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 73–101.
- Reyes-Gómez, Juan David/Obeidat, Bader Yousef 2016: The Effect of Strategic Orientation on Organizational Performance: The Mediating Role of Innovation. A meta-analytic path analysis. In: *International Journal of Communications, Network and System Sciences* 09 (11), 478–505.
- Riesebrodt, Martin 2007: *Cultus und Heilsversprechen*. München: Beck.
- Riley, Richard D. 2009: Multivariate meta-analysis. The effect of ignoring within-study correlation. In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 172 (4), 789–811.
- Rizvi, Mohd Ahsan Kabir/Hossain, Mohammad Zakir 2016: Relationship Between Religious Belief and Happiness. A Systematic Literature Review. In: *Journal of religion and health*, 1561–1582.
- Rosa, Hartmut/Strecker, David/Kottmann, Andrea 2009: *Soziologische Theorien*. Konstanz: UVK.
- Rosenthal, Robert 2010: *Meta-analytic procedures for social research*. Newbury Park: Sage.
- Rosopa, Patrick J./Kim, Brian 2017: Robustness of statistical inferences using linear models with meta-analytic correlation matrices. In: *Human Resource Management Review* 27 (1), 216–236.
- Rothstein, Hannah R./Hopewell, Sally 2009: Grey Literature. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 103–125.
- Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael 2005: Publication Bias in Meta-Analysis. In: Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael (Hrsg.): *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester, West Sussex: Wiley, 1–8.
- RStudio Team 2015: *RStudio: Integrated Development for R [Computer software]*. Boston: RStudio, Inc. In: <http://www.rstudio.com/>.
- Ruddat, Michael/Schulz, Marlene 2010: Möglichkeiten und Grenzen einer integrativen Metaanalyse in der Soziologie. *35 (1)*, 289–322.
- Ryff, Carol D. 1989: Happiness is everything, or is it? Explorations on the meaning of psychological well-being. In: *Journal of personality and social psychology* 57 (6), 1069–1081.
- Ryff, Carol D./Boylan, Morozink 2014: Linking Happiness to health: comparisons between hedonic and eudaimonic well-being. In: Bruni, Luigino/Porta, Pier L. (Hrsg.): *Handbook of Research Methods and Applications in Happiness and Quality of Life*. Cheltenham, Gloucestershire: Edward Elgar Publishing, 53–70.
- Sabatier, C./Mayer, B./Friedlmeier, M./Lubiewska, K./Trommsdorff, G. 2011: Religiosity, Family Orientation, and Life Satisfaction of Adolescents in Four Countries. In: *Journal of Cross-Cultural Psychology* 42 (8), 1375–1393.
- Sagar, Ambuj D./Najam, Adil 1998: The human development index. A critical review. In: *Ecological Economics* 25 (3), 249–264.
- Salsman, John M./Fitchett, George/Merluzzi, Thomas V./Sherman, Allen C./Park, Crystal L. 2015: Religion, spirituality, and health outcomes in cancer: A case for a meta-analytic investigation. In: *Cancer* 121 (21), 3754–3759.
- Salsman, John M./Pustejovsky, James E./Jim, Heather S. L./Munoz, Alexis R./Merluzzi, Thomas V./George, Login/Park, Crystal L./Danahauer, Suzanne C./Sherman, Allen C./Snyder, Mallory A./Fitchett, George 2015: A meta-analytic approach to examining the correlation between religion/spirituality and mental health in cancer. In: *Cancer* 121 (21), 3769–3778.

- Saroglou, Vassilis 2002: Religion and the five factors of personality. A meta-analytic review. In: *Personality and Individual Differences* 32 (1), 15–25.
- Saroglou, Vassilis/Delpierre, Vanessa/Dernelle, Rebecca 2004: Values and religiosity. A meta-analysis of studies using Schwartz's model. In: *Personality and Individual Differences* 37 (4), 721–734.
- Sawatzky, Richard/Ratner, Pamela A./Chiu, Lyren 2005: A Meta-Analysis of the Relationship Between Spirituality and Quality of Life. In: *Social indicators research* 72 (2), 153–188.
- Scalco, Andrea/Noventa, Stefano/Sartori, Riccardo/Ceschi, Andrea 2017: Predicting organic food consumption. A meta-analytic structural equation model based on the theory of planned behavior. In: *Appetite* 112, 235–248.
