Entwicklungshintergründe: Grundlagen und Methoden

European Nursing care Pathways

© TIMDAVIDCOLLECTION/stock.adobe.com

September 2024

Dr. Pia Wieteck
Sebastian Kraus (M.Sc.)
Dr. Qiumei Jiang-Siebert
Alexandra Mannes (B.Sc. OT)
Dr. Natalie Hubenthal
Sabine Kopp (B.Sc.)

Thieme RECOM

Inhaltsverzeichnis

1. Die evidenzbasierte Weiterentwicklung von ENP ... 3
2. Die systematische Suche nach der bestmöglichen Evidenz ... 4
2.1 Strategie und Kernelemente der systematischen, internationalen Literatursuche ... 4
2.1.1 Die Fragestellung(en) ... 5
2.1.2 Ableitung von Suchbegriffen ... 7
2.1.3 Ein- und Ausschlusskriterien ... 7
2.1.4 Definition der Fachdatenbanken und weiterer Anlaufstellen zur Recherche ... 8
2.1.5 Entwicklung von Suchphrasen und Einflussmöglichkeiten auf den Rechercheverlauf ... 10
2.1.6 Abstract Screening und Volltextbeschaffung ... 13
2.2 Kriterien und Instrumente für die Evidenzbewertung, Evidenzselektion und Datenextraktion ... 15
2.2.1 Critical Appraisal von Publikationen im ENP-Entwicklungsprozess ... 15
2.2.2 Evidenzsynthese und Bewertung einzelner Items/Handlungskonzepte ... 28
2.2.3 Evidenztabelle zur Studienbewertung im Rahmen der ENP-(Weiter-) Entwicklung ... 29
2.3 Der weitere Verlauf des ENP-Entwicklungsprozesses ... 41
Literatur ... 42
Anlage I: Exemplarisch befüllte Evidenztabelle im Rahmen der ENP-Weiterentwicklung seit 2021 ... 48
Anlage II: Checklisten zum Critical Appraisal verschiedener Publikations- und Studientypen seit 2021 ... 49
Anlage III: Protokoll zur Dokumentation einer systematischen Literaturrecherche ... 67

Entwicklungshintergründe: Grundlagen und Methoden

1. Die evidenzbasierte Weiterentwicklung von ENP

Unter den European Nursing care Pathways (ENP) ist eine standardisierte Pflegefachsprache zu verstehen, die monohierarchische Strukturen aufweist und das Pflegefachwissen in Form sogenannter Praxisleitlinien klassifiziert und zur Verfügung stellt. Anhand der nachfolgenden Abbildung kann der systematische Weiterentwicklungsprozess von ENP mit seinen grundsätzlichen Schritten nachvollzogen werden, der kontinuierlich verbessert wurde und wird. Es wird in der Regel jährlich eine neue Datenbankversion von ENP zur Verfügung gestellt, Buchveröffentlichungen werden für gewöhnlich im Dreijahresrhythmus realisiert.

Abbildung 1: Schematisierte Darstellung des systematischen Weiterentwicklungsprozesses von ENP

Beeinflusst durch gesundheitspolitische Entscheidungen, Anwenderfeedback sowie neue wissenschaftliche Erkenntnisse aus der Pflege und benachbarten Disziplinen der Gesundheitsversorgung wird jährlich entschieden, welche ENP-Praxisleitlinien einer systematischen Überprüfung und ggf. einer Überarbeitung unterzogen werden. Die Entwicklungsqualität von pflegediagnosenbezogenen Praxisleitlinien ist für die Nutzer(innen) der Pflegeklassifikation bedeutend, da diese im pflegerischen Entscheidungsfindungsprozess eine hohe Sicherheit für evidenzbasiertes Arbeiten ermöglicht. Umso wichtiger erscheint in diesem Zusammenhang daher auch die möglichst transparente Darlegung der wissenschaftlichen Arbeitsmethoden und Vorgehensweisen zur fachlich-inhaltlichen Weiterentwicklung bestehender bzw. Erarbeitung neuer ENP-Praxisleitlinien.

2. Die systematische Suche nach der bestmöglichen Evidenz

Die systematische Weiter- bzw. Neuentwicklung einer ENP-Praxisleitlinie verläuft in der Regel in einem mehrstufigen Prozess. Unter anderem folgende Fragestellungen sind dabei relevant:

- Handelt es sich bei dem beobachteten Phänomen um ein pflegediagnostisches Konzept?
- Finden sich in der Literatur evidenzbasierte Kennzeichen/Risikofaktoren/Ursachen, Ressourcen, welche die Pflegediagnose bestätigen bzw. im ursächlichen/beeinflussenden Zusammenhang mit der Pflegediagnose stehen?
- Welche pflegerischen Zielsetzungen stehen im Fokus?
- Können die Handlungskonzepte zur Behebung und/oder Linderung eines pflegerischen Problems einer Pflegediagnose in der Literatur bestätigt werden?
- Welcher unter anderem zeitliche Ressourcenverbrauch ist bei einer neuen Pflegemaßnahme anzusetzen und welche Verknüpfungen (Mappings) mit anderen Systemen und anderen Instrumenten sind erforderlich?
- …

Ein zentrales Kernelement zur Beantwortung dieser Fragen ist in der Planung und Durchführung einer internationalen, systematischen Literaturrecherche und -bewertung zu sehen. Abhängig von der jeweiligen Fragestellung, die bei der Entwicklung und/oder Aktualisierung der ENP-Praxisleitlinie im Vordergrund steht, wird im ersten Schritt eine umfassende Suchstrategie entwickelt, mit der im Verlauf weitreichend, und möglichst ohne „blinde Flecken“ zu hinterlassen, nach relevanten Publikationen gesucht wird. Gleichzeitig soll es eine sorgsam durchdachte Recherchestrategie ermöglichen, adäquat und vorbereitet auf unterschiedliche Eventualitäten im Verlauf der Literatursuche reagieren zu können, etwa hinsichtlich Qualität (z. B. zu viele thematisch abweichende Suchergebnisse) aber auch Quantität der Suchergebnisse bzw. Trefferzahlen (z. B. offenkundig zu wenig Suchergebnisse bei zu enger Recherchestrategie bzw. zu viele und nicht mehr handhabbare Trefferzahlen bei zu offener Recherchestrategie). Die nachfolgenden Kapitel widmen sich ausführlich sowohl der Entwicklung einer Suchstrategie, der eigentlichen Durchführung der Literaturrecherche sowie insbesondere auch der individuellen Evidenzbewertung der recherchierten Publikationen im Sinne eines Critical Appraisal in Abhängigkeit von der Art der Publikation bzw. des zugrunde liegenden Forschungsdesigns.

2.1 Strategie und Kernelemente der systematischen, internationalen Literatursuche

Im Rahmen der ENP-Weiterentwicklung werden die klassischen Schritte einer systematischen, internationalen Literatursuche befolgt (vgl. z.B. Mayer, Raphaelis, & Kobleder, 2021; Nordhausen & Hirt, 2020) und nach bestem Gewissen umgesetzt. Diese lassen sich wie folgt zusammenfassen:

1. Eingrenzung des Themas
2. Formulierung möglichst konkreter und spezifischer Forschungsfragen bzw. -hypothesen
3. Ableitung geeigneter Suchbegriffe aus den Forschungsfragen und Vermehrung derselben (z. B. Synonyme, englische Übersetzungen etc.)

4. Festlegen der Ein- und Ausschlusskriterien (z. B. Ein-/Ausschluss bestimmter Studiendesigns, Eingrenzung von Veröffentlichungszeiträumen und Sprachen der Publikationen etc.)
5. Definieren und Bestimmen der für die Recherche zu nutzenden Fachdatenbanken und weiterer Anlaufstellen zur Recherche
6. Dokumentation der Suchstrategie sowie des gesamten Rechercheverlaufs
7. Durchführung der Literaturrecherche
8. Sichtung und Bewertung der Suchergebnisse (Abstract Screening)
9. Beschaffung und Management der relevanten Literatur
10. Critical Appraisal der Volltexte (vgl. Kapitel 2.2)

Alle Schritte und Entscheidungen werden dokumentiert und an wichtigen Stellen im Rechercheprozess im Entwicklerteam konsentiert. Hierzu wurde ein Protokoll für die Dokumentation einer systematischen Literaturrecherche erstellt (siehe Anlage III), das bei jeder Umsetzung einer systematischen Recherche zu führen ist. Je nach Zahl und Qualität der Suchergebnisse wird die Suchstrategie beim Abstract Screening mit dem Ziel, ein handhabbares und gleichzeitig möglichst umfassendes Ergebnis zu erreichen verfeinert. Der prozesshafte Verlauf einer systematischen Literaturrecherche kann im Verlauf also auch wieder um einige Schritte zurückgeworfen werden. Bei kleineren, punktuellen Fragestellungen wie beispielsweise der Überprüfung einer möglichen Ergänzung eines Kennzeichens zu einer Pflegediagnose kann auch entschieden werden, dass eine „kleine" Literatursuche durchgeführt wird. In diesem Fall kann das Protokoll zur systematischen Literaturrecherche vernachlässigt und die gefundenen Quellen entsprechend den in Kapitel 2.2 beschriebenen Kriterien bewertet werden. Alle am Entwicklungsprozess beteiligten Personen haben eine entweder pflegepraktische oder therapeutische Ausbildung sowie einen Abschluss in mindestens einem pflegebezogenen bzw. therapeutischen Studiengang.

2.1.1 Die Fragestellung(en)

Die Fragestellung spielt bei der Evidenzsuche und Datenextraktion eine entscheidende Rolle. Je mehr Klarheit darüber herrscht, wonach gesucht werden soll, desto erfolgreicher verläuft in der Regel die Recherche. Ausgangspunkt für den Forschungsprozess zur Erstellung bzw. Aktualisierung einer ENP-Praxisleitlinie und/oder Elementen derselben ist daher die möglichst explizite und präzise Formulierung einer oder mehrerer Fragestellungen. Probate Mittel zur Einengung eines Themas bzw. zur Konkretisierung einer Fragestellung bestehen etwa in einer orientierenden Literaturrecherche (Einlesen in einen Themenbereich mit einem zwar gezielten, aber nur begrenzt systematischen Vorgehen) sowie in der Spezifizierung verschiedener Rahmenbedingungen bzw. Faktoren, in die das Thema von Interesse eingebettet ist. Entsprechende Aspekte können beispielsweise sein:

- Die Zielgruppe bzw. der Personenkreis von Interesse
- Das Setting
- Der Kontext
- Die (Pflege-) Handlung, das Phänomen, die Situation (...) von Interesse
- Klassisch: die W-Fragen: Wer, wann, was, wie, warum?
- Die Abgrenzung zu benachbarten bzw. artverwandten Themengebieten

Grundsätzlich ist bei der Formulierung einer Fragestellung zu berücksichtigen, dass sie einerseits beantwortbar sein muss, dabei aber in ihrem Aufbau nicht bereits ein (vermutetes) Ergebnis vorwegnehmen darf. Auch ist unbedingt darauf zu achten, dass eine Fragestellung keine Vorannahmen zum Ausdruck bringt und immer so gut wie möglich ergebnisoffen formuliert sein sollte, um nicht (unbewusst) auf die Vorannahme hinzuarbeiten. So wäre beispielsweise die Fragestellung „Welches der Pflegekonzepte A oder B wirkt sich vorteilhafter auf die Lebensqualität von Menschen mit Demenz aus" der Fragestellung „Wirkt sich Pflegeintervention A tatsächlich besser als Pflegeintervention B auf die Lebensqualität von Menschen mit Demenz aus" vorzuziehen.

Insbesondere im klinischen Kontext und bei Fragestellungen rund um pflegerische Interventionskonzepte auch innerhalb ENP bietet sich das in der evidenzbasierten Medizin bereits seit langem fest etablierte PICO-Schema[1] zur Strukturierung und Konkretisierung von Fragestellungen an (Behrens & Langer, 2004; Panfil, Ivanović, & Conrad, 2011). PICO steht als Akronym für die Parameter „**P**atient / Population" – „**I**ntervention" – „**C**ontrol / Comparison" – „**O**utcome". Nach diesem Prinzip sollte eine Fragestellung Aspekte zu all diesen Parametern abfragen:

- P (Population / Patient): Beschreibung der Zielgruppe
- I (Intervention): Technologie, Verfahren oder Maßnahme als Gegenstand der Untersuchung bzw. Recherche
- C (Control / Comparison): Alternative, mit der die Technologie, das Verfahren oder die Maßnahme verglichen werden kann
- O (Outcome): Zielgröße. Was soll erreicht werden und womit kann das Ergebnis gemessen werden?

Stehen bei der ENP-Entwicklung pflegerische Interventionskonzepte im Fokus des Interesses, so ist das PICO-Schema grundsätzlich in unveränderter Form anwendbar. Der pflegediagnostische Kontext ist jedoch in jedem Fall mit zu berücksichtigen, wie nachstehende Beispiele illustrieren sollen:

	Fragestellung
Population / Pflegeempfänger(in) / pflegediagnostischer Kontext	Haben an Demenz erkrankte Patient(innen) im Krankenhaus
Intervention	durch die Einrichtung eines Nacht-Cafés
Kontrolle	im Vergleich zu Kliniken ohne ein solches Angebot
Ergebnis	veränderte Resultate beim Mini Mental Status Test (MMST)?

Tabelle 1: Beispiel 1 für eine Fragestellung zu Pflegeinterventionen nach dem PICO-Schema

	Fragestellung
Population / Pflegeempfänger(in) / pflegediagnostischer Kontext	Erfahren Pflegeempfänger(innen), die an einer Fatigue leiden
Intervention	durch komplementäre Pflegemaßnahmen der Akupressur
Kontrolle	im Vergleich zu keinen oder anderen komplementären Pflegemaßnahmen
Ergebnis	eine Veränderung der subjektiv empfundenen Müdigkeit?

Tabelle 2: Beispiel 2 für eine Fragestellung zu Pflegeinterventionen nach dem PICO-Schema

Da sich die Neu- bzw. Weiterentwicklung von ENP-Praxisleitlinien jedoch nicht ausschließlich auf pflegerische Interventionskonzepte bezieht, sondern auch die pflegediagnostischen Komponenten in Form der Diagnose selbst sowie zugehöriger Kennzeichen, Ursachen und Ressourcen einen wesentlichen Fokus darstellen, kann das PICO-Schema dahingehend adaptiert werden, dass es auch als Hilfestellung für die Erarbeitung von Fragestellungen in diesem Kontext anwendbar ist, wie folgendes Beispiel zeigt:

	Fragestellung
Population / Pflegeempfänger(in) / pflegediagnostischer Kontext	Pflegeempfänger(innen), die an einer Fatigue leiden
Beobachtbare / messbare Kennzeichen und Symptome	weisen welche messbaren, beobachtbaren bzw. berichteten Kennzeichen und Symptome auf
Ergebnis	die eine eindeutige Verifizierung des Vorliegens der Pflegediagnose „...leidet an einer Fatigue" erlauben?

Tabelle 3: Beispiel für eine Fragestellung zur Recherche nach Kennzeichen für eine Pflegediagnose in Anlehnung an das PICO-Schema

[1] Im deutschsprachigen Raum auch synonym als PIKE-Schema bezeichnet als Akronym für die deutschsprachigen Begriffe Patient, Intervention, Kontrollintervention und Ergebnis.

2.1.2 Ableitung von Suchbegriffen

Eine möglichst konkrete und präzise formulierte Fragestellung bildet eine ideale Ausgangsbasis für die Ableitung von Suchbegriffen bzw. Schlagwörtern, die bei einer Datenbankrecherche genutzt werden können. In der Regel lässt sich eine Fragestellung in mehrere inhaltliche Facetten bzw. Kernkonzepte aufteilen, die mit Blick auf die abgeleiteten Suchbegriffe auch als individuelle Kernkonzepte erfasst werden sollten. Ein Beispiel:

„Wie wirkt sich die Anwendung von ***Snoezelen*** im Vergleich zur ***integrativen Validation*** auf das Auftreten von ***herausfordernden Verhaltensweisen*** bei ***dementen Pflegeheimbewohner(inne)n*** aus?“

Als Kernkonzepte und zentrale Suchbegriffe dieser beispielhaften Fragestellung lassen sich identifizieren und ableiten:

- Pflegeheimbewohner(innen) mit Demenz (Population)
- Snoezelen (Intervention)
- Integrative Validation (Kontrollintervention)
- Herausfordernde Verhaltensweisen (Ergebnis)

Die Vermehrung der zentralen Suchbegriffe zu jedem Kernkonzept der Forschungsfrage erhöht die Aussicht auf relevante Treffer meist deutlich und reduziert gleichzeitig die Gefahr, „blinde Flecken“ bei der Durchführung der Literaturrecherche zu hinterlassen! Relevante Stellschrauben sind hierbei insbesondere Synonyme, Ober- und Unterbegriffe sowie englische Übersetzungen und deren Synonyme. Auch die Beachtung verschiedener Flexionsformen. Alle Kernelemente/Kernkonzepte der Fragestellung(en) sollten nach diesem Vorgehen zunächst für sich betrachtet werden, um die resultierenden Suchbegriffe im Anschluss sinnvoll zu Suchphrasen verbinden zu können. Als praktisches Hilfsmittel zur Identifikation und Vermehrung geeigneter Suchbegriffe und Schlagworte hat sich die Anlage einer tabellarischen Matrix bewährt. Im nachfolgenden, die eben vorgestellte Fragestellung aufgreifenden Beispiel sind die Felder nur exemplarisch und nicht vollständig ausgefüllt.

	Kernkonzept 1 Pflegeheim-bewohner(in)	**Kernkonzept 2 Herausfordern des Verhalten**	**Kernkonzept 3 Demenz**	**Kernkonzept 4 (Integrative) Validation**	**Kernkonzept 5 Snoezelen**
Oberbegriff	Pflegeheim, Pflege-bedürftige	Verhalten	Kognitive Beeinträchti-gung	Kommunikation	Sensorische Stimulation, Umgebungs-gestaltung
Unterbegriff	--	Wandering, Aggressivität, ...	(Morbus) Alzheimer	Akzeptanz	--
Synonyme	--	Verhaltens-auffälligkeit	--	--	--
englische Übersetzung	resident, nursing home, ...	challenging behavior, defiantness, ...	dementia, Alzheimer disease, mental disorder, ...	validation, ...	Snoezelen, snoezeling

Tabelle 4: Beispiel einer tabellarischen Matrix zur Identifikation und Vermehrung von relevanten Suchbegriffen

2.1.3 Ein- und Ausschlusskriterien

Die formulierten Ein- und Ausschlusskriterien zu den Fragestellungen sind bestimmend für die Datenextraktion. Ziel dabei ist die Klärung des formalen Rahmens der Recherche und Fokussierung auf das, was wirklich von Interesse und praktikabel ist (zusätzliche Filterung). Mögliche Kriterien hierbei sind

unter anderem: Die Sprache der Publikationen, der Erscheinungszeitraum, verschiedene Publikations-/ Studientypen von Interesse, die Fokussierung auf bestimmte Altersgruppen, Geschlechter, Settings und /oder Professionen sowie die Auswahl der zu nutzenden Fachdatenbanken und weiterer Suchorte. Ein wichtiges Kriterium bei der Auswahl ist, dass alle genutzten Ein- bzw. Ausschlusskriterien begründbar sein müssen und mit der entsprechenden Argumentation entsprechend dokumentiert werden. Für das angeführte Beispiel könnten sinnvolle Ein- bzw. Ausschlusskriterien etwa wie folgt lauten.

Einschlusskriterien	Ausschlusskriterien
Altersgruppe: > 65 Jahre (Grund: präsenile Demenzen sind sehr selten)	Publikationstypen: Ausschluss von Erfahrungsberichten, Einzelfallstudien u.Ä. (Grund: nur wenig Aussagekraft hinsichtlich Wirksamkeit erwartbar)
Erscheinungszeitraum der Publikationen: ab 1990 (Grund: weder Validation noch Snoezelen wurden vorher in nennenswertem Rahmen angewandt)	
Sprachen der Publikationen: Deutsch und Englisch (Grund: professionelle Übersetzungen sind sehr teuer und zeitaufwendig)	

Tabelle 5: Beispielhafte Darstellung von Ein- und Ausschlusskriterien

Die endgültige Formulierung von Ein- und Ausschlusskriterien ist oftmals erst im Verlauf der Durchführung der Literaturrecherche möglich, da vielfach erst hier thematische bzw. formale Aspekte auftauchen, welche die Frage nach einer Berücksichtigung oder einem Ausschluss erneut aufwerfen können.

2.1.4 Definition der Fachdatenbanken und weiterer Anlaufstellen zur Recherche

Ein zentraler Aspekt bei der Erarbeitung einer Suchstrategie ist die Frage danach, „wo“ mit möglichst großen Erfolgsaussichten nach Literatur recherchiert wird. Neben klassischen Anlaufstellen wie etwa Bibliothekskatalogen oder Bibliografien steht insbesondere mit Blick auf die Recherche über das Internet eine enorme und nahezu unerschöpfliche Vielfalt an Instrumenten und Optionen zur Recherche zur Verfügung. Naheliegend ist der Kontext des wissenschaftlichen Arbeitens, dass einer (meist unstrukturierten) Recherche in allgemeinen Suchmaschinen wie Google hier eine wesentlich niedrigere Bedeutung zukommt als etwa der elektronischen Recherche in einer Fachdatenbank. Folgende Abbildung stellt die wesentlichsten Möglichkeiten und Anlaufstellen zur systematischen Literaturrecherche zusammenfassend vor:

Abbildung 2: Wesentliche Anlaufstellen und Möglichkeiten zur systematischen Literaturrecherche (Quelle: eigene Darstellung in Anlehnung an Simon, 2018)

Die im Rahmen der ENP-Entwicklung bedeutsamsten Anlaufpunkte zur Literaturrecherche sind elektronische Fachdatenbanken wie beispielsweise PubMed/MEDLINE, CINAHL (Cumulative Index to Nursing and Allied Health Literature), die Cochrane Library, GeroLit oder PsycNet und viele weitere mehr. Daneben stellen pflegerische und medizinische Leitlinien einschlägiger Verbände, Gesellschaften und Herausgeber sehr bedeutsame Wissensquellen für die ENP-Entwicklung dar, da es sich hierbei um systematisch entwickelte Übersichtsarbeiten handelt, die in einem aufwändigen Verfahren erarbeitet und geprüft werden sowie evidenzbasierte Empfehlungen zur Verfügung stellen. Sie haben keinen bindenden bzw. verpflichtenden Charakter, tragen jedoch in kompakter Form den Wissensstand zu einer meist sehr spezifischen Fragestellung zusammen und bieten evidenzbasierte Lösungswege zum jeweiligen Thema an. Beispiele für relevante Bezugsquellen von Leitlinien sind etwa die Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften (AWMF), die Leitlinien- und Standarddatenbank des Zentrums für Qualität in der Pflege (ZQP), das National Institute for Health and Clinical Excellence (NICE), die Nursing Best Practice Guidelines der Registered Nurses' Association of Ontario (RNAO) oder das Scottish Intercollegiate Guideline Network (SIGN).

Die systematische Literatursuche in Fachdatenbanken, Leitlinienkatalogen, Bibliotheken und weiteren strukturierten Anlaufpunkten kann im Bedarfsfall um die Recherche auf weiteren Wegen ergänzt werden. Beispielhaft ist hier etwa „Google Scholar“ zu nennen, ein spezielles Derivat der weit verbreiteten bekannten Suchmaschine Google, die der allgemeinen Recherche nach wissenschaftlichen Dokumenten gilt. Auch das sog. „Schneeballsystem“, im Englischen footnote chasing genannt, kann eine sinnvolle Ergänzung der Recherchestrategie sein. Hierunter ist die Suche in den Literaturangaben neu recherchierter oder bereits bekannter Fachpublikationen zu verstehen. Insbesondere bei systematischen Übersichtsarbeiten kann dieses Vorgehen erfolgversprechend sein, um weitere für die eigene Fragestellung relevante Publikationen oder auch neue Suchbegriffe für weitere Recherchedurchläufe aufzuspüren.

Im Rahmen der Dokumentation der systematischen Literaturrecherche sollten folgende Punkte zur Auswahl der Fachdatenbanken und weiterer Anlaufstellen auf dem Rechercheprotokoll festgehalten werden:

- Die Auswahl elektronischer Fachdatenbanken mit kurzer Begründung
- Die Auswahl von Bibliothekskatalogen / Freihandaufstellungen mit kurzer Begründung
- Die Auswahl von Leitlinienkatalogen (z. B. AWMF etc.) mit kurzer Begründung
- Die Auswahl weiterer Rechercheorte (z. B. Fachzeitschriften, graue Literatur, Suchmaschinen im Internet, Schneeballsystem etc.) mit kurzer Begründung

Thieme RECOM

2.1.5 Entwicklung von Suchphrasen und Einflussmöglichkeiten auf den Rechercheverlauf

Ein wesentlicher Schritt, um die Durchführung der systematischen Literaturrecherche möglichst zielgerichtet, handhabbar und vollständig ohne „blinde Flecken“ zu gestalten, aber auch, um auf oft nicht exakt vorhersehbare Suchergebnisse sowohl hinsichtlich Qualität als auch hinsichtlich Quantität adäquat reagieren zu können (zu viele Treffer, zu wenige Treffer, die meisten Ergebnisse am Thema vorbei, ...), ist die Generierung von verschiedenen Suchphrasen. In diesem Zuge ist es ratsam, neben den mutmaßlich „eigentlichen“ Suchphrasen auch Alternativen mitzudenken, um die Suchläufe z. B. einzuengen oder zu erweitern. Zur Entwicklung von Suchphrasen sollten folgende Techniken bzw. Aspekte berücksichtigt werden:

- **Feldsuche** (Nutzung des gegliederten Aufbaus der bibliographischen Angaben in einer Datenbank, z. B. Suche nach Autor, nach Titel, nach Abstract, ...)
- **Filterung der Suche**, im Wesentlichen nach den Ein- / Ausschlusskriterien (z. B. Veröffentlichungszeitraum, Sprache, Altersgruppen, ...)
- Nutzung der **Schlagwortsuche (bzw. MeSH-Terms[2])**, die eine Publikation charakterisieren bzw. die schnelle Erschließung der Inhalte einer Publikation ermöglichen. Die Schlagworte selbst müssen dabei nicht zwingend im eigentlichen Text der Publikation enthalten sein.
- **Trunkierung und Maskierung**, bedeutet die Nutzung von Platzhaltern, um die Eingabe aller möglichen grammatischen Varianten eines Suchbegriffs bzw. von Wortkombinationen zu vermeiden. Die Trunkierung (häufig mit dem Zeichen *) sucht nach Wortstämmen mit beliebiger Endung bzw. Worterweiterung, die Maskierung ersetzt hingegen ein einziges Zeichen innerhalb eines Suchbegriffs (häufig mit dem Zeichen # oder ? umgesetzt)
- **Boolesche Operatoren**, welche die zielgerichtete Verknüpfung der Suchbegriffe miteinander ermöglichen. Die Booleschen Operatoren sind eine der mächtigsten Funktionen bei der Durchführung einer systematischen Literaturrecherche, weswegen deren Einsatz lieber doppelt als nur einfach geprüft werden sollte (schnelle Fehlerquelle). Die wichtigsten Operatoren sind:
    - **AND** (selten auch das deutsche UND. Zeigt nur Treffer, die alle derart verknüpften Suchbegriffe enthalten – „sowohl als auch“)
    - **OR** (selten auch das deutsche ODER. Zeigt Treffer, die entweder den Suchbegriff A oder den Suchbegriff B oder den Suchbegriff C (...) enthalten – „entweder oder“)
    - **NOT** (selten auch das deutsche NICHT. Schließt alle Treffer aus den Suchergebnissen aus, die den oder die entsprechenden Begriffe enthalten – „auf keinen Fall“).

Abbildung 3: Die Booleschen Operatoren und Ihre Auswirkungen bei der Literaturrecherche (Quelle: eigene Darstellung)

- Die **Phrasensuche**, mit der die eingegebenen Suchbegriffe in ausschließlich der angegebenen Reihenfolge gesucht und berücksichtigt werden. Oft sehr hilfreich bei der Suche nach Konzepten mit einer feststehenden Kombination aus mehreren Wörtern (z. B. „pflegende Angehörige“). Oft wird die Phrasensuche durch das Setzen der Suchbegriffe in Anführungszeichen („“) umgesetzt oder durch die Aktivierung einer entsprechenden Option in der Suchmaske der Datenbank (z. B. Funktion „exakte Wortfolge“).

[2] MeSH steht als Akronym für „Medical Subject Headings” und beschreibt einen Thesaurus zur Erfassung von Zeitschriftenartikeln und Büchern in der Medizin sowie den Biowissenschaften. Er kommt vor allem bei der Datenbank MEDLINE (PubMed) zum Einsatz.

In vielen Fällen ist es zielführend, die Erstellung von Suchphrasen zur Recherche in Fachdatenbanken und Katalogen zunächst auf die einzelnen Kernkonzepte der Forschungsfrage aufzuteilen, um sie im Rahmen der letzten Suchdurchgänge schließlich zusammenzubringen. Zur Konstruktion der Teilsuchphrasen wie auch der abschließenden Gesamtsuchphrase bietet sich abermals die Nutzung einer Matrix an, in der sowohl die jeweiligen Suchbegriffe als auch die verschiedenen Recherchetechniken wie Feldsuche, Trunkierung und die Nutzung der Booleschen Operatoren zusammengebracht werden können. Nachfolgende Matrizes greifen die Konstruktion von Teil- bzw. einer Gesamtsuchphrase beispielhaft und unvollständig anhand der bisher behandelten Fragestellung auf.

Teilsuchphrase zu Kernkonzept 1 Pflegeheimbewohner(in)			
Verknüpfender Operator (AND, OR)	Suchbegriff (ggf. trunkiert, maskiert oder als Phrase)	Feldsuche? (Titel, MeSH, ...)	Alternativen?
	Hauptbegriff resident	ggf. Feldsuche in Titel und Abstract	
OR	Synonym 1 „nursing home“	ggf. Feldsuche in Titel und Abstract	
OR	Synonym 2 „elder care“	ggf. Feldsuche in Titel und Abstract	„elderly care“, „geriatric care“
OR	Synonym 3		
Bei diesem Kernkonzept auszuschließende Begriffe?			
NOT	Hospital OR „outpatient care“ OR “home care”		

Tabelle 6: Konstruktion einer Teilsuchphrase zu einem Kernkonzept der Fragestellung

Teilsuchphrase zu Kernkonzept 2 Herausforderndes Verhalten			
Verknüpfender Operator (AND, OR)	Suchbegriff (ggf. trunkiert, maskiert oder als Phrase)	Feldsuche? (Titel, MeSH, ...)	Alternativen?
	Hauptbegriff „challenging behavior“	ggf. Feldsuche in Titel und Abstract	
OR	Synonym 1 defiantness	ggf. Feldsuche in Titel und Abstract	
OR	Synonym 2 wandering	ggf. Feldsuche in Titel und Abstract	
OR	Synonym 3 aggressi*		
Bei diesem Kernkonzept auszuschließende Begriffe?			
NOT			

Tabelle 7: Konstruktion einer Teilsuchphrase zu einem weiteren Kernkonzept der Fragestellung

Sind zu allen Kernkonzepten der Fragestellung Suchphrasen und mögliche Alternativen erstellt, gilt es, alle Facetten zu einer oder mehreren alternativen Gesamtsuchphrasen zusammenzuführen. Hierzu lässt sich oftmals die Suchhistorie in den einzelnen Fachdatenbanken nutzen (z. B. kombiniere „Suchlauf zu Kernkonzept 1“ mit dem Operator „AND“ mit dem „Suchlauf nach Kernkonzept 2“). Im Sinne einer nachvollziehbaren Dokumentation sowie der oft gegebenen Notwendigkeit, die finalen Suchläufe anderweitig darzulegen, sollen auch alle „Gesamtsuchen“ sowie alternative Suchdurchgänge (falls das Suchergebnis zu breit, zu eng oder thematisch unangemessen ist) soweit möglich im Vorfeld überlegt und tabellarisch dokumentiert werden. Bei der Kombination der Teilsuchphrasen zu einem oder mehreren Gesamtsuchläufen kommt in aller Regel auch die Anwendung der Ein- und Ausschlusskriterien zum Einsatz.

Gesamtsuchphrase – Variante 1

	Verknüpfender Operator (AND, OR, NOT)		Verknüpfender Operator (AND, OR, NOT)	
Teilsuchphrase zu Kernkonzept 1	*AND*	*Teilsuchphrase zu Kernkonzept 2*	*AND*	*Teilsuchphrase zu Kernkonzept 3*
Bei der Gesamtsuchphrase auszuschließende Begriffe?				
			NOT	
Zu nutzende Suchfilter (Ein- / Ausschlusskriterien)				
Publikationszeitraum	Sprache	Studien- / Publikationsdesigns	Altersgruppe	...
Gesamtsuchphrase – Variante 2				
	Verknüpfender Operator (AND, OR, NOT)		Verknüpfender Operator (AND, OR, NOT)	
Teilsuchphrase zu Kernkonzept 1	*AND*	*Teilsuchphrase zu Kernkonzept 2*	*AND*	*Teilsuchphrase zu Kernkonzept 3*
Bei der Gesamtsuchphrase auszuschließende Begriffe?				
			NOT	
Zu nutzende Suchfilter (Ein- / Ausschlusskriterien)				
Publikationszeitraum	Sprache	Studien- / Publikationsdesigns	Altersgruppe	...