- Scammacca, Nancy/Roberts, Greg/Stuebing, Karla K. 2014: Meta-Analysis With Complex Research Designs. Dealing With Dependence From Multiple Measures and Multiple Group Comparisons. In: *Review of Educational Research* 84 (3), 328–364.
- Scherer, Ronny/Siddiq, Fazilat/Tondeur, Jo 2019: The technology acceptance model (TAM): A meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education. In: *Computers & Education* 128, 13–35.
- Schewe, Anna F./Hülshager, Ute R./Maier, Günter W. 2014: Metaanalyse–praktische Schritte und Entscheidungen im Umsetzungsprozess. In: *Zeitschrift für Arbeits- und Organisationspsychologie A&O* 58 (4), 186–205.
- Schimack, Ulrich 2008: The Structure of Subjective Well-Being. In: Eid, Michael (Hrsg.): *The science of subjective well-being*. New York NY: The Guilford Press, 97–123.
- Schlaegel, Christopher/Koenig, Michael 2014: Determinants of Entrepreneurial Intent. A Meta-Analytic Test and Integration of Competing Models. In: *Entrepreneurship Theory and Practice* 38 (2), 291–332.
- Schmidt, Frank L./Hunter, John Edward 2015: *Methods of meta-analysis*. 3. Auflage. Los Angeles: Sage.
- Schreiber, Judith A./Brockopp, Dorothy Y. 2012: Twenty-five years later - what do we know about religion/spirituality and psychological well-being among breast cancer survivors? A systematic review. In: *Journal of cancer survivorship: research and practice* 6 (1), 82–94.
- Schudey, Alexander P./Jensen, Ove/Kraicz, Nils D. 2013: Expatriates-Training = Expatriates-Einsatzserfolg? Eine Metaanalyse. In: *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 65 (6), 518–552.
- Sedikides, Constantine/Gebauer, Jochen E. 2009: Religiosity as self-enhancement: a meta-analysis of the relation between socially desirable responding and religiosity. In: *Personality and social psychology review* 14 (1), 17–36.
- Sergent, Adam 2015: *Moderation and Mediation of the Spirituality and Subjective Well-Being Relation*. Dissertation. Colorado State University. Fort Collins.
- Seybold, Kevin S./Hill, Peter C. 2001: The Role of Religion and Spirituality in Mental and Physical Health. In: *Current Directions in Psychological Science* 10 (1), 21–24.
- Shadish, William R./Haddock, Keith C. 2009: Combining Estimates of Effect Sizes. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 257–293.
- Shadish, William R./Sweeney, R. B. 1991: Mediators and moderators in meta-analysis. There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. In: *Journal of consulting and clinical psychology* 59 (6), 883–893.
- Sheng, Zitong/Kong, Wenmo/Cortina, Jose M./Hou, Shuofei 2016: Analyzing matrices of meta-analytic correlations: current practices and recommendations. In: *Research synthesis methods* 7 (2), 187–208.

- Sherkat, Darren E./Ellison, Christopher G. 1999: Recent developments and current controversies in the sociology of religion. In: *Annual Review of Sociology* 25 (1), 363–394.
- Sherman, Allen C./Merluzzi, Thomas V./Pustejovsky, James E./Park, Crystal L./George, Logan/Fitchett, George/Jim, Heather S. L./Munoz, Alexis R./Danhauer, Suzanne C./Snyder, Mallory A./Salsman, John M. 2015: A meta-analytic review of religious or spiritual involvement and social health among cancer patients. In: *Cancer* 121 (21), 3779–3788.
- Shor, Eran/Roelfs, David J. 2013: The Longevity Effects of Religious and Nonreligious Participation. A Meta-Analysis and Meta-Regression. In: *Journal for the Scientific Study of Religion* 52 (1), 120–145.
- Smith, Timothy B./Bartz, Jeremy/Scott Richards, P. 2007: Outcomes of religious and spiritual adaptations to psychotherapy. A meta-analytic review. In: *Psychotherapy Research* 17 (6), 643–655.
- Smith, Timothy B./McCullough, Michael E./Poll, Justin 2003: Religiousness and depression: evidence for a main effect and the moderating influence of stressful life events. In: *Psychological bulletin* 129 (4), 614–636.
- Stavrova, Olga/Fetchenhauer, Detlef/Schlosser, Thomas 2013: Why are religious people happy? The effect of the social norm of religiosity across countries. In: *Social science research* 42 (1), 90–105.