Tabelle 8: Beispiel für eine Matrix zur Erstellung von Suchphrasen zur Zusammenführung der Suchläufe zu einzelnen Kernkonzepten der Fragestellung

Eine sinnvolle optionale oder ggf. ergänzende Alternative zur tabellarischen Darstellung der Suchphrase(n) ist ggf. die rein textliche Umsetzung bzw. der rein textliche Entwurf derselben in Form einer datenbankspezifischen Syntax. Hierbei ist wichtig, ähnlich wie in der Mathematik eine akkurate Klammersetzung sowie die Besonderheiten und Techniken der jeweiligen Datenbank zu berücksichtigen. Klammern drücken dabei den Vorrang in der Bearbeitung der Suchphrase durch die Datenbank aus, Elemente innerhalb einer Klammer werden zunächst abgefragt, anschließend die Elemente außerhalb einer Klammer. In der Regel ist auch die Verschachtelung mehrerer Klammerpaare möglich. Sollen einzelne Begriffe in spezifischen Suchfeldern (z. B. nur im Titel, nur im Abstract, ...) gesucht werden, so wird dies je nach verwendeter Fachdatenbank unterschiedlich ausgedrückt, oft werden hierzu eckige Klammern [] mit dem entsprechenden Suchfeld genutzt. Bei der manuellen Zusammenstellung von Suchphrasen über eine Syntax ist grundsätzlich große Achtsamkeit und idealerweise Überprüfung durch eine zweite Person erforderlich, da sich auf diesem Weg sehr leicht Fehler mit massiven Auswirkungen auf die Suchergebnisse einschleichen können. Zur Illustration ein kurzes Beispiel einer Recherchesyntax mit Bezug auf die beispielhafte Fragestellung:

Abbildung 4: Textuelle Erstellung einer mehrere Kernkonzepte verbindenden Suchphrase in Form einer datenbankspezifischen Syntax

Zur Nachvollziehbarkeit des Rechercheverlaufs inklusive der jeweiligen Änderungen an Quantität und Qualität der Suchergebnisse in verschiedenen Fachdatenbanken (und ggf. mittels anderer Suchorte) empfiehlt es sich, auch die Historie aller durchgeführten Suchläufe zu dokumentieren. Hierzu bieten alle größeren Fachdatenbanken und Kataloge die Möglichkeit, den Sucherverlauf einzusehen und zu kopieren (beispielhaft vgl. Screenshot) oder auch für eine spätere Nutzung abzuspeichern. Letzteres erfordert in der Regel die Anlage eines meist kostenfreien Benutzerkontos bei der jeweiligen Fachdatenbank.

History and Search Details — Download — Delete

Search	Actions	Details	Query	Results	Time
#2	•••	>	Search: **(dementia[title/abstract] OR Alzheimer [title/abstract] OR "cognitive impairment"[title/abstract]) AND ("nursing home" OR eldercare OR "geriatric nursing") NOT hospital***	4,094	06:04:31
#1	•••	>	Search: **dementia**	242,516	03:33:48

Abbildung 5: Beispiel für die Aufzeichnung des Rechercheverlaufs, hier in der Fachdatenbank PubMed (https://pubmed.ncbi.nlm.nih.gov/)

2.1.6 Abstract Screening und Volltextbeschaffung

Ein entsprechendes Nutzerkonto bei den einschlägigen Fachdatenbanken kann sich auch als sehr hilfreiches Verwaltungswerkzeug erweisen, wenn es darum geht, nach der erfolgreichen Zusammenstellung der finalen Suchphrasen die Ergebnisse zu sichten und zu entscheiden, welche der Suchergebnisse im Volltext zu beschaffen und im weiteren Verlauf tiefergehend zu analysieren und zu bewerten sind. Das Mittel der Wahl zur Entscheidung diesbezüglich ist in den allermeisten Fällen ein sog. „Abstract Screening“.

Mittels dieses Schrittes der systematischen Literaturrecherche werden alle Suchergebnisse der finalen Recherchedurchläufe daraufhin geprüft, inwieweit sie die inhaltlichen und sonstigen Einschlusskriterien erfüllen. Hierzu wird für alle Suchtreffer die meistens unmittelbar in der Fachdatenbank einsehbare Zusammenfassung/Kurzfassung der jeweiligen Publikation – der sog. Abstract – nach einem vorab zu bestimmenden Muster bzw. Fragenkatalog entsprechend den definierten Ein- / Ausschlusskriterien studiert und bewertet. Alle Publikationen, welche die Einschlusskriterien erfüllen bzw. bei denen allein aufgrund des Abstracts keine sichere Entscheidung zu treffen ist, werden im Anschluss nach Möglichkeit im Volltext beschafft. Alle anderen Suchergebnisse, die bereits während des Screenings sicher und begründet ausgeschlossen werden können, erfahren im Verlauf keine weitere Berücksichtigung. Wann

immer möglich sollte das Abstract-Screening durch zwei Personen in unabhängiger Art und Weise erfolgen, sodass die Screening-Ergebnisse im Anschluss verglichen und im Bedarfsfall bei großen Abweichungen diskutiert werden können. Sowohl Ausschlüsse als auch Einschlüsse aus den Suchergebnissen werden mit Angabe einer kurzen Begründung auf dem Rechercheprotokoll festgehalten.

Neben der textuellen Darstellung[3], wie viele Treffer initial durch Recherche in verschiedenen Fachdatenbanken und andere Anlaufstellen gefunden wurden, bei wie vielen Treffern es sich um Duplikate handelt und wie viele Suchergebnisse durch das Abstract Screening sowie die sich anschließende Volltextanalyse ausgeschlossen wurden, empfiehlt es sich zudem, entsprechend etwa den Empfehlungen des PRISMA-Statements (Moher et al., 2011; Ziegler & König, 2011), ein Flussdiagramm zur Beschreibung der verschiedenen Phasen der Literaturrecherche bzw. der Phasen des Ausschlusses von Sucherergebnissen/Publikationen anzufertigen. Nachfolgende Abbildung stellt ein solches Diagramm exemplarisch dar.

Abbildung 6: Flussdiagramm zur Darstellung des Ausschlusses von Publikationen im Rechercheprozess (Quelle: eigene Darstellung in Anlehnung an Ziegler & König, 2011, S. e11)

Zur Beschaffung potenziell relevanter Volltexte wird neben den verfügbaren (internetbasierten) Zugängen zu Bibliotheken, Online-Ausgaben von Zeitschriften und verschiedenen Herausgebern/Verlagen im Bedarfsfall auch auf Dokumentenlieferdienste wie beispielsweise Subito[4] zurückgegriffen. Die Verwaltung und das Management aller Volltexte sowie sämtlicher bibliographischer Angaben erfolgt mit dem Literaturverwaltungsprogramm EndNote.

[3] Z. B.: „Von insgesamt <XX> Suchtreffern wurden insgesamt <YY> Publikationen als relevant für die Anschaffung im Volltext bewertet. Von diesen <YY> relevanten Publikationen konnten im Verlauf <ZZ> tatsächlich als Volltext beschafft werden“.
[4] Vgl. https://www.subito-doc.de/

2.2 Kriterien und Instrumente für die Evidenzbewertung, Evidenzselektion und Datenextraktion

Bei der (Weiter-) Entwicklung von ENP-Praxisleitlinien wird häufig analog dem Vorgehen der Leitlinienentwicklung der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) vorgegangen. Es wird in der Regel ein iterativer, hierarchischer Rechercheprozess für die Erstellung bzw. Überarbeitung von ENP-Praxisleitlinien gewählt (Schmucker et al., 2017). Das bedeutet, dass bevorzugt bereits „vorbewertete" Evidenz analog der sog. 6S-Hierarchie (Dicenso, Bayley, & Haynes, 2009) genutzt wird. Die 6S-Hierarchie vorbewerteter Evidenz sieht vor, dass nacheinander die Berücksichtigung a) von hochwertigen Leitlinien, von b) aggregierter Evidenz (systematische Übersichtsarbeiten) und c) Primärliteratur vorgenommen wird. Kann die Evidenz von pflegerischen Handlungskonzepten und/oder pflegediagnostischen Zusammenhängen nicht basierend auf hochwertigen Leitlinien/Übersichtsarbeiten gestützt werden, wird die entsprechende Primär- und weitere Fachliteratur herangezogen. Abhängig davon, um welche Art von Publikation bzw. welches zugrunde liegende Forschungsdesign es sich handelt, werden entsprechende Bewertungsverfahren (Critical Appraisal) zur Einstufung der Qualität genutzt. Selbige werden in den folgenden Kapiteln genauer vorgestellt.

2.2.1 Critical Appraisal von Publikationen im ENP-Entwicklungsprozess

Im Rahmen der ENP-Entwicklung wurden unterschiedliche Bewertungsformulare für das Critical Appraisal genutzt (vgl. Anlage II), die im Zeitraum von 2004 bis 2021 im Einsatz waren. Diese Bewertungsformulare wurden im Jahr 2021 von vollständig neu entwickelten Varianten abgelöst und bestimmen seither die (Weiter)Entwicklungsprozesse. Wesentliche Gründe für die Neuentwicklung der Bewertungsinstrumente im Entwicklungsprozess von ENP sind die vermehrt internationale Ausrichtung der neu gewählten Ansätze sowie zahlreiche neue Erkenntnisse in diesem Zusammenhang über den Zeitverlauf. Es gibt eine beträchtliche Anzahl an verschiedenen Bewertungsinstrumenten zur Einschätzung der Qualität von Studienarbeiten bzw. Veröffentlichungen (Browman et al., 2017; Cochrane Deutschland (Ed.) et al., 2021; Dreier et al., 2010; Ma et al., 2020; Platz 2018; Thole et al., 2007). In einem Bericht zu einem Health Technology Assessment (HTA) aus dem Jahr 2010 wurden für Interventions-, Beobachtungs- und Diagnosestudien sowie systematische Reviews und Meta-Analysen bereits 125 unterschiedliche Instrumente mit großen Varianzen hinsichtlich deren Güte identifiziert (Dreier et al., 2010). Die Verfasser(innen) beschreiben zudem unterschiedliche Operationalisierungen der Studienqualität in ihrer systematischen Literatursuche. Sie empfehlen bei der Bewertung, die methodische Qualität einer Forschungsarbeit von dem Risiko eines systematischen Verzerrungseffekts (Bias) sowohl vor, während als auch nach der Durchführung zu differenzieren und gesondert zu bewerten. Zudem ist eine explizite Trennung der Berichts- und Studienqualität anzuraten, da eine Vermischung zu Fehleinschätzungen führen kann. Die Beurteilung von Studien bezieht sich zu einem großen Teil auf die interne Validität, also die Gültigkeit der getroffenen Aussagen. Verschiedene Studientypen bzw. Forschungsdesigns sind dabei für unterschiedliche Fehler bezüglich des Risikos einer Verzerrung anfällig. Beispielsweise sind zur Evaluation der Wirksamkeit von Maßnahmen der Akupressur in der Pflege zur Reduktion von Übelkeit die Struktur- und Beobachtungsgleichheit zwischen der Interventions- und der Kontrollgruppe sehr bedeutend, während die Bewertung von systematischen Übersichtsarbeiten maßgeblich von der Qualität der ausgewählten Einzelstudien sowie der gewählten Ein- bzw. Ausschlusskriterien abhängt.

Art und Umfang der Studien, die im Rahmen der Neu- bzw. Weiterentwicklung von ENP-Praxisleitlinien genutzt werden, hängen von den jeweiligen Fragestellungen sowie den verfügbaren Forschungsarbeiten ab. Wird z. B. gezielt nach Belegen für einen Ursachenzusammenhang zu einer Pflegediagnose gesucht, ist die Fragestellung meist sehr „scharf & eng" formuliert. Die Rechercheergebnisse bei sehr eng formulierten Fragestellungen sind in der Regel weniger umfangreich, umgekehrt führen abstraktere Fragestellungen oftmals zu einer deutlich größeren Zahl an Suchergebnissen. Im Laufe des iterativ aufgesetzten Rechercheprozesses zur Weiter- bzw. Neuentwicklung einer ENP-Praxisleitlinie kann es sich daher ergeben,

dass eine weitere Literatursuche mit dem Fokus auf beispielsweise ein spezifisches Handlungskonzept zusätzlich aufgesetzt wird. Alle Publikationen, auf denen ein Item in ENP basiert, sind zuvor bezüglich ihrer Qualität bewertet worden. Der komplexe Prozess der kritischen Würdigung jeder genutzten Veröffentlichung wird abhängig von der Art der Referenz mit dem jeweils passenden Bewertungsinstrument realisiert. Abhängig von der jeweiligen Fragestellung bei der ENP-Entwicklung kommen unterschiedliche Studienarten bzw. Forschungsdesigns in Frage. Bei pflegerischen Interventionskonzepten haben etwa randomisiert kontrollierte Studien (englisch randomized controlled trial – RCT) eine hohe Relevanz, da hier bei adäquater Durchführung das geringste Verzerrungsrisiko besteht. Idealerweise sind zudem systematische Übersichtsarbeiten und/oder Metaanalysen verfügbar, die als Primärstudien RCTs nutzen. Bei den anderen Fragestellungen, wie etwa der Recherche nach Kennzeichen und/oder der Ätiologie zu einer Pflegediagnose oder auch zu Erfahrungen eines Pflegeempfängers werden andere Studientypen wie beispielsweise die verschiedenen Arten qualitativen Studiendesigns bedeutend sein. Für die allermeisten Studien- und Publikationstypen stehen wie bereits angesprochen verschiedenste Instrumente zum Critical Appraisal bereit. Nachfolgende Liste stellt die Auswahl der für die ENP-Entwicklung genutzten Instrumente zusammenfassend dar, bevor jeder einzelne Ansatz und ggf. auch deren Adaptionen für den internen Gebrauch genauer erläutert werden.

2.2.1.1 Systematische Reviews (SR) & Meta-Analysen – Bewertungssystematik nach AMSTAR 2 (Assessment of Multiple SysTemAtic Reviews)

2.2.1.2 Beobachtungsstudien (Fall-Kontroll-, Kohorten-, Querschnittsstudien) – Critical Appraisal Tools des JBI (Joanna Briggs Institute)

2.2.1.3 Randomisiert kontrollierte Studien (RCTs) – Bewertungssystematik in Anlehnung an die Cochrane-Empfehlungen nach Platz (2021)

2.2.1.4 Qualitative Forschungsarbeiten – Critical Appraisal Checklist for Qualitative Research: Bewertungssystematik des JBI für qualitative Forschung (z. B. für Grounded Theory, Inhaltsanalyse ...)

2.2.1.5 Leitlinien und Clinical Practice Guidelines – Bewertungssystematik in Anlehnung an das AGREE-II-Instrument

2.2.1.6 Mixed-Methods-Forschungsarbeiten – Bewertungssystematik (auszugsweise) des MMAT (Mixed Methods Appraisal Tool)

2.2.1.7 Fachbücher / Lehrbücher – Eigene Umsetzung der Bewertungskriterien der Bibliothek der Cornell Universität in Ithaca, New York

2.2.1.8 Graue Literatur und Webseiten – Bewertungskriterien der AACODS Checkliste (Authority, Accuracy, Coverage, Objectivity, Date, Significance)

Die für jede ENP-Praxisleitlinie relevante Literatur wird systematisch nach den vorgegebenen Qualitätskriterien in Abhängigkeit vom Studiendesign bzw. vom Publikationstyp bewertet. Der Einsatz von Checklisten unterstützt dabei den Bewertungsprozess im ENP-Entwicklerteam. Anschließend wird die Qualität der Evidenz nach einem Klassifikationsschema zusammenfassend angegeben (Evidenzgrad, Level of Evidence). Die Beschreibung des Aufbaus der Evidenztabelle und deren Nutzung findet im Anschluss in Abschnitt 2.2.3 statt.

Die genutzten Instrumente zum Critical Appraisal werden im Vorfeld vertiefend vorgestellt. Die konkrete Umsetzung der einzelnen Instrumente zum Critical Appraisal in Checklisten für die ENP-Weiterentwicklung ist in Anlage II einsehbar, ein beispielhaft befülltes Instrument stellt die nachstehende Abbildung vor:

Thieme RECOM

Critical Appraisal von randomisiert kontrollierten Studien (RCTs) gemäß Platz (2021)[1]

Publikation/ Studie: **Miranzadeh, S. et al.** (2014). A New mouthwash for Chemotherapy Induced Stomatitis. *Nursing and Midwifery Studies, 3*(3), pp. e20249. doi: 10.17795/nmsjournal20249

Reviewer 1: Wieteck, Pia

Reviewer 2: Kraus, Sebastian

Critical Appraisal von randomisiert kontrollierten Studien (RCTs) gemäß Platz (2021)	Ja	Nein	unklar	Nicht anwendbar
1. Wurden die Ein- und Ausschlusskriterien klar definiert? (externe Validität)	☒	☐	☐	☐
2. Gibt es eine klare Definition und ein adäquates Assessment der Zielgrößen (Outcomes) der Studie (klinische Relevanz)?	☒	☐	☐	☐
3. Wurden Nebenwirkungen und Akzeptabilität der Interventionen berichtet?	☒	☐	☐	☐
4. Fand eine adäquate „Follow-up"-Untersuchung statt (Erfassung von Langzeiteffekten)?	☐	☐	☐	☒
5. Wird eine klare Definition und Beschreibung der experimentellen und der Kontrollbedingungen gegeben (Studienkontrast)?	☒	☐	☐	☐
6. Werden die Teilnehmer randomisiert zugeteilt (Selektionsbias)?	☒	☐	☐	☐
7. Wurde die Zuordnung unter Verschluss gehalten („allocation concealment", Selektionsbias)?	☒	☐	☐	☐
8. Waren die Experimental- und Kontrollgruppen zu Beginn der Studie vergleichbar (Selektionsbias)?	☒	☐	☐	☐
9. Waren die Patienten sowie das Personal während der Behandlung verblindet, und waren außerhalb der untersuchten Effekte die Behandlungen in den randomisierten Gruppen vergleichbar (Durchführungsbias)?	☒	☐	☐	☐
10. Wurden die Studienergebnisse verblindet erhoben („detection bias")?	☒	☐	☐	☐
11. War die Berichterstellung nicht selektiv („reporting bias")?	☒	☐	☐	☐
12. Waren die Ergebnisdaten (fast) komplett („attrition bias")?	☒	☐	☐	☐
13. Wurde eine „Intention-to-treat"-Analyse berichtet („attrition bias")?	☒	☐	☐	☐
14. Unterstützen die Ergebnisse die berichteten Schlussfolgerungen ausreichend?	☒	☐	☐	☐

[1] **Platz, T.** (2021). Evidence-based Practice Guidelines for the German Society for Neurology (DGN) and the German Society for Neurorehabilitation (DGNR): Methods for systematic evidence-to-decision process (in German). *Fortschritte der Neurologie - Psychiatrie, 89*(09), pp. 415-423. doi: https://doi.org/10.1055/a-1309-1856

Abbildung 7: Beispiel für eine ausgefüllte Checkliste zum Critical Appraisal von Studien (hier: für randomisiert kontrollierte Studien)

2.2.1.1 Systematische Reviews (SR) & Meta-Analysen – Bewertungssystematik nach AMSTAR 2

In der Medizin werden bei der Auswahl geeigneter Therapien oder diagnostischer Verfahren zur Ermittlung der externen Evidenz häufig systematische Reviews als Grundlage herangezogen. Die genutzten Methoden und Verfahren eines systematischen Reviews sind längst etabliert. Dies gilt auch für eine ganze Reihe von Instrumenten, um die Güte von SRs zu bewerten. Für die Einordnung der Qualität von systematischen Übersichtsarbeiten gibt es zahlreiche Bewertungsverfahren (Zeng et al., 2015). Schmucker und Kolleg(inn)en (2017) führen alleine in ihrem Manual zur *„Bewertung des Verzerrungsrisikos von systematischen Übersichtsarbeiten"* 20 verschiedene Instrumente auf. Das AMSTAR (Assessment of Multiple SysTemAtic Reviews) zur Bewertung von systematischen, auf randomisiert kontrollierten Studien (RCTs) basierenden Übersichtsarbeiten wird als ein validiertes und am häufigsten verwendeten Instrument eingestuft, auch außerhalb des Anwendungsbereichs der Medizin (Lühnen, Berger-Höger, & Richter, 2021; Schmucker et al., 2017). Ebenso wurden sog. „reporting statements" für Systematic Reviews entwickelt, die verschiedene Kriterien zum Bericht bzw. zur Veröffentlichung einer systematischen Übersichtsarbeit bereitstellen. Ein häufig genutztes Beispiel hierfür ist etwa das PRISMA-Statement: „Preferred Reporting Items for Systematic reviews and Meta-Analyses" (Moher et al., 2011; Thangatorai, Lim, & Nalliah, 2018; Ziegler & König, 2011).

Die Güte des von einer kanadischen Arbeitsgruppe entwickelten AMSTAR ist in Studien belegt worden. Im Vergleich dazu sind die Messeigenschaften einer im Verlauf veröffentlichten revidierten Version des Instruments (R-AMSTAR) noch nicht hinreichend untersucht (Pieper et al., 2015). AMSTAR steht auch als Online-Anwendung zur Verfügung (Shea et al., 2017). Die 11 Leitfragen des ursprünglichen Instrumentes werden mit „ja", „nein", „unklar" oder „nicht anwendbar" eingestuft. Eine Gesamteinschätzung ist nicht vorgesehen. Die revidierte Fassung (R-AMSTAR) hinterlegt hingegen Score-Punkte zwischen 1 und 4 für die einzelnen Leitfragen mit dem Ziel, eine bessere Quantifizierung der Gesamtgüte der Publikation zu erreichen (Schmucker et al., 2017). Neben der revidierten Fassung existiert noch eine weitere überarbeitete Fassung, genannt AMSTAR 2. Im Fokus dieser Weiterentwicklung stand die Nutzungserweiterung von AMSTAR auch für Übersichtsarbeiten, die sowohl nicht-randomisierte Studien mit berücksichtigen und/oder auch Meta-Analysen (Shea et al., 2017). Da in der Pflege häufig nicht-randomisierte Studien vorliegen und infolgedessen Übersichtsarbeiten auch derartige Studien aufnehmen, scheint es in diesem Zusammenhang sinnvoll und zielführend, AMSTAR 2 zu verwenden. Gegenüber der ursprünglichen Variante umfasst AMSTAR 2 insgesamt 16 Items, die abhängig vom einzelnen Bewertungskriterium mit „ja", „teilweise erfüllt", „nein" oder „nicht anwendbar" zu bewerten sind. Die Checkliste mit den konkreten Ausführungen zur Auslegung kann kostenfrei im Internet abgerufen werden[5] (Shea et al., 2017). AMSTAR 2 wurde bislang noch nicht umfassend validiert. Bisherige Prüfungen untersuchten die Interrater-Reliabilität mittels Kappa-Statistiken, der überwiegend eine akzeptable Güte bescheinigt wird. Auch die Usability von AMSTAR 2 wurde evaluiert, wonach die Dauer der Anwendung einen Zeitbedarf zwischen 15 und 32 Minuten erfordert. Dieses lässt sich rechtfertigen, da 10 Fragen aus dem ursprünglichen Instrument beibehalten wurden und die Entwicklung auf einem breiten Expertenkonsens und Anwenderfeedback basiert (Shea et al., 2017). Während die erste Version von AMSTAR in einer übersetzten deutschsprachigen Version verfügbar ist (Schmucker et al., 2017), existiert für AMSTAR 2 bislang keine offizielle deutsche Fassung.

Zur Unterstützung bei der Anwendung von AMSTAR 2 existieren zu allen 16 Bewertungskriterien Erklärungen und Interpretationshilfen, die im Rahmen der ENP-Entwicklung bei der Literaturbewertung genutzt werden.

Die Vergabe von Score-Punkten zur Quantifizierung und Einstufung der Gesamtgüte einer systematischen Übersichtsarbeit bzw. einer Meta-Analyse ist bei AMSTAR 2 wie auch bei der ursprünglichen Fassung nicht vorgesehen. Stattdessen wird eine zusammenfassende Bewertung auf Basis des Urteils der AMSTAR 2 anwendenden Person vorgeschlagen. Ein Raster für eine solche Einstufung wird von Shea und Kollegen

[5] vgl. https://amstar.ca/Amstar_Checklist.php (Zugriff am 02.03.2022)

(2017) vorgeschlagen (vgl. nachstehende Tabelle), es wird jedoch darauf verweisen, dass auch andere Verfahren genutzt werden können.

Bewertung des allgemeinen Vertrauens in die Ergebnisse der Bewertung eines SR / einer Meta-Analyse	
Hoch	**Keine oder eine unkritische Schwäche:** Die systematische Übersicht liefert eine genaue und umfassende Zusammenfassung der Ergebnisse der verfügbaren Studien, die sich auf die Forschungsfrage beziehen.
Mäßig/Moderat	**Mehr als eine unkritische Schwäche:** Die systematische Übersicht weist mehr als eine Schwäche auf, aber keine kritischen Mängel. Sie kann eine genaue Zusammenfassung der Ergebnisse der verfügbaren Studien liefern, welche in der Untersuchung einbezogen wurden.
Gering/niedrig	**Ein kritischer Mangel mit oder ohne unkritische Schwächen:** Die Überprüfung weist einen kritischen Fehler (Risk of Bias) auf und bietet möglicherweise keine genaue und umfassende Zusammenfassung der verfügbaren Studienlage, die sich auf die Forschungsfrage beziehen
Sehr niedrig	**Mehr als ein kritischer Fehler mit oder ohne unkritische Schwächen:** Die Überprüfung hat mehr als einen kritischen Fehler und bietet keine umfassende Zusammenfassung der verfügbaren Studien.

Tabelle 9: Zusammenfassende Bewertung des Gesamtvertrauens in die Ergebnisse einer systematischen Übersichtsarbeit / einer Meta-Analyse (Quelle: eigene Darstellung und Übersetzung nach Shea et al., 2017, S. 6)

Für die zusammenfassende Bewertung einer Studie bzw. Publikation wird im Rahmen der ENP-Entwicklung ein an GRADE (Grading of Recommendations, Assessment, Development and Evaluation) angelehntes Einstufungsverfahren genutzt, das viele Ähnlichkeiten mit diesem Schema zur Gesamtbewertung aufweist (vgl. Kapitel 2.2.3).

2.2.1.2 Beobachtungsstudien (Fall-Kontroll-, Kohorten-, Querschnittsstudien) – Tools des Joanna Briggs Instituts (JBI) zum Critical Appraisal

Beobachtungsstudien verfolgen sehr unterschiedliche Zielsetzungen. Das Kontinuum reicht von der Entdeckung neuer Zusammenhänge bis zur Bestätigung bereits bekannter Erkenntnisse. Die drei häufigsten Studiendesigns, die vor dem Hintergrund der verschiedenen Zielsetzung angewendet werden (Fall-Kontroll-, Kohorten- oder Querschnittsstudien), können durch das weitverbreitete STROBE Statement (Strengthening the Reporting of Observational Studies in Epidemiology) bewertet werden. Bei der Entwicklung des STROBE Statements wurde auf frühere Erfahrungen im Rahmen der Entwicklung von Instrumenten zur Berichterstattung von Beobachtungsstudien zurückgegriffen. Hier ist im Besonderen das CONSORT Statement (Consolidated Standards of Reporting Trials) zu erwähnen, das vorwiegend zur Berichterstattung für RCTs entwickelt wurde (Cuschieri, 2019a; Moher et al., 2010). Unter Berücksichtigung empirischer Belege und theoretischer Überlegungen hat eine Gruppe von Forschern das STROBE Statement entwickelt. Das Instrument umfasst insgesamt 22 Items, wovon einige spezifisch auf die Forschungsdesigns der Fall-Kontroll-, Kohorten- bzw. Querschnittsstudie zugeschnitten sind. Mit dem STROBE Statement wird Autoren ein unterstützendes Instrument für eine möglichst optimale Berichtsverfassung von Forschungsergebnissen angeboten. Zwar können die Kriterien des STROBE Statements auch zur Bewertung und Interpretation von Forschungsarbeiten und der Einschätzung deren Güte herangezogen werden (Cuschieri, 2019b; Vandenbroucke et al., 2014; von Elm et al., 2014), was dem eigentlichen Anwendungszweck jedoch entgegensteht. Diese Zweckentfremdung wird vielfach kritisch betrachtet (Haile, 2021). Das STROBE Statement in der deutschen Fassung kann zur Orientierung und weiteren Anregung oder im Rahmen der Berichterstattung einer ENP-Entwicklungsarbeit genutzt werden.

Für das Critical Appraisal von Beobachtungsstudien sind zahlreiche spezifische Instrumente entwickelt worden, die vielfach auch Besonderheiten eines jeweiligen Studiendesigns berücksichtigen. Dreier und

Kolleg(inn)en (2010) berichten hinsichtlich Instrumenten zum Critical Appraisal und zum Reporting von Beobachtungsstudien von einer Vielfalt von 30 Instrumenten, eine Zahl, die seither nochmals angewachsen ist. Exemplarisch können hier die Checklisten des „**S**cottish **I**ntercollegiate **G**uidelines **N**etworks“ (SIGN), das ROBINS-I Tool (**R**isk **o**f **B**ias **I**n **N**on-randomized **S**tudies of Interventions), die Newcastle-Ottawa Scale (NOS) oder das AXIS-Tool (Appraisal tool for Cross-Sectional Studies) aufgeführt werden (Cochrane Deutschland (Ed.) et al., 2021; Ma et al., 2020). Das ROBINS-I-Tool wird beispielsweise im Rahmen der Leitlinienentwicklung der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) zur Evidenzsynthese genutzt (Cochrane Deutschland (Ed.) et al., 2021; Ma et al., 2020). Es ist ein mehrstufiges, komplexes Werkzeug zur Bewertung des systematischen Verzerrungsrisikos (Risk of Bias) der Ergebnisse von nicht-randomisierten Interventionsstudien. Wie auch zahlreiche andere Instrumente unterliegt ROBINS-I verschiedenen Einschränkungen (Page, McKenzie, & Higgins, 2018). Das Tool beschränkt sich auf quantitative Studien, in denen der Nutzen oder Schaden einer Intervention untersucht wird. Hierzu zählen z. B. Kohorten-, Fall-Kontroll-Studien oder Vorher-Nachher Studien.

Nach Durchsicht verschiedener verfügbarer Instrumente zur Bewertung von Beobachtungsstudien wurde entschieden, im Kontext der ENP-Entwicklung die Checklisten des in Adelaide (Australien) ansässigen Joanna Biggs Instituts (JBI) zu nutzen ((Ma et al., 2020). Diese sind eine anerkannte Methode und decken das erforderliche Spektrum im Bereich von Beobachtungsstudien umfassend ab (Munn et al., 2020). Zudem sind die Instrumente des JBI in der Pflege und Medizin sowie weiteren Professionen des Gesundheitswesens häufig verwendet und etabliert (Dietrich et al., 2016; Hopper et al., 2019; Porritt, Gomersall, & Lockwood, 2014; Stratil et al., 2021; Williams, Boylan, & Nunan, 2020). Zu jedem Critical Appraisal des JBI stehen zudem Erklärungen bzw. Interpretationshilfen für die einzelnen Bewertungskriterien zur Verfügung.

Beobachtungsstudien lassen sich grundsätzlich in analytische und deskriptive Studien unterteilen. Deskriptive Studien sind beschreibende Studien, bei denen die erhobenen Daten ausgewertet werden, aber keine kausalen Zusammenhänge untersucht werden. Querschnittsstudien (auch: Cross-sectional studies) und Langzeitstudien (auch: Longitudinal studies) zählen zu den deskriptiven Beobachtungsstudien. Analytische Studien widmen sich hingegen der Suche nach dem Ursache-Wirkungs-Prinzipien und/oder nach Zusammenhängen oder bestimmten Risikofaktoren (Exposition). Zu den analytischen Verfahren zählen Fall-Kontroll-Studien (auch: Case-control studies) und Kohortenstudien (auch: Cohort studies).

Fall-Kontroll-Studien können sowohl retrospektiv als auch prospektiv aufgesetzt werden. In der Studie werden Personen mit einer bestimmten Krankheit oder dem Vorliegen eines bestimmten pflegerelevanten Phänomens mit nicht erkrankten Personen bzw. solchen ohne das Vorliegen des pflegerelevanten Phänomens verglichen. Die kritische Bewertung von Fall-Kontroll-Studien im Rahmen der ENP-Entwicklung erfolgt mit der „Critical Appraisal Checklist for Case Control Studies“ des JBI (Moola et al., 2020).

In einer **Kohortenstudie** werden die Teilnehmer(innen) nach festgelegten Merkmalen ausgewählt und die Stichprobe über einen längeren Zeitraum beobachtet. Erhoben werden beispielsweise Risikofaktoren der Proband(inn)en (z. B. Adipositas, Diabetes) und ob, beziehungsweise wie bestimmte Ereignisse (z. B. ein Dekubitus) im Vergleich zu nicht exponierten Proband(inn)en auftreten. Für diese Art von Forschungsdesign steht von JBI folgendes Critical Appraisal zur Verfügung: „Checklist for Cohort Studies“ (Moola et al., 2020).

Bei **Querschnittsstudien** werden zu einem definierten Zeitpunkt einmalig klinische oder pflegerische Befunde oder Parameter innerhalb einer Stichprobe im Sinne einer „Momentaufnahme“ erhoben und beschrieben. Das Bewertungsinstrument des JBI zur Beurteilung von Studien mit diesem Forschungsdesign ist die „Checklist for Analytical Cross Sectional Studies“ (Moola et al., 2020).

Zur Bewertung von Publikationen, denen ein **quasi-experimentelles Forschungsdesign** zugrunde liegt (bedeutet alle experimentellen Studien, bei denen die Gruppen nicht-zufällig und damit ohne

Randomisierung zustande gekommen sind), wird ebenfalls eine entsprechende Checkliste des JBI genutzt, die „Checklist for Quasi-Experimental Studies“ genutzt.

Alle Critical Appraisal Tools des JBI samt Interpretationshilfen sind über eine eigene Internetpräsenz verfüg- und abrufbar: https://jbi.global/critical-appraisal-tools (Stand: 24.05.2024).

2.2.1.3 Randomisiert kontrollierte Studien (RCTs) – Bewertungssystematik in Anlehnung an die Cochrane-Empfehlungen nach Platz (2021)

Auch für randomisiert kontrollierte Studien (RCTs) stehen eine Vielzahl verschiedener Instrumente sowohl zur Berichterstattung als auch zum Critical Appraisal zur Verfügung (Zeng et al., 2015). Ein weit verbreitetes Beispiel für das Reporting von RCTs ist etwa das CONSORT Statement (Consolidated Standards of Reporting Trials) (Cuschieri, 2019a; Moher et al., 2010). Das insgesamt 25 Items umfassende Instrument stellt eine Weiterentwicklung des vorausgegangenen SORT Statement (Standards of Reporting Trials) dar, das für eine Reihe von Schwächen häufig kritisiert wurde. Da RCTs in der Medizin und bisweilen auch in der Pflege vielfach als Goldstandard zur Bewertung von Interventionskonzepten angesehen werden und es auch bei diesem Forschungsdesign bei methodischen Mängeln zu verzerrten Ergebnissen kommen kann, ist eine gute Berichterstattung der Studienergebnisse bedeutend. Im Rahmen der ENP-Entwicklung kann das CONSORT Statement daher ergänzend genutzt werden, um eine gute Berichterstattung zu erkennen. Ergänzend hierzu empfiehlt sich das Studium der Erläuterungen und Auslegung zu den Fragestellungen bzw. Items des CONSORT Statements (vgl. hierzu Knippschild et al., 2015; Moher et al., 2010).