- Steffen, Patrick R./Masters, Kevin S./Baldwin, Scott 2016: What Mediates the Relationship Between Religious Service Attendance and Aspects of Well-Being? In: *Journal of religion and health* 56 (1), 158–170.
- Steinmetz, Holger/Isidor, Rodrigo 2013: *Meta-Analytical Structural Equation Modeling: An Easy Introduction to the Two-Step Approach*.
- Steinmetz, Holger/Isidor, Rodrigo/Baeuerle, Naissa 2012: Testing the Circular Structure of Human Values: A Meta-Analytical Structural Equation Modelling Approach. In: *Survey Research Methods* 6 (1), 61–75.
- Sterne, Jonathan/Becker, Betsy J./Egger, Matthias 2005: The Funnel Plot. In: Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael (Hrsg.): *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester, West Sussex: Wiley, 75–98.
- Sterne, Jonathan/Egger, Matthias 2005: Regression Methods to Detect Publication and Other Bias in Meta-Analysis. In: Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael (Hrsg.): *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester, West Sussex: Wiley, 99–110.
- Stolz, Jörg/Buchard, Emmanuelle 2014: Religionssoziologie. In: *VSH-Bulletin* (1), 55–63.
- Sutton, Alexander J. 2005: Evidence Concerning the Consequences of Publication and Related Biases. In: Rothstein, Hannah R./Sutton, Alexander J./Borenstein, Michael (Hrsg.): *Publication bias in meta-analysis. Prevention, assessment and adjustments*. Chichester, West Sussex: Wiley, 175–192.
- Sutton, Alexander J. 2009: Publication Bias. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 435–452.
- Tang, Ryan W./Cheung, Mike W.-L. 2016: Testing IB theories with meta-analytic structural equation modeling. In: *Review of International Business and Strategy* 26 (4), 472–492.
- Tanner-Smith, Emily E./Tipton, Elizabeth 2014: Robust variance estimation with dependent effect sizes. Practical considerations including a software tutorial in Stata and spss. In: *Research synthesis methods* 5 (1), 13–30.
- Tay, Louis/Li, Miao/Myers, David G./Diener, Ed 2014: Religiosity and Subjective Well-Being: An International Perspective. In: Kim-Prieto, Chu (Hrsg.): *Religion and Spirituality Across Cultures. Cross-Cultural Advancements in Positive Psychology*, Bd. 9. Dordrecht: Springer Netherlands, 163–175.

- Thoresen, Carl E. 1999: Spirituality and health: is there a relationship? In: *Journal of health psychology* 4 (3), 291–300.
- Thornton, A./Lee, P. 2000: Publication bias in meta-analysis: its causes and consequences. In: *Journal of clinical epidemiology* 53 (2), 207–216.
- Tiliouine, Habib 2009: Measuring Satisfaction with Religiosity and Its Contribution to the Personal Well-Being Index in a Muslim Sample. In: *Applied Research in Quality of Life* 4 (1), 91–108.
- Tipton, Elizabeth 2015: Small sample adjustments for robust variance estimation with meta-regression. In: *Psychological methods* 20 (3), 375–393.
- Unterrainer, H. F./Lewis, A. J./Fink, A. 2014: Religious/Spiritual Well-being, personality and mental health: a review of results and conceptual issues. In: *Journal of religion and health* 53 (2), 382–392.
- Ura, Karma 2012: *A short guide to gross national happiness index*. Thimphu: Centre for Bhutan Studies.
- Urban, Dieter/Fiebig, Joachim 2015: *Quantitative Meta-Analyse zur Überprüfung sozialwissenschaftlicher Hypothesen*. Weinheim: Beltz-Juventa.
- Urban, Dieter/Mayerl, Jochen 2014: *Strukturgleichungsmodellierung*. Wiesbaden: Springer Fachmedien.
- Urban, Dieter/Mayerl, Jochen 2018: *Angewandte Regressionsanalyse*. Wiesbaden: Springer Fachmedien.
- Utsch, Michael/Klein, Constantin 2011: Religion, Religiosität, Spiritualität. Bestimmungsversuche für komplexe Begriffe. In: Klein, Constantin/Berth, Hendrik/Balck, Friedrich (Hrsg.): *Gesundheit - Religion - Spiritualität. Konzepte, Befunde und Erklärungsansätze*. Gesundheitsforschung. Weinheim: Juventa-Verlag, 25–46.