Neben Checklisten zur Berichterstattung bestehen auch zahlreiche Instrumente zum Critical Appraisal von randomisiert kontrollierten Studien. Das Cochrane Manual zur Bewertung des systematischen Verzerrungsrisikos (Risk of Bias) von randomisierten Studien etwa ist im Rahmen der Erstellung von Leitlinien weit verbreitet. Es liegt mittlerweile in einer überarbeiteten zweiten Version aus dem Jahr 2021 vor (Cochrane Deutschland (Ed.) et al., 2021; Schmucker et al., 2016). Grundsätzlich kann das Instrument sowohl für individuell randomisierte Studien mit parallelen Gruppen als auch für Cluster-randomisierte Studien mit parallelen Gruppen oder individuell randomisierte Cross-Over-Studien verwendet werden (Cochrane Deutschland (Ed.) et al., 2021). Basierend auf den Vorgaben der Cochrane-Empfehlung zum Risk of Bias hat Platz (2021) ein Critical Appraisal für RCTs abgeleitet. Vor dem Hintergrund der Anlehnung der Bewertungskriterien an die Cochrane-Empfehlungen und dem bereits erprobten Einsatz im Rahmen der Leitlinienentwicklungen der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V. (AWMF) wird dieses Instrument für die Bewertung von RCTs im Rahmen der ENP-Entwicklung genutzt.

Platz schlägt vor, nach der Bewertung der 14 Fragen die Studie gesamthaft bezüglich ihrer Validität zu bewerten, sehr gute Validität (++), gute Validität (+), mangelhafte Validität (-), geringe Validität (--) (Platz, 2021). Die Gesamtbewertung einer genutzten Publikation wird in ähnlicher Form auch im Rahmen der ENP-Neu-/Weiterentwicklung einheitlich für alle Forschungsdesigns und Publikationstypen umgesetzt, jedoch basierend auf einer anderen Grundlage in Form des GRADE-Systems (vgl. Kapitel 2.2.3).

2.2.1.4 Qualitative Forschungsarbeiten – Bewertungssystematik in Anlehnung an JBI Critical Appraisal Checklist for Qualitative Research

In der evidenzbasierten Pflegeforschung nehmen qualitative Forschungsdesigns einen zunehmend wichtigen Stellenwert ein. Dies liegt in der Notwendigkeit begründet, komplexe Sachverhalte, Erfahrungen, Perspektiven und Lebenswelten von Pflegeempfänger(inne)n, deren Angehörigen und Fachpersonen im Kontext verschiedener gesundheits- und krankheitsbezogener Phänomene zu beschreiben und nachzuvollziehen. Durch qualitative Forschung können explorative Einblicke gewonnen werden, welchen Fragestellungen nach dem „Wie“ oder „Warum“ zugrunde liegen (Fringer & Schrems, 2018; Williams, Boylan, & Nunan, 2019). Je nach Fragestellung und Zielsetzung steht in der qualitativen Forschung ein weites Spektrum an verschiedenen Forschungsdesigns zur Verfügung, beispielsweise

phänomenologische, ethnographische oder die Methode der Grounded Theory (Behrens & Langer, 2016). Im Rahmen der ENP-Entwicklung ist es dabei unerlässlich, qualitative Studien hinsichtlich ihrer Güte und Eignung im Kontext verschiedener Praxisleitlinien zu prüfen und zu bewerten. Hierbei müssen Kernkriterien bzw. Gütekriterien qualitativer Forschungsdesigns beachtet werden, um die Glaubwürdigkeit einer Studie und ihrer Forschungsergebnisse zu bewerten (Moorley & Cathala, 2019; Noble & Smith, 2015; Williams et al., 2020). Zur kritischen Bewertung der Studien wurden zahlreiche Instrumente bzw. Checklisten entwickelt, zu denen es keinen internationalen Konsens über die beste Eignung im Hinblick auf die Bewertung qualitativer Studien gibt (Hannes, 2011; Munthe-Kaas et al., 2019). Ein international etabliertes Critical Appraisal Tool ist die von Mitarbeitenden des Joanna Briggs Instituts entwickelte „Critical Appraisal Checklist for Qualitative Research", die vom wissenschaftlichen Ausschuss des Instituts im Anschluss an ein Peer-Review-Verfahren genehmigt wurde (Lockwood, Munn, & Porritt, 2015; Williams et al., 2020). Diese Checkliste kann zur Bewertung der Gesamtheit qualitativer Studien verwendet werden und berücksichtigt nicht mögliche Besonderheiten verschiedener qualitativer Forschungsansätze. Das Critical Appraisal Tool samt Interpretationshilfe ist über die eigene Internetpräsenz des JBI verfüg- und abrufbar: https://jbi.global/critical-appraisal-tools (Stand 13.07.2022).

Die Notwendigkeit der Entwicklung methodenspezifischer und wissenschaftlich untersuchter Instrumente zur Beurteilung qualitativer Studien steht im aktuellen Diskurs evidenzbasierter Gesundheitsforschung (Williams et al., 2020). Aus diesem Grund erfolgt die kritische Beurteilung von qualitativen Studien im Zuge der ENP-Entwicklung vorerst nicht methodenspezifisch entsprechend des jeweiligen qualitativen Studiendesigns, jedoch wird die Beobachtung des Fortschritts hierzu kontinuierlich weitergeführt, um mögliche Ergänzungen der Bewertungsinstrumente vornehmen zu können.

2.2.1.5 Leitlinien und Clinical Practice Guidelines – Bewertungssystematik in Anlehnung an das AGREE-II-Instrument

Die methodische Qualität des Erstellungsprozesses von Leitlinien ist häufig nicht transparent und auch die enthaltenen Aussagen und Empfehlungen sind nicht eindeutig zu interpretieren. Vor diesem Hintergrund sind wenig überraschend auch für die Bewertung von Leitlinien und Clinical Practice Guidelines eine Reihe an verschiedenen Instrumenten sowohl zum Critical Appraisal als auch zur Berichterstattung entwickelt worden. Ein Beispiel für ein Instrument zum Reporting von Leitlinien ist die RIGHT-Checkliste (Reporting Items for practice Guidelines in HealThcare), die von einem multidisziplinären, internationalen Team mit politischen Entscheidungsträgern im Rahmen eines vierstufigen methodischen Vorgehens entwickelt wurde, um die Qualität der Publikation von Leitlinien zu verbessern (Chen et al., 2017).

Ein etabliertes Beispiel, das bei der kritischen Beurteilung von Leitlinien Anwendung findet, ist das AGREE-Tool (Appraisal of Guidelines for Research and Evaluation), ein international verbreitetes sowie validiertes Verfahren (Burgers et al., 2004; Chiappini et al., 2017; MacDermid et al., 2005; Werner et al., 2016). AGREE wurde im Jahr 2009 in einer weiterentwickelten Fassung (AGREE II) veröffentlicht (Browman et al., 2017). Diese überarbeitete Version löst unter anderem die vormaligen Ja/Nein-Antwortmöglichkeiten der einzelnen Bewertungskriterien durch eine siebenstufige Rating-Skala von *„stimmt überhaupt nicht"* (1) bis *„stimmt voll und ganz"* (7) ab. Zudem wurde eine Kalkulationsgrundlage entwickelt, um die bewerteten Leitlinien miteinander bezüglich ihrer Güte zu vergleichen. Sämtliche Änderungen im Vergleich zur ersten Fassung von AGREE wie auch Interpretationshilfen und Erläuterungen zu den sechs Domänen und 23 Items des AGREE II sind im Benutzerhandbuch beschrieben (Browman et al., 2017). Für den deutschsprachigen Raum existiert sowohl eine übersetzte Originalfassung (The AGREE Collaboration (Ed.), 2001), als auch eine adaptierte und ergänzte Fassung des ursprünglichen AGREE-Instrumentes zur Bewertung von Leitlinien, genannt DELBI (Deutsches Instrument zur methodischen Leitlinien-Bewertung) (Kopp et al., 2008; Thole et al., 2007). Das DELBI wurde bereits bei der initialen Veröffentlichung im Jahr 2005 um eine zusätzliche siebte, im weiteren Verlauf (2008) schließlich um eine weitere achte Domäne ergänzt und bietet gegenüber der Ja/Nein-Bewertung des ersten AGREE eine Einschätzung der einzelnen Items auf einer

Skala von 1 (trifft überhaupt nicht zu) bis 4 (trifft vollständig zu). Die beiden ergänzten Domänen fokussieren zum einen die Anwendbarkeit der zu beurteilenden Leitlinie im deutschen Gesundheitswesen (Domäne 7) sowie die methodische Exaktheit der Leitlinien-Entwicklung bei der Verwendung von existierenden Leitlinien (Domäne 8), die Items der anderen Domänen decken sich im Wesentlichen mit den AGREE-Items. Bei der Entwicklung des DELBI wurden die Erfahrungen der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF), des Ärztlichen Zentrums für Qualität in der Medizin (ÄZQ) und ihrer Partner, die Entwickler des AGREE sowie des internationalen Netzwerks für Leitlinienentwicklungen eingebunden (Kopp et al., 2008).

Darüber hinaus steht auch das aktuellere AGREE II in einer systematisch übersetzten deutschen Fassung zur Verfügung, die inhaltlich und mit Blick auf die Anwendung dem englischsprachigen Original entspricht (The AGREE Next Steps Consortium (Ed.), 2013). Nicht zuletzt scheint auch eine Neufassung des DELBI auf den Weg gebracht worden zu sein, genannt DELBI 2.0, das wiederum auf einer für den deutschsprachigen Raum adaptierten Form des AGREE II basiert. Offenbar hat es DELBI 2.0 jedoch nicht geschafft, den Status einer Konsultationsfassung zu verlassen (Dippmann et al., 2013) und wurde bis heute nicht offiziell veröffentlicht[6]. Eine Recherche nach einer Begründung hierfür oder einer Skizzierung der Zusammenhänge blieb leider erfolglos. Dies erscheint auch dahingehend bedauerlich und verwirrend, als dass die derzeit (Stand März 2022) aktuelle Fassung eine Gültigkeit des ersten DELBI nur bis Mitte 2015 ausweist:

„*Das Deutsche Instrument zur methodischen Leitlinien-Bewertung (DELBI) wurde am 7. Juni 2005 veröffentlicht (Fassung 2005/2006), am 22. September 2008 ergänzt (Fassung 2005/2006 + Domäne 8) und ist bis zur nächsten Überarbeitung bzw. spätestens bis zur Mitte des Jahres 2013 gültig. Ein vollständiges Aktualisierungsverfahren wurde entsprechend eingeleitet. Nach Überprüfung durch die Herausgeber erfolgte zunächst eine Verlängerung der Gültigkeit bis Mitte 2015*“ (Kopp et al., 2008, S. 4).

Vor dem geschilderten Hintergrund und insbesondere dem unsicheren Status einer aktuellen Version des DELBI-Instrumentes wird zum Critical Appraisal von Leitlinien und/oder Clinical Practice Guidelines im Rahmen der ENP-Entwicklung daher eine leicht adaptierte Form der deutschen Fassung von AGREE II herangezogen.

Wie bei vielen der bisher vorgestellten Bewertungsinstrumenten existieren auch für AGREE II Interpretationshilfen und Erläuterungstexte für die einzelnen 23 Items, die in der Anleitung zur Anwendung ausführlich dargestellt sind (The AGREE Next Steps Consortium (Ed.), 2013).

Alle Items von AGREE II werden laut Manual auf einer siebenstufigen Likert-Skala bewertet (1: trifft überhaupt nicht zu bis 7: trifft vollständig zu). Das Bewertungsprozedere sieht darüber hinaus vor, dass für jede der sechs Domänen ein sog. Domänenwert berechnet wird, diese dann unabhängig voneinander betrachtet und nicht zu einem Gesamt-Qualitätsscore für die zu beurteilende Leitlinie zusammengefasst werden sollen. Hierzu werden die einzelnen Werte für jedes Item einer Domäne addiert und mit der Anzahl der Rater multipliziert und anschließend relativ (prozentual) als Anteil der maximal möglichen Punktzahl der jeweiligen Domäne (auch hier mit der Anzahl an Ratern multipliziert) ausgedrückt. Dieses Vorgehen soll den Vergleich von Leitlinien ermöglichen und die Entscheidung hinsichtlich der Empfehlung einer Leitlinie unterstützen. Ein Cut-off für einen „guten“ bzw. „schlechten“ Domänenwert wurde vom AGREE-Konsortium bewusst nicht gewählt, diese Entscheidung soll seitens des Raters im jeweiligen Kontext der Leitlinie erfolgen (The AGREE Next Steps Consortium (Ed.), 2013). Die Gesamtbewertung der jeweiligen Leitlinie erfolgt im Nachgang der Einzelbewertung der sechs Domänen bzw. 23 Kriterien mit zwei weiteren, abschließenden Items. Hier wird einerseits die Gesamtqualität der Leitlinie auf einer ebenfalls siebenstufigen Skala (1: geringstmögliche Qualität bis 7: höchstmögliche Qualität) eingeschätzt und

[6] So ist auch unter dem offiziell angegebenen Internetauftritt von DELBI nach wie vor nur die um die achte Domäne ergänzte Fassung von 2008 auffindbar, alle Hinweise auf DELBI 2.0 sind nicht mehr auffindbar (vgl. https://www.leitlinien.de/hintergrund/leitliniengrundlagen#delbi – letzter Zugriff am 08.03.2022)

andererseits eine Empfehlung des Raters im Hinblick auf die Anwendung der Leitlinie abgegeben (ja – ja mit folgenden Änderungen – nein).

Dieses Bewertungsschema soll auch im Rahmen der ENP-Entwicklung grundsätzlich weiterhin möglich bleiben, weswegen die detaillierte Bewertungspunktzahl eines Items auch immer mit vermerkt werden sollte. Da im Rahmen der ENP-Weiterentwicklung zur zusammenfassenden Evidenzbewertung einer Publikation eine Methodik in Anlehnung an GRADE verwendet wird (vgl. Kapitel 2.2.3) und darüber hinaus bei der Beurteilung einer Publikation durch mehrere Rater die Interrater-Reliabilität in Form der prozentualen Übereinstimmungsrate berechnet wird, werden die beiden zusammenfassenden Items zur Gesamtbewertung aus AGREE II nicht mit übernommen. Zudem werden für die interne Nutzung von AGREE II mit Blick auf die Bestimmung der Interrater-Reliabilität die sieben Antwortstufen bei der Bewertung der einzelnen Kriterien ergänzend zu insgesamt drei zusammenfasst: 1–3 (tendenziell nichtzutreffend), 4 (neutral) sowie 5–7 (tendenziell zutreffend).

2.2.1.6 Mixed-Methods-Forschungsarbeiten – Bewertungssystematik (auszugsweise) des MMAT (Mixed Methods Appraisal Tool)

Unter dem Begriff Mixed Methods wird im Regelfall die Verbindung quantitativer wie auch qualitativer Forschungsmethoden in einem kombinierten Studiendesign verstanden. Dabei geht Mixed-Methods-Forschung noch deutlich weiter als andere die beiden Forschungsstränge verbindenden Ansätze wie beispielsweise die Triangulation oder die Methodenkombination (Fringer, 2018; Kelle, 2019). Wesentliches Merkmal der Mixed Methods ist die systematische und gezielte Integration und Synthese qualitativer sowie quantitativer Elemente sowohl in den Schritt der Datenerhebung als auch in die der Datenanalyse sowie der Berichterstattung über die Ergebnisse desselben Forschungsvorhabens.

Während die Kombination von quantitativen und qualitativen Forschungsmethoden innerhalb ein und derselben empirischen Studie in Disziplinen wie der Sozial- oder Erziehungsforschung seit langer Zeit gängige Praxis ist, spielen Mixed Methods in den Gesundheitswissenschaften und auch im Speziellen der Pflegewissenschaft erst seit der jüngeren Vergangenheit eine zunehmend wichtige Rolle (Kelle, 2019; Knappertsbusch, Langfeldt, & Kelle, 2021; Niederberger & Peter, 2018). Dabei erscheint dieser Ansatz gerade im für die Pflege(-forschung), die in einem komplexen System der interdisziplinären und sektorenübergreifenden Gesundheitsversorgung agiert, von hoher Bedeutung, da die Versorgung von pflegebedürftigen Menschen ein breites Spektrum an pflegediagnostischen Perspektiven wie auch Interventionskonzepten verfolgt. Nicht selten weisen pflegerische Versorgungszusammenhänge ein hohes Maß an Komplexität auf, oft beeinflussen sich verschiedene pflegerische Phänomene gegenseitig. Monomethodische Forschungsansätze, die sich dieser Komplexität ausschließlich mit quantitativen oder qualitativen Methoden zu nähern versuchen, können ihr oftmals nicht in ausreichendem Maße gerecht werden (Quasdorf & Holle, 2018). Die seit vielen Jahren beobachtbare Kontroverse zwischen Vertreter(inne)n der quantitativen bzw. qualitativen Methodentradition, oft mit gegenseitiger Absprache von Validität und Wissenschaftlichkeit des jeweiligen Vorgehens, erschwert die Erfassung des komplexen pflegerischen Versorgungsgeschehens zusätzlich (Kelle, 2019; Quasdorf & Holle, 2018). Mixed-Methods-Studien stellen Niederberger & Peter (2018) zufolge einen vielversprechenden Ansatz dar, da ihr Erkenntnispotenzial sowohl in der Erfassung kausaler Zusammenhänge als auch im intersubjektiven Verstehen vielfältiger (auch neuer) Phänomene liegt und sie darüber hinaus eine subjektorientierte und evidenzbasierte Annährung an die (Pflege-)Praxis erlauben.

Mit der zunehmenden Bedeutung und Verbreitung des Mixed-Methods-Ansatzes in der Gesundheits- und Pflegeforschung geht auch der steigende Bedarf an geeigneten Werkzeugen zur Bewertung entsprechender Forschungsarbeiten einher. Die Recherche nach geeigneten Instrumenten zum Critical Appraisal für Mix-Methods-Studien brachte eine Reihe verschiedener Ansätze zutage. Beispielhaft erwähnt sei etwa das mit Blick auf die zu bewertenden Artikel recht breit aufgestellte „Quality Assessment with Diverse Studies“ (QuADS) (Harrison et al., 2021), das sowohl Kriterien für qualitative wie auch quantitative

Arbeiten enthält. Zur Beurteilung von Studien, die nicht nur ein monomethodisches Vorgehen verfolgen, sondern eine Methodenkombination oder den Mixed-Methods-Ansatz wählen, wird empfohlen, alle Kriterien des Instruments anzuwenden. Da das QuADS jedoch keine expliziten Bewertungskriterien für das zentrale Merkmal von Mixed-Methods-Forschung enthält, genauer die systematische Integration und Synthese quantitativer wie auch qualitativer Elemente in mehreren oder allen relevanten Phasen des Forschungsprozesses, kam es für die Nutzung im Kontext der Weiterentwicklung von ENP nicht in die engere Auswahl.

Sehr spezifisch innerhalb des Feldes der Mixed-Methods-Forschung agieren hingegen das „Mixed Methods Impact Evaluation Tool“ (MMIE) sowie das „Mixed Methods Systematic Reviews Appraisal Tool“ (MMSR) (Jimenez et al., 2018a, 2018b). Diese Instrumente fokussieren im Speziellen Mixed-Methods-Arbeiten zur Evaluierung der Wirkung/Effekte von Interventionen bzw. Systematic Reviews, die einen Mixed Methods Ansatz verfolgen, um eine Evidenzsynthese für eine bestimmte Fragestellung herzustellen. Während bei beiden Instrumenten der besondere integrative Faktor von quantitativen wie auch qualitativen Aspekten der Mixed-Methods-Forschung durch spezifische Bewertungskriterien Berücksichtigung findet, ist die Güte des MMIE bzw. MMSR bislang noch unzureichend untersucht. Dieser Umstand sowie der für die ENP-Weiterentwicklung zu detaillierte Fokus auf bestimmte Designs der Mixed-Methods-Forschung waren wesentlich für die Entscheidung, auch diese beiden Bewertungswerkzeuge derzeit nicht weiter zu verfolgen.

Die Wahl für das Critical Appraisal von Mixed-Methods-Studien im Rahmen der Weiterentwicklung von ENP-Praxisleitlinien fiel letztlich auf das „Mixed Methods Appraisal Tool“ (MMAT) (Hong et al., 2018; Pluye et al., 2009; Taylor & Hignett, 2014). Ursprünglich im Jahr 2006 entwickelt, erschien die aktuelle Revision des MMAT im Jahr 2018. Es bewertet sowohl die quantitativen als auch die qualitativen Aspekte einer Mixed-Methods Studie anhand fünf verschiedener Kategorien mit jeweils fünf Bewertungskriterien. Diese beziehen sich auf die Qualität der folgenden Forschungsdesigns: qualitative Forschung, RCTs, nicht randomisierte quantitativ orientierte Studien, quantitativ-deskriptive Studien sowie explizite Mixed-Methods-Studien. Diesen fünf Kategorien sind zwei allgemeine Bewertungskriterien für alle Studiendesigns vorangestellt. Alle Bewertungskriterien werden mit „Ja“ / „Nein“ / „Nicht beurteilbar“ bewertet, zusätzlich ist für jedes Bewertungskriterium die Möglichkeit zur Kommentierung gegeben. Es wird dringend davon abgeraten, aus der Bewertung einzelner Kriterien einen Gesamtscore zu berechnen. Stattdessen wird empfohlen, die vorgenommenen Bewertungen der einzelnen Kriterien detaillierter zu beschreiben. Zentrales Einsatzgebiet des MMAT sind in erster Linie empirische Studien, nicht hingegen theoretische Forschung oder Reviews. Das Instrument sollte von mindestens zwei Personen unabhängig voneinander angewandt werden.

Die Anwendung des MMAT erfolgt in einem zweistufigen Verfahren. Im Vorfeld der jeweiligen Bewertung der genannten Kategorien werden zunächst mittels eines Flussdiagramms die für die gegebene Mixed-Methods-Forschungsarbeit passenden methodischen Studiendesignelemente ausgewählt.

Abbildung 8: Flussdiagramm zur Auswahl methodischer Studiendesignelemente zum Critical Appraisal einer Mixed-Methods-Forschungsarbeit gemäß dem MMAT (Quelle: Hong et al., 2018, S. 8)

Wird beispielsweise neben einem qualitativen Ansatz ein quantitativer Strang in Form einer randomisiert-kontrollierten (Teil-)Studie verfolgt, wären die Kategorien 1, 2 und 5 des MMAT als Bewertungskriterien heranzuziehen, um sowohl die quantitativen wie auch die qualitativen Elemente einerseits für sich zu bewerten, als auch um diese Besonderheiten des Mixed-Methods-Ansatzes kritisch zu reflektieren. Zu allen Kategorien und Bewertungskriterien werden darüber hinaus Erklärungen bzw. Interpretationshilfen bereitgestellt, welche die Bedeutung und Intention des jeweiligen Kriteriums genauer ausführen. Neben dem englischsprachigen Original ist das MMAT auch in den Sprachen Französisch, Brasilianisch, Türkisch und Chinesisch verfügbar, eine offizielle deutsche Übersetzung existiert bislang nicht. Hinsichtlich der Güte des MMAT wurden bislang die Aspekte der Nützlichkeit, der Zuverlässigkeit, der Effizienz, der Inhaltsvalidität sowie der Interrater-Reliabilität in einzelnen Studien geprüft[7].

Für das Critical Appraisal von Mixed-Methods-Forschungsarbeiten im Zusammenhang mit der Weiterentwicklung von ENP wird das MMAT nur in Ausschnitten herangezogen, die englischsprachige Originalfassung des gesamten Instruments kann in der zugehörigen Publikation eingesehen werden (Hong et al., 2018). Genauer wird die fünfte Kategorie des Instruments, die sich explizit mit den Spezifika dieses integrativen Forschungsdesigns beschäftigt, zur Bewertung genutzt. Die Begründung hierfür ist darin zu sehen, dass für die Bewertung der anderen Kategorien des MMAT (qualitative Forschung, RCTs, nicht randomisierte quantitativ orientierte Studien, quantitativ-deskriptive Studien) bereits teils deutlich differenziertere und besser geprüfte Instrumente zur Verfügung stehen und ausgewählt wurden, die in den vorangegangenen Kapiteln ausführlich dargestellt wurden. Die Kombination der fünften Kategorie des MMAT mit den spezifischeren Bewertungsrastern für quantitative bzw. qualitative Forschungsdesigns lässt auf eine noch aussagekräftigere Einschätzung einer Mixed-Methods-Forschungsarbeit hoffen.

[7] Eine Übersicht über die zugehörigen Publikationen ist im Internet unter http://mixedmethodsappraisaltoolpublic.pbworks.com/w/page/127425458/Publications%20on%20the%20MMAT veröffentlicht (Abruf am 23.02.2022).

2.2.1.7 Fachbücher/Lehrbücher – eigene Umsetzung der Bewertungskriterien der Bibliothek der Cornell Universität in Ithaca, New York

Im Rahmen der ENP-Entwicklung werden im Bedarfsfall auch Fach- und Lehrbücher aus der Pflege und benachbarten Professionen genutzt, insbesondere, wenn sich die Forschungslage zu einem bestimmten Thema noch als weitgehend oder völlig unergründet erweist. Insofern es sich bei fachspezifischen Buchpublikationen bzw. Buchkapiteln um eine Veröffentlichungsform von Studien oder Forschungsarbeiten handelt oder eine literaturbasierte Übersicht zu einem Thema erstellt wurde, werden die bereits beschriebenen Bewertungskriterien des jeweiligen Forschungsdesigns genutzt. In allen anderen Fällen ergibt sich der Bedarf für eine spezifische Bewertung des Publikationstyps Fach-/Lehrbuch bzw. von Kapiteln aus denselben. Die Recherche nach geeigneten Bewertungsinstrumenten aus dem Kontext der Pflege- bzw. Gesundheitsversorgung für diesen spezifischen Zweck brachte jedoch nur wenige Ergebnisse in Form von zwei verschiedenen Ansätzen zutage. Hier handelt es einerseits um die „Checklist for Text and Opinion“ aus der Critical-Appraisal-Serie des Joanna Briggs Instituts (Joanna Briggs Institute (Ed.), 2017). Die Checkliste umfasst insgesamt sechs Kriterien, die mit „Ja“, „Nein“, „Unklar“ oder „Nicht anwendbar“ eingestuft werden. Abschließend erfolgt eine Gesamtbewertung zum Ein- bzw. Ausschluss der jeweiligen Buchveröffentlichung bzw. zum Bedarf vor einer Entscheidung, weitere Informationen einholen zu müssen. Wenngleich die „Checklist for Text and Opinion“ einem kontrollierten Entwicklungsprozess unterliegt und wiederholt evaluiert wurde und gemäß Definition für text- und meinungsbasierte Evidenz anwendbar ist, die aus Expert(inn)enmeinungen, Konsens, Kommentaren, Annahmen und Behauptungen entstammt und in Fachzeitschriften, Magazinen, Monografien und Berichten publiziert wird (Aromataris & Munn, 2020; McArthur et al., 2015), fiel die Entscheidung zur Bewertung von Fach- und Lehrbüchern gegen dieses Instrument. Der wesentliche Grund hierfür ist in der in den allermeisten Fällen andersartigen Natur von in Fach- und Lehrbüchern vermittelten Inhalten zu sehen, die keine Meinungen, Konsens oder Kommentare zu vermitteln versuchen, sondern einen (oft auch nicht-wissenschaftlich) für die Praxis und Ausbildung aufbereiteten Wissenskorpus zu einem oder mehreren Themen. Dementsprechend erweisen sich auch die Items der Checklist for Text and Opinion als nicht durchgängig passend zur Bewertung von Fach- und Lehrbüchern im klassischen Sinn[8].

Vor diesem Hintergrund fiel die Wahl auf ein Instrument, das aus einem allgemeinen Kriterienkatalog zur Bewertung von Informationsquellen der Bibliothek der Cornell Universität in Ithaca, New York[9] heraus konstruiert wurde und anlassbezogen spezifisch zur Bewertung von Büchern/Buchkapiteln entwickelt wurde (Datta, Funnell, & Ramuscak, 2012). Leider blickt dieses Instrument dem Stand der Recherche nach auf keine systematische Entwicklungsgeschichte und keinerlei Güteprüfungen zurück. Nachdem jedoch Fach-/Lehrbücher bzw. Kapitel hieraus, die nicht mit einem der anderen, spezifischen Instrumente aus den vorangegangenen Kapiteln bewertet werden können, ohnehin und letztlich unabhängig vom Bewertungsergebnis auf einer sehr niedrigen Evidenzstufe anzusiedeln sind, wurde die Zielgerichtetheit der Bewertungskriterien für den Kontext Lehr-/Fachbuch als bedeutsamer eingestuft als der wissenschaftliche Entwicklungsstatus des Instruments. Hierin ist auch die Begründung zu sehen, weswegen die bei Datta und Kollegen (2012) vorgestellte Umsetzung der Bewertungskriterien auch nochmals an die Bedarfe der ENP-Entwicklung adaptiert und mit Blick auf den Umfang reduziert wurde. Für den internen Gebrauch wurden die Items zudem nach bestem Gewissen vom englischsprachigen Original ins Deutsche übersetzt. Das Ergebnis sind insgesamt zehn mit „ja“, „teils“ oder „nein“ einzustufende Bewertungskriterien, die in einer zusammenfassenden Gesamtbewertung (stark, mäßig, schwach) resultieren.

[8] Als Beispiel für derartige Passungenauigkeit ist beispielhaft Item 3 aufzuführen, das frei ins Deutsch übersetzt wie folgt lautet: „Stehen die Interessen der betroffenen Bevölkerung im Mittelpunkt der Stellungnahme?“

[9] Vgl. https://guides.library.cornell.edu/critically_analyzing/home (Abruf am 23.02.2022)

2.2.1.8 Graue Literatur und Webseiten – Bewertungskriterien der AACODS Checkliste (Authority, Accuracy, Coverage, Objectivity, Date, Significance)

Im Rahmen der ENP-Entwicklung spielen Informationen aus Webseiten und grauer Literatur, die sich nicht mittels eines der bisher vorgestellten Instrumente bewerten lassen, nur eine untergeordnete bzw. unterstützende Rolle. Unter „grauer Literatur“ sind in diesem Zusammenhang alle Publikationen zu verstehen, die in gedruckter und/oder elektronischer Form veröffentlicht wurden, die nicht durch ein kommerzielles Verlagswesen veröffentlicht wurden. Typischerweise zählen hierunter etwa akademische Abschluss-/Qualifikationsarbeiten, Kongressberichte, Firmenschriften und dergleichen mehr, aber auch generische Internetauftritte/Webseiten (Paez, 2017; Woods, Phillips, & Dudash, 2020). Insofern sich kein anderes, spezifischeres Instrument zur Bewertung der jeweils bereitgestellten Informationen anwenden lässt, wird im Folgenden auf die sog. „AACODS-Checkliste“ zur Bewertung grauer Literatur zurückgegriffen, um alle verbleibenden Veröffentlichungsarten und Publikationen aufzugreifen, die nicht mit den bislang vorgestellten Instrumenten bedient werden können. Die AACODS-Checkliste wurde im Jahr 2010 an der Flinders University (Adelaide, Australien) veröffentlicht (Tyndall, 2008, 2010). Die AACODS besteht aus sechs weit gefassten Kategorien, aus deren Abkürzungen sich auch der Name des Instruments in Form eines Akronyms ergibt:

- Autorität / Kompetenz (Authority)
- Genauigkeit (Accuracy)
- Umfang / Reichweite (Coverage)
- Objektivität (Objectivity)
- Datum / Veröffentlichungszeitraum (Date)
- Bedeutung / Signifikanz (Significance)

Die AACODS findet mittlerweile regelmäßige Anwendung zur Bewertung von grauer Literatur, auch im Zusammenhang mit der Erstellung von Wissenssynthesen im gesundheitlichen Kontext (Vaska et al., 2019).

2.2.2 Evidenzsynthese und Bewertung einzelner Items/Handlungskonzepte

Eine Evidenzsynthese zu einzelnen Handlungskonzepten einer Pflegediagnose wird aktuell in ENP zwar bei der Literaturbewertung ermittelt, aber nicht in den Veröffentlichungen ausgewiesen. Die Handlungskonzepte einer Praxisleitlinie bilden die fachlich-inhaltlich sinnvollen pflegerischen Handlungsoptionen zu einer Pflegediagnose ab, eine Priorisierung der Handlungskonzepte und/oder Offenlegung der gefundenen Evidenzlage zu den einzelnen Handlungskonzepten wird derzeit nicht veröffentlicht. Das synthetisierte Evidenzniveau zu einzelnen Handlungskonzepten wird jedoch in der internen ENP-Entwicklungsdokumentation mitgeführt. Hintergrund ist, dass in der Pflege die Forschungslage in den allermeisten Fällen noch zu eingeschränkt ist, um zu den Handlungskonzepten im Kontext von Pflegediagnosen die (externe) Evidenzlage auszuweisen. Zahlreiche pflegerische Handlungskonzepte befinden sich noch auf der Ebene „best practice“. Auf der anderen Seite würde mit dem Ausweisen der Evidenzstärke zu den Handlungskonzepten das Risiko entstehen, dass ggf. der Entscheidungsfindungsprozess der Pflegepersonen im Abwägungsprozess zwischen ihrer internen und der externen Evidenz einseitig beeinflusst wird. Dies birgt dahingehend eine Gefahr, als dass die in der individuellen Versorgungssituation eines pflegebedürftigen Menschen angebrachte Wahl an Pflegemaßnahmen aus vielerlei Gründen von der an sich „evidentesten“ begründet abweichen kann. In den Studien bzw. Fachveröffentlichungen nicht bestätigte Handlungskonzepte werden konsequent nicht in ENP aufgenommen bzw. mit einer ENP-Pflegediagnose verknüpft, auch wenn entsprechende Interventionen in der Pflegepraxis (noch) angewendet werden oder eine Bitte um Aufnahme aus der

Pflegepraxis an die ENP-Entwicklung herangetragen werden. Bei der Evidenzsynthese und Bewertung einzelner Items/Handlungskonzepte wird sichergestellt, dass es sich um „best practice“ oder evidenzbasierte Handlungskonzepte/Items handelt, die eine entsprechende Gültigkeit (mindestens eine niedrige Evidenz) nachweisen konnten. Auch bei der Evidenzsynthese von Kennzeichen, Ursachen zur Beschreibung einer pflegediagnostischen Aussage wird die Evidenzstärke nicht zu den einzelnen Items ausgewiesen. Die ENP-Entwicklung konzentriert sich auf eine möglichst vollständige Abbildung in der Literatur bestätigter Zusammenhänge und/oder des bestätigten Auftretens entsprechender Items. Dies dient dem Ziel des Erreichens der erforderlichen „Vollständigkeit“ zur Abbildung eines pflegerischen Zusammenhangs, auch um im weiteren Verlauf diese Datengrundlage für die Pflegeforschung nutzen zu können. Eine große Herausforderung in diesem Zusammenhang ist das Finden des geeigneten Grades an Granularität (Detailliertheit bzw. Abstraktionsgrad), auch um inhaltliche Überschneidungen zu vermeiden.