- Vaas, Rüdiger/Blume, Michael 2009: *Gott, Gene und Gehirn*. Stuttgart: Hirzel.
- Valentine, Jeffrey C. 2009: Judging the quality of primary research. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 129–145.
- van Cappellen, Patty/Toth-Gauthier, Maria/Saroglou, Vassilis/Fredrickson, Barbara L. 2016: Religion and Well-Being. The Mediating Role of Positive Emotions. In: *Journal of Happiness Studies* 17 (2), 485–505.
- van den Noortgate, Wim/López-López, José A./Marin-Martinez, Fulgencio/Sanchez-Meca, Julio 2013: Three-level meta-analysis of dependent effect sizes. In: *Behavior research methods* 45 (2), 576–594.
- van der Jagt-Jelsma, W./Vries-Schot, M. R. de/Jong, Rint de/Hartman, C. A./Verhulst, F. C./Klip, H./van Deurzen, P A M/Buitelaar, J. K. 2015: Religiosity and mental health of pre-adolescents with psychiatric problems and their parents: the TRAILS study. In: *European psychiatry* 30 (7), 845–851.
- Veenhoven, Ruut 2007: Quality of Life research. In: Bryant, Clifton D./Peck, Dennis L. (Hrsg.): *21st century sociology. A reference handbook*. Gale virtual reference library. Thousand Oaks, Calif: Sage, 54–62.
- Veenhoven, Ruut 2008: Sociological Theories of Subjective Well-Being. In: Eid, Michael (Hrsg.): *The science of subjective well-being*. New York NY: The Guilford Press, 44–61.
- Veenhoven, Ruut 2010: The four qualities of Life Ordering concepts and measures of the good life. In: McGillivray, Mark/Clarke, Matthew (Hrsg.): *Understanding human well-being*. Tokyo: United Nations University Press, 74–100.
- Veenhoven, Ruut/Ehrhardt, Joop 1995: The cross-national pattern of happiness. Test of predictions implied in three theories of happiness. In: *Social indicators research* 34 (1), 33–68.

- Ventis, W. Larry 1995: The Relationships Between Religion and Mental Health. In: *Journal of Social Issues* 51 (2), 33–48.
- Viechtbauer, Wolfgang 2010: Conducting Meta-Analyses in R with the metafor Package. In: *Journal of Statistical Software* 36 (3).
- Viechtbauer, Wolfgang/Cheung, Mike W.-L. 2010: Outlier and influence diagnostics for meta-analysis. In: *Research synthesis methods* 1 (2), 112–125.
- Viechtbauer, Wolfgang/López-López, José A./Sanchez-Meca, Julio/Marin-Martinez, Fulgencio 2015: A comparison of procedures to test for moderators in mixed-effects meta-regression models. In: *Psychological methods* 20 (3), 360–374.
- Visser, Anja/Garssen, Bert/Vingerhoets, Ad 2010: Spirituality and well-being in cancer patients: a review. In: *Psycho-Oncology* 19 (6), 565–572.
- Viswesvaran, Chockalingam/Ones, Deniz S. 1995: Theory Testing. Combining Psychometric Meta-Analysis and Structural Equations Modeling. In: *Personal Psychology* 48 (4), 865–885.
- Wagner, Michael/Weiß, Bernd 2003: Bilanz der deutschen Scheidungsforschung. Versuch einer Meta-Analyse. In: *Zeitschrift für Soziologie* 32 (1), 29–49.
- Wagner, Michael/Weiß, Bernd 2014: Meta-Analyse. In: Baur, Nina/Blasius, Jörg (Hrsg.): *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer VS, 1117–1126.
- Ward, Andrew 2010: The Relationship Between Religiosity and Religious Coping to Stress Reactivity and Psychological Well-Being. Dissertation. Georgia State University. Atlanta.
- Weis, Robert/Speridakos, Elena C. 2011: A Meta-Analysis of hope enhancement strategies in clinical and community settings. In: *Psychology of Well-Being: Theory, Research and Practice* 1 (1), 5.
- Weiß, Bernd/Wagner, Michael 2008: Potentiale und Probleme von Meta-Analysen in der Soziologie. In: *Sozialer Fortschritt* 57 (10-11), 250–256.
- White, Howard 2009: Scientific communication and literature retrieval. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 51–71.
- White, Sarah C./Blackmore, Chloe (Hrsg.), 2016: *Cultures of Wellbeing*. Basingstoke, Hampshire: Palgrave Macmillan.