Zusammengefasst: Grundsätzlich werden keine Pflegeinterventionen und/oder Items im Bereich der Kennzeichen/Risikofaktoren/Ursachen aufgenommen, bei denen eine negative Effektstärke und/oder Bedeutung im Kontext der jeweiligen ENP-Praxisleitlinie nachgewiesen ist. Ebenso werden keine Items aufgenommen, die nicht mit entsprechenden vertrauenswürdigen Referenzen belegt werden konnten.

2.2.3 Evidenztabelle zur Studienbewertung im Rahmen der ENP-(Weiter-) Entwicklung

Zur Dokumentation der Ergebnisse des kritischen Bewertungsprozesses (Critical Appraisal) bei der ENP-Entwicklung wird eine sogenannte „Evidenztabelle“ genutzt. Die Evidenztabelle ist eine wichtige Methode zur Darstellung der Qualität der für die ENP-Entwicklung genutzten Publikationen. Sowohl für die Mitglieder des Entwicklungsteams als auch für interessierte außenstehende Personen ist es bedeutend, prägnante Evidenzzusammenfassungen einsehen zu können, was auch zur Förderung der Transparenz des ENP-Weiterentwicklungsprozesses beiträgt. Die neue Evidenztabelle ist in grober Orientierung an den Leitlinienentwicklungsprozess der Deutschen Gesellschaft für Neurologie (DGN) und der Deutschen Gesellschaft für Neurorehabilitation (DGNR) vom ENP-Entwicklerteam erarbeitet worden (Platz, 2021). Wie bereits dargelegt, wird abhängig vom Studiendesign ein Critical Appraisal zu einer vorliegenden Publikation durchgeführt und die Bewertung bzw. die Ergebnisse in komprimierter Form in der Evidenztabelle dokumentiert. Die Gesamtbetrachtung der eingeschätzten Fragen wird ebenfalls in der Evidenztabelle dokumentiert wie auch zentrale Ergebnisse und Informationen zur Studie selbst. Nachfolgend wird die Grundstruktur der Evidenztabelle vorgestellt.

2.2.3.1 Aufbau der Evidenztabelle

Die zur ENP-Weiterentwicklung erarbeitete Evidenztabelle basiert auf einem Microsoft Excel-Dokument mit klar definierten Strukturen und Funktionen. In ihr festgehalten werden sowohl die zentralen Informationen und Kernergebnisse zur für die Weiterentwicklung herangezogenen Studien und Publikationen, die Bewertung deren Qualität (Evidenzbewertung) sowie die Schlussfolgerungen für ENP. Eine differenzierte Funktion zum Filtern der eingetragenen Informationen und Daten ermöglich es, zu spezifischen Aspekten einen schnellen Überblick zu bekommen, selbst wenn die Evidenztabelle mit Dutzenden von Referenzen befüllt ist. Nachfolgende Abbildung zeigt die wesentlichen Elemente der Evidenztabelle, die im Anschluss kurz erläutert werden.

Abbildung 9: Inhalte der Evidenztabelle zur Bewertung und Analyse von Publikationen

Die Felder im oberen Bereich der Evidenztabelle spiegeln die Filterfunktionen wider, mit denen mittels weniger Mausklicks gezielt bestimmte Inhalte ein- oder ausgeblendet werden können, um je nach gegebenem Interesse rasch Informationen oder Daten bündeln zu können.

Die nachfolgenden Spalten der Evidenztabelle spiegeln die eigentlichen Inhalte wider:

- **Ref.Nr.:** Hier wird die laufende Arbeitsnummer der Publikation beginnend bei 1 festgehalten, um im Rahmen der ENP-Weiterentwicklungsarbeiten bzw. in anderen Dokumenten einen einfach zu handhabenden Querverweis zu ermöglichen.
- **Autor(en):** Hier werden die Namen der Autor(inn)en bzw. Herausgeber(innen) der jeweiligen Publikation festgehalten. Bei mehr als drei Autor(inn)en wird in Anlehnung an den Zitationsstil APA 7th (entsprechend der 7. Ausgabe des Publikationshandbuchs der American Psychological Association) nur der/die erste Autor(in) genannt und mit „et al." auf die anderen Mitwirkenden verwiesen.
- **Jahr:** Hier wird das Erscheinungsjahr der Publikation bzw. Studie festgehalten.
- **Studientyp:** Hier werden in prägnanter Form Angaben über das Forschungsdesign der jeweiligen Publikation oder Studie festgehalten.
- Die Spalte **Population, TN-Zahlen, ggf. Studienanzahl** wird mit Angaben zu den wesentlichen Charakteristika der Studie bzw. Publikation in Abhängigkeit vom jeweiligen Studientyp befüllt. Bei Reviews etwa die Zahl an eingeschlossenen Einzelstudien, oder aber Angaben über Teilnehmer(innen)zahlen und/oder Gruppengrößen bei empirischen Arbeiten.
- In die Spalten **Evidenzbewertung (Rater 1 bzw. Rater 2)** wird unter Berücksichtigung aller in den nachfolgenden Kapitel 2.2.3.2 sowie 2.2.3.3 beschriebenen Gesichtspunkten zur Einstufung sowie zur eventuellen Auf- oder Abwertung die Gesamteinschätzung des Evidenzgrades für die jeweilige Publikation bzw. Studie eingetragen. Die Evidenzbewertung wird wann immer möglich für mindestens 10% aller Publikationen durch eine zweite Person in unabhängiger Form vorgenommen. Bei einer Übereinstimmung von 80% oder mehr bei den Bewertungsergebnissen durch beide Rater werden die weiteren Evidenzbewertungen alleinig von der hauptverantwortlichen Person für die Bearbeitung des zuvor im Team festgelegten Themas durchgeführt. Bei einer geringeren Übereinstimmungsquote wird die Bewertung durch zwei Personen fortgeführt.
- **Fragestellung, Methodik, Outcomemessung und Hauptergebnisse:** Diese Spalte ist für kurze und essenzielle Informationen gedacht, die zum einen den Forschungsschwerpunkt der jeweiligen Publikation behandeln und zum anderen den inhaltlichen Schwerpunkt (z.B. die Fragestellungen) prägnant beschreiben. Darüber hinaus ist kurz anzugeben, wie und ggf. wann (z.B. Follow-Up) die Ergebnisse/Outcomes der Arbeit erfasst bzw. gemessen wurden (Methodik). Schließlich werden hier auch die zentralen Hauptergebnisse der Publikation bzw. Studie kompakt beschrieben. Ergänzend hierzu werden je nach Studientyp kurze Angaben zu den wesentlichen Effektstärken (z.B. p-Wert) bzw. weiteren/anderen statistischen Resultaten gemacht.
- **Schlussfolgerung und ENP-Relevanz:** Diese Spalte dient der Formulierung der zentralen Schlussfolgerungen der Ergebnisse für die konkrete Aufnahme/Änderung in ENP im Sinne des Transfers der Erkenntnisse/Ergebnisse in passende ENP-Items.

- **ENP-Item-Typ:** Hier wird angeben, welche Art von ENP-Items durch die Veröffentlichung aufgenommen, geändert oder auch stillgelegt werden (Kennzeichen, Ursachen, Ressourcen, Pflegeziele, Pflegeinterventionen, etc.). Diese Angabe ermöglicht ein Filtern der gesamten Evidenztabelle nach Publikationen hinsichtlich der jeweiligen Item-Art.
- **Quelle (formatiert aus EndNote):** In diesem Bereich wird schließlich die vollständige Zitation der Publikation bzw. Studie im korrekten Zitationsstil direkt aus dem Literaturverwaltungsprogramm EndNote aufgenommen.

Eine im Arbeits- und Bewertungsprozess befindliche und daher noch nicht vollständig befüllte Evidenztabelle wird in Anlage I exemplarisch vorgestellt.

Wichtig für die Überführung der Einzelergebnisse aus den Bewertungssystemen (vgl. Kapitel 2.2.1) in eine gesamthafte Einschätzung hinsichtlich Evidenzgrad und Zuverlässigkeit der jeweiligen Publikation ist dabei die Betrachtung der Bedeutung und Auswirkung einzelner Fragen/Kriterien auf die Zuverlässigkeit der pflegerelevanten Aussagen. Sind z. B. bei den Fragestellungen kritische Bereiche, die mit hoher Wahrscheinlichkeit die Gültigkeit der Aussagen beeinflussen als „nicht erfüllt“ eingestuft worden, so hat dies eine gravierendere Bedeutung, als wenn eine Frage mit „nicht erfüllt“ eingeschätzt wurde, die in der Gesamtbetrachtung nur eine begrenzte Bedeutung auf die Validität der Studienaussagen/-ergebnisse hat.

2.2.3.2 Einleitende Überlegungen zu Einstufungsverfahren im Rahmen der Evidenzbewertung

Über die vergangenen Jahrzehnte hinweg hat sich eine Reihe verschiedener Evidenz-Ranking-Systematiken herausgebildet, die sich je nach Anwendungsgebiet wiederholt auch verschiedener Kritik gegenübersahen. So ist insbesondere im Kontext der Pflegeforschung bekannt, dass die in der Medizin oft als evidenter Goldstandard angesehenen randomisiert-kontrollierten Studien nicht zwangsläufig und automatisch als höherwertiger mit Blick auf den Evidenzgehalt anzusehen sind als beispielsweise Beobachtungsstudien (Behrens & Langer, 2004; Evans, 2003; Grypdonck, 2004; Lavin et al., 2002). Bei mancherlei pflegerelevanten Fragestellungen ist der „beste“ Evidenzbeweis nicht über RCTs zu belegen oder RCTs sind aus ethischen Gründen nicht durchführbar (Roberts & DiCenso, 1999). Zunehmend wird anerkannt, dass sowohl Beobachtungsstudien als auch Fallserien einen guten Beitrag zur Evidenzgewinnung bei Interventionen leisten können und auch qualitative sowie Mixed-Methods-Forschungsarbeiten wertvolle Belege für die Entwicklung von beispielsweise Pflegediagnosen liefern. Welches Studiendesign im Rahmen der ENP-Entwicklung die beste Antwort auf konkrete Fragestellungen geben kann, ist immer abhängig vom Kontext der jeweiligen Fragestellung (Petticrew & Roberts, 2003). Gerade bei pflegediagnostischen Phänomenen, wie beispielsweise das Leiden im Zusammenhang von Trauer oder Einsamkeit, eignen sich qualitative Studien aus der Phänomenologie oder Beobachtungsstudien, um etwa Kennzeichen aus der Erlebenswelt der Betroffenen zur Beschreibung eines pflegediagnostischen Konzeptes abzuleiten (Panke-Kochinke, 2012; Ploeg, 1999). Zusammenfassend eignen sich für die Klärung pflegediagnostischer Phänomene eher Beobachtungsstudien, Querschnittsstudien, qualitative Studien, Konzeptanalysen und dergleichen, während für Recherche nach sinnvollen pflegerischen Behandlungskonzepten randomisiert kontrollierte Studien (RCTs) bevorzugt werden sollten. Geht es um Fragestellungen über Ursachenzusammenhänge oder Auswirkungen von z. B. Umwelteinflüssen auf unsere Gesundheit, so können sich auch Kohortenstudien oder Fall-Kontroll-Studien gut eignen. Je nach Fragestellung werden in systematischen Übersichtsanalysen und Meta-Analysen mehrere Forschungsarbeiten zusammenfassend bewertet, um eine Forschungsfrage zu beantworten. Ergebnisse von Meta-Analysen und systematische Übersichtsarbeiten haben bei einer guten Auswahl und Qualität der einbezogenen Primärstudien in der Regel den höchsten Evidenzgrad. Nachfolgender Überblick stellt die wesentlichsten Studiendesigns und deren zentrale Merkmale ohne Anspruch auf Vollständigkeit vor.

Studiendesigns	
Blickwinkel/Merkmale	**Beschreibung, Besonderheiten**
Zeitliche Perspektive	**Retrospektiv:** Beispielsweise nachträgliche Auswertung von Daten aus Patient(inn)enakten, Krankenhausinformationssystemen, Pflegeprozessdokumentation u. ä. **Prospektiv:** Datenerfassung laufend ab Studienbeginn über einen gewissen Zeitraum, festgelegte Methode und anschließend Auswertung
Beobachtungsdauer	**Querschnittsstudie:** „Momentaufnahme", z. B. einmalige Messung oder Befragung **Längsschnittstudie:** Betrachtung im zeitlichen Verlauf bzw. mehrfache Messung mit zeitlichem Abstand
Kontrollgruppe	**Ohne Kontrollgruppe** (z. B. Fallserien, Anwendungsbeobachtungen, Befragungen, Dokumentenanalysen) **Mit Kontrollgruppe** (z. B. RCTs)
Triangulation	Es werden verschiedene Methoden in der Forschungsarbeit eingesetzt, um unterschiedliche Blickwinkel zusammenzuführen. Dabei kann es sich um qualitative und quantitative Studiendesigns handeln.
Analytische Studien widmen sich der Suche nach Ursachen-Wirkungs-Prinzipien wie z. B. Zusammenhänge bestimmter Risikofaktoren für ein Pflegephänomen.	**Kohortenstudie (Cohort Study):** Beobachtet Teilnehmer(innen) nach festgelegten Merkmalen in einer ausgewählten Stichprobe über einen längeren Zeitraum. **Fall-Kontroll-Studie (Case Control):** Vergleicht Pflegeempfänger mit und ohne einem spezifischen Pflegephänomen in Hinblick z. B. auf Einflussgrößen, Risikofaktoren, Kennzeichen usw.
Deskriptive Studien sind rein beschreibende Studien, Daten werden dargestellt und ausgewertet, ohne dass der Versuch unternommen wird, Ursachen oder kausale Zusammenhänge zu erklären.	**Beobachtungsstudie:** Erhebt zu einem bestimmten Zeitpunkt ein pflegerelevantes(n) Ereignis/Befund/Vorkommen; kein Kausalitätsnachweis. Die Beobachtungsstudie ist eine Art von epidemiologischem Studiendesign und kann z. B. in Form einer Kohorten-, einer Fall-Kontroll- oder Querschnittsstudie angelegt werden (Cuschieri, 2019b). **Epidemiologische Studie:** erforscht z. B. Risikofaktoren und ihre Verteilung in der Bevölkerung, meist Beobachtungsstudie **Einzelfallstudien; Fallberichte (Case Report); Fallserien (Case Series)** **Korrelationsstudien:** Geeignet, um Hypothesen zu bilden. Es wird anhand von Daten geprüft, ob z. B. ein Zusammenhang mit bestimmten beobachteten Faktoren und z. B. einer Pflegediagnose besteht. **Querschnittsstudie z.B. Prävalenz-/Inzidenzstudien:** Zum Beispiel über das Auftreten von unerwünschten Ereignissen in einer Einrichtung oder das Auftreten von Pflegephänomen in einer Population
Experimentelle Studien umfassen eine Reihe von Forschungsdesigns, bei denen Forscher eine oder mehrere Variablen manipulieren und jede Veränderung bei anderen Variablen kontrollieren und messen. Es wird dabei untersucht, inwieweit ein bestimmter Faktor (Variable) eine Situation/ein Verhalten oder einen Zustand beeinflusst.	**Randomisiert kontrollierte Interventionsstudie (RCT):** zusätzlich zufällige Verteilung der Proband(inn)en auf die Gruppen; (doppelt) verblindet randomisiert kontrolliert: zusätzliche Verblindung der Proband(inn)en (und Behandler(innen)) **Quasi-experimentelle Studie:** Haben einen experimentellen Aufbau, es fehlen aber wesentliche Merkmale wie z.B. Zufallsverteilung der Versuchspersonen oder die Kontrollgruppe. Es gibt z.B. kein zufälliges Auswahlverfahren der Studienteilnehmer(innen), es wird mindestens eine unabhängige Variable aktiv durch die experimentelle Anordnung des Studiendesigns untersucht. **Crossover-Trial Interventionsstudie/Cross-over-Studie:** Hierbei werden die Interventions- und Kontrollgruppen in der Studienmitte gewechselt **Interventionsstudie nicht-randomisiert kontrolliert:** Proband(inn)en werden verschiedenen Interventionen zugeordnet kontrolliert: Studie legt Intervention fest, Kontrollgruppe vorhanden **Vorher-Nachher-Studie:** Die Teilnehmer(innen) einer Studie werden z. B. vor und nach einer Intervention untersucht. **Multizentrische Studie:** Es handelt sich um klinische Studien, die in mehreren Einrichtungen gleichzeitig mit einem multizentrischen Studiendesign mit mehreren Forschungsteams realisiert werden.