- WHO o. J.: *Verfassung der Weltgesundheitsorganisation*.
- Wilson, David B. 2009: Systematic Coding. In: Cooper, Harris M./Hedges, Larry V./Valentine, Jeffrey C. (Hrsg.): *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation, 159–176.
- Wilson, Sandra Jo/Polanin, Joshua R./Lipsey, Mark W. 2016: Fitting meta-analytic structural equation models with complex datasets. In: *Research synthesis methods* 7 (2), 121–139.
- Witter, Robert A./Stock, William A./Okun, Morris A./Haring, Marilyn J. 1985: Religion and Subjective Well-Being in Adulthood. A Quantitative Synthesis. In: *Review of Religious Research* 26 (4), 332.
- Wong, Y. Joel/Rew, Lynn/Slaikeu, Kristina D. 2006: A systematic review of recent research on adolescent religiosity/spirituality and mental health. In: *Issues in mental health nursing* 27 (2), 161–183.
- Worthington, Everett L./Sandage, Steven J. 2001: Religion and spirituality. In: *Psychotherapy: Theory, Research, Practice, Training* 38 (4), 473–478.
- Wu, Andrew/Wang, Jing-Yu/Jia, Cun-Xian 2015: Religion and Completed Suicide: a Meta-Analysis. In: *PloS one* 10 (6), 1-14.
- Wu, Meng-Jia/Becker, Betsy J. 2013: Synthesizing regression results: a factored likelihood method. In: *Research synthesis methods* 4 (2), 127–143.

- Yonker, Julie E./Schnabelrauch, Chelsea A./Dehaan, Laura G. 2012: The relationship between spirituality and religiosity on psychological outcomes in adolescents and emerging adults: a meta-analytic review. In: *Journal of adolescence* 35 (2), 299–314.
- Yoon, Eunju/Chang, Christine Chih-Ting/Clawson, Angela/Knoll, Michael/Aydin, Fatma/Bargisgian, Laura/Hughes, Kelly 2014: Religiousness, spirituality, and eudaimonic and hedonic well-being. In: *Counselling Psychology Quarterly* 28 (2), 132–149.
- Yu, Jia Joya/Downes, Patrick E./Carter, Kameron M./O'Boyle, Ernest H. 2016: The problem of effect size heterogeneity in meta-analytic structural equation modeling. In: *The Journal of applied psychology* 101 (10), 1457–1473.
- Yu, Jia Joya/Downes, Patrick E./Carter, Kameron M./O'Boyle, Ernest 2018: The heterogeneity problem in meta-analytic structural equation modeling (MASEM) revisited: A reply to Cheung. In: *The Journal of applied psychology* 103 (7), 804–811.
- Yuan, Ke-Hai 2016: Meta analytical structural equation modeling: comments on issues with current methods and viable alternatives. In: *Research synthesis methods* 7 (2), 215–231.
- Yuan, Ke-Hai/Kano, Yutaka 2018: Meta-Analytical SEM: Equivalence Between Maximum Likelihood and Generalized Least Squares. In: *Journal of Educational and Behavioral Statistics* 43 (6), 693–720.
- Zarzycka, Beata/Rydz, Elżbieta 2014: Centrality of Religiosity and Sense of Coherence. A Cross-sectional Study with Polish Young, Middle and Late Adults. In: *International Journal of Social Science Studies* 2 (2), 126–136.
- Zhang, Ying 2011: Meta-analytic Structural Equation Modeling (MASEM): Comparison of the multivariate methods. Dissertation. Florida State University. Tallahassee.
- Zinnbauer, Brian J./Pargament, Kenneth I./Cole, Brenda/Rye, Mark S./Butter, Eric M./Belavich, Timothy G./Hipp, Kathleen M./Scott, Allie B./Kadar, Jill L. 1997: Religion and Spirituality. Unfuzzifying the Fuzzy. In: *Journal for the Scientific Study of Religion* 36 (4), 549.
- Zuckerman, Miron/Silberman, Jordan/Hall, Judith A. 2013: The relation between intelligence and religiosity: a meta-analysis and some proposed explanations. In: *Personality and social psychology review* 17 (4), 325–354.
- Zwingmann, Christian/Klein, Constantin 2012: Deutschsprachige Fragebögen zur Messung von Religiosität und Spiritualität. In: *Spiritual Care - Zeitschrift für Spiritualität in den Gesundheitsberufen* 1 (3), 7–21.