Thieme RECOM

Qualitative Studiendesigns	**Qualitative Inhaltsanalyse:** Beruht auf der Analyse schriftsprachlich fixierter Daten, z. B. Interviews, Dokumentationen, Videomaterial mit gesprochenem Wort usw. **Interpretative Studien**, z. B. Kommunikationsanalysen, Narrationsanalysen, Dokumentationsanalysen **Grounded Theory:** Qualitative Forschungsmethode mit der Besonderheit, dass die Datensammlung und die Datenauswertung einander abwechseln. Der Forscher identifiziert, verfeinert und integriert Kategorien aus dem Datenmaterial, um eine Theorie abzuleiten und/oder zu bestätigen. **Delphi-Methode:** Ist ein Verfahren, um aus Expert(inn)enmeinungen Prognosen zu gewinnen und/oder Konsens und Dissens zwischen Expert(inn)enmeinungen deutlich zu machen. Über ein hochstrukturiertes Gruppenkommunikationsverfahren wird eine Lösung, ein Konsens für komplexe Problemstellungen/Fragestellungen aus den Einzelbeiträgen erarbeitet. **Phänomenologisches Studiendesign:** Hierbei geht es um Forschungsdesigns, die z. B. bestimmte pflegerelevante Phänomene wie Trauer, Einsamkeit, Hoffnungslosigkeit beschreiben (Mehrholz, 2010). Phänomenologische Studien sind gerade im Kontext der Pflegediagnostik bedeutend. Diese Studien zählen auch zu den deskriptiven Ansätzen. **Concept Analysis:** Sind sehr hilfreich, um beispielsweise pflegerische Phänomene in der Pflegepraxis genauer zu beschreiben und zu fassen
Zusammenfassung von Einzelstudien	**Review / Übersichtsarbeit:** Auswertung der relevanten aktuellen Literatur mit dem Ziel, ähnliche Studien mit derselben Fragestellung zusammenfassend zu bewerten. Bei einem „systematischen" Review nach einem festen Standard werden idealerweise Studien mit dem gleichen Endpunkt und auf hohem Niveau z. B. RCTs miteinander bewertet, um eine Evidenzaussage treffen zu können. **Metaanalyse:** Systematisches Review, das die Ergebnisse der Einzelstudien mithilfe statistischer Verfahren zusammenfasst. Die einbezogenen Primärstudien sollten sich bezüglich ihrer Studienpopulation möglichst ähnlich sein, und alle Studien sollten im besten Fall den gleichen primären Endpunkt haben. **Multizentrische Studien:** Klinische Studien, die an mehreren Einrichtungen parallel durchgeführt werden und deren wissenschaftliche Aussagekraft durch die Beteiligung verschiedener Untersucher(innen) erhöht wird.

Tabelle 10: Übersicht über die wichtigsten Studiendesigns und deren zentrale Merkmale (Quelle: eigene ergänzte Darstellung in Anlehnung an Rojahn, 2016, S. 17). Zur Erweiterung genutzte Publikationen: (Bartholomeyczik et al., 2008; Behrens & Langer, 2004; Moorley & Cathala, 2019; Morse et al., 1996; Niederberger & Dreiack, 2020; Panke-Kochinke, 2012; Roberts & DiCenso, 1999; Rojahn, 2016).

Verschiedene Studiendesigns bergen unterschiedliche Risiken eines systematischen Fehlers bzw. systematischen Verzerrungseffekts (Risk of Bias). Da es sehr unterschiedliche Fehlerquellen bei der Generierung von Wissen geben kann, lässt sich keine eindimensionale Rangfolge von Studientypen bezüglich des Level of Evidence[10] festlegen. Ein systematisches Verzerrungsrisiko ist beispielsweise die Verfälschung der Studienergebnisse durch eine unangemessene Auswahlmethode von Studienteilnehmenden (Selektionsbias). RCTs versuchen diesen möglichen Verzerrungseffekt durch die Zufallsauswahl von Teilnehmer(inne)n und die Verblindung sowohl der teilnehmenden als auch der die Studie durchführenden Personen zu verhindern. Würde es nur diesen einen Irrtumsfehler geben, wäre die Rangfolge (Evidenzhierarchie) eindimensional zu lösen. Zu den vielen anderen Möglichkeiten eines systematischen Fehlers, die bei einer Studie zu nicht validen Ergebnissen führen können, gehören beispielsweise die nicht angemessene Auswahl von Messinstrumenten oder ein fehlender theoretischer Rahmen für die Forschungsarbeit. Daher ist es nicht verwunderlich, dass selbst in der Medizin die klassische „Hierarchie der Evidenz" umstritten ist (Petticrew & Roberts, 2003). Die Evidenzhierarchie ist dennoch ein häufig genutztes Instrument, um basierend auf den Studiendesigns die Güte und Zuverlässigkeit im Rahmen der klinischen Entscheidungsfindung ableiten und einschätzen zu können. Nachfolgend

[10] Level of Evidence (Evidenzgrad/Evidenzlevel): Wissenschaftlicher Nachweis, der den Grad der Verallgemeinerbarkeit bzw. der Gültigkeit von Forschungsergebnissen nach einem Kriterienraster ausweist (Bartholomeyczik et al., 2008).

dargestellt die häufig verwendete Evidenzhierarchie, die sich in zahlreichen insbesondere medizinischen Veröffentlichungen mit einigen Adaptionen wiederfindet.

Klassische Evidenzhierarchie in der Medizin	
1	Systematischer Review und Meta-Analyse
2	Randomisiert kontrollierte Studien mit definitiven Ergebnissen
3	Randomisiert kontrollierte Studien mit nicht eindeutigen Ergebnissen
4	Kohortenstudien
5	Fall-Kontroll-Studien
6	Querschnittsstudien
7	Einzelfallstudien

Tabelle 11: Klassische Evidenzhierarchie in der Medizin (Quelle: eigene Darstellung in Anlahenung an Petticrew & Roberts, 2003)

Beim Vergleich der verschiedenen Evidenzhierarchien sind Ähnlichkeiten zu erkennen, ebenso ist zu erkennen, dass qualitative Forschungsdesigns in den allermeisten Fällen nicht repräsentiert sind. In der Evidenzhierarchie des Scottish Intercollegiate Guidelines Network (SIGN) wird bereits ein qualitativer Umsetzungsgrad der Studie mit einbezogen.

Evidenzhierarchie nach dem SIGN Grading System	
1++	Qualitativ hochwertige Meta-Analysen, systematische Überprüfungen von RCTs oder RCTs mit einem sehr geringen Risiko der Verzerrung
1+	Gut durchgeführte Meta-Analysen, systematische Übersichten oder RCTs mit geringem Risiko einer Verzerrung
1-	Meta-Analysen, systematische Übersichten oder RCTs mit einem hohen Risiko der Verzerrung
2++	Hochwertige systematische Übersichten von Fall-Kontroll- oder Kohortenstudien Qualitativ hochwertige Fall-Kontroll- oder Kohortenstudien mit einem sehr geringen Risiko von Verwechslungen oder Verzerrungen und einer hohen Wahrscheinlichkeit, dass der Zusammenhang kausal ist
2+	Gut durchgeführte Fall-Kontroll- oder Kohortenstudien mit einem geringen Risiko von Verwechslungen oder Verzerrungen und einer mäßigen Wahrscheinlichkeit, dass der Zusammenhang kausal ist
2-	Fall-Kontroll- oder Kohortenstudien mit einem hohen Risiko von Verwechslungen oder Verzerrungen und einem signifikanten Risiko, dass der Zusammenhang nicht kausal ist
3	Nicht-analytische Studien, z. B. Fallberichte, Fallserien
4	Expert(inn)enmeinung

Tabelle 12: Evidenzhierarchie gemäß dem SIGN Grading System (Quelle: eigene Darstellung in Anlehnung an Harbour & Miller, 2001)

Das Oxford Centre for Evidence-Based Medicine schlägt vor, die Evidenzhierarchie nicht pauschal vom Studiendesign, sondern in Abhängigkeit von der zugrundeliegenden Fragestellung aufzubauen (Durieux, Vandenput, & Pasleau, 2013; Oxford Centre for Evidence-Based Medicine, 2009). Mittlerweile besteht in diesem Zusammenhang auch zunehmend Einigkeit darüber, dass Evidenzaussagen im Kontext der Interventionsforschung nicht grundsätzlich und ausschließlich auf RCTs basieren müssen, um eine höhere Evidenzstufe zu erreichen (Behrens & Langer, 2004; Gastinger et al., 2002; Petticrew & Roberts, 2003). Es gibt sogar wissenschaftliche Hinweise darauf, dass RCTs nicht immer die oberste Stufe der methodischen Leiter der Evidenz und Beobachtungsstudien die unteren Stufen besetzen (Petticrew & Roberts, 2003). Zudem entsteht vermehrt Konsens darüber, dass quantitative Untersuchungen nur dann Sinn machen und einen wirklichen Nutzen für die jeweiligen Zielgruppen generieren können, wenn diese in qualitative Forschung und Erkenntnisse eingebettet sind (Grypdonck, 2004). Es gab bereits mehrere Versuche, auch die Pflegeforschung in eine eigene Evidenzhierarchie zu überführen. So haben z. B. Lavin et al. (2002) eine pflegespezifische Einordnung und deren Notwendigkeit zur Beachtung der zugrunde liegenden Forschungsfrage(n) basierend auf den Unterschieden zwischen Evidence-Based Nursing und Evidence-Based Medicine herausgearbeitet. Ein weiterer Ansatz zur Definition einer Evidenzhierarchie im Kontext

Thieme RECOM

der Pflegeforschung stammt von Evans (2003). Er betrachtet drei Dimensionen: die Wirksamkeit, die Angemessenheit und die Durchführbarkeit von Interventionen im Gesundheitswesen. Hier werden etwa Beobachtungsstudien ebenfalls wie RCTs auf der Stufe „Gut“ vorgeschlagen.

Level of Evidence	Wirksamkeit	Angemessenheit	Durchführbarkeit
Exzellent	Systematische Übersichtsarbeiten Multizentrische Studien	Systematische Übersichtsarbeiten Multizentrische Studien	Systematische Übersichtsarbeiten Multizentrische Studien
Gut	RCT Beobachtungsstudien	RCT Beobachtungsstudien Interpretative Studien*	RCT Beobachtungsstudien Interpretative Studien*
Mäßig/ moderat	Nicht-kontrollierte Studiendesigns mit ausgeprägten Ergebnissen Vorher-Nachher-Studien Nicht randomisierte kontrollierte Studien	Deskriptive Studien Fokusgruppen	Deskriptive Studien Interventionsstudien Vorher-Nachher Studien Fokusgruppen
Niedrig/ schlecht	Beschreibende Studien Fallserien Expert(inn)enmeinung Studien von schlechter methodologischer Qualität	Expert(inn)enmeinung Fallserien Studien von schlechter methodologischer Qualität	Expert(inn)enmeinung Fallserien Studien von schlechter methodologischer Qualität

Tabelle 13: Evidenzhierarchie nach Evans zur Evaluationen von Interventionen im Gesundheitswesen (Quelle: eigene Darstellung in Anlehnung an Evans, 2003)

* **Interpretative Studien:** Hierunter fallen beispielsweise Dokumentationsanalysen, Narrationsanalysen, Konversationsanalysen, Inhaltsanalysen, hermeneutische Methoden, Phänomenologie, Grounded Theory

Es wird in Auseinandersetzung mit der Thematik schnell deutlich, dass neben den bisher aufgezeigten Überlegungen auch die Durchführungsqualität einer publizierten Studie von hoher Bedeutung zur Einschätzung deren Qualität ist (Behrens & Langer, 2016). So ist ein RCT oder eine systematische Übersichtsarbeit nicht per se mit einer hohen Qualität einzustufen. Die nachfolgende Tabelle gibt eine grobe Übersicht über häufige systematische Fehlerquellen bei Studien, welche die Bewertung der Evidenz beeinflussen und mindern können.

Häufige systematische Fehlerquellen bei der Durchführung von Studien	
Fehlerquelle	**Erklärung**
Fallzahl zu klein	Bei zu kleinen Stichproben können vorhandene Unterschiede evtl. nicht nachgewiesen werden.
Ungenaue Fragestellung oder fehlende Theorieableitung	Es werden ungenaue Fragen gestellt oder es fehlt das theoretische Einbetten in einen Rahmen. Abhilfe: Theorierahmen ggf. PIKE-Fragenschema
Fehlende Messwerte	Problem z. B. bei Meta-Analysen, bei denen ggf. geschätzte Werte eingesetzt werden, um weitere Berechnungen möglich zu machen, Messungen sind wegen inhomogener Studienlage nicht möglich.
Keine angemessene Auswahl der Messinstrumente	Gerade bei Interventionsstudien geht es darum, den Outcome/die Veränderung zu ermitteln. Werden hier ungenaue oder nicht geeignete Instrumente eingesetzt, um die Effekte zu messen kann es sein, dass etwas anderes gemessen wird als ursprünglich intendiert. Abhilfe: ggf. qualitative Forschung, um geeignete Messinstrumente zu entwickeln.
Störgröße, Störfaktor (confounder)	Parameter oder Risikofaktor, der sowohl mit der Exposition/Intervention als auch mit der Zielgröße (z. B. Erkrankung) assoziiert ist. Bsp.: Alter, Geschlecht, Nikotinkonsum, zusätzliche Erkrankungen

Thieme RECOM

	Abhilfe: im experimentellen Studiendesign z. B. Randomisierung, in Datenanalyse z. B. Stratifizierung
Schwankungen im Krankheitsverlauf	Besserung oder Verschlechterung im Krankheitsverlauf, unabhängig von der Intervention, beeinflusst durch nichterfasste Variablen/Einflüsse z. B. spontane Besserung von Erkältungskrankheiten, Fortschreiten eines Diabetes, Schübe bei Multipler Sklerose oder Rheuma Abhilfe: Kontrollgruppe
Selektionsbias (selection bias)	Systematische Unterschiede in der Auswahl der Proband(inn)en oder der Zusammensetzung der Gruppen, z. B. Männer vs. Frauen, Patient(inn)en mit leichten vs. schweren Symptomen Abhilfe: Randomisierung, verdeckte Zuordnung
Behandlerbias (performance bias)	Systematische Unterschiede zwischen den Gruppen bezüglich z. B. Pflege, Aufmerksamkeit etc. Abhilfe: Verblindung der Proband(inn)en und Behandler(innen)
Erinnerungsbias (recall bias)	Erinnerungsbedingte Verzerrung, z. B. in retrospektiven Studien: Probanden, deren Erinnerung an eine mögliche Exposition ungenau ist, geben diese eher an, wenn sie erkrankt sind als wenn nicht. Weitere Beispiele können ungünstige, lückenhafte Datenlage bei der Dokumentenanalysen sein.
Beobachterbias (observer / ascertainment bias)	Systematische Unterschiede in der Bewertung, weil die Gruppenzuordnung nicht (ausreichend) verblindet ist.
Informationsbias (detection bias)	Systematische Unterschiede zwischen den Gruppen in der Bestimmung des Outcomes, z. B.: Hormonbehandlung bei Frauen führt zu mehr Arztbesuchen (z. B. wegen Blutungen), wodurch auch mehr Gebärmutter-Karzinome gefunden werden – obwohl Hormone keine Karzinome begünstigen.
Abweichungen vom Studienplan (attrition bias, drop out, loss to follow up)	Systematische Unterschiede zwischen den Gruppen bei Abweichungen vom Studienprotokoll, z. B. wenn in der Interventionsgruppe mehr Proband(inn)en die Studie wegen Nebenwirkungen abbrechen als in der Placebogruppe. Abhilfe: Intention-to-treat-Analyse (ITT)
Publikationsbias (publication / reporting bias)	Systematische Unterschiede zwischen veröffentlichten und unveröffentlichten Ergebnissen, weil z. B. signifikante, erwünschte und spektakuläre Ergebnisse häufiger, schneller und hochrangiger publiziert werden als andere. Abhilfe: Nutzung der entsprechenden Leitlinien zur Veröffentlichung und Transparenz des Forschungsprozesses

Tabelle 14: Häufige systematische Fehlerquellen bei der Durchführung von Studien (Quelle: eigene Darstellung in Anlehnung an Grypdonck, 2004; Rojahn, 2016)

In den bisherigen Auseinandersetzungen wird deutlich, dass die Einstufung einer Studie bzw. deren Veröffentlichung in eine Evidenzhierarchie ein komplexes Konstrukt ist und nicht allein auf das Studiendesign reduziert werden darf. Darüber hinaus ist die Frage nach der Angemessenheit des Studiendesigns für die Glaubwürdigkeit der Evidenz kritisch zu hinterfragen, besonders vor dem Hintergrund der zahlreichen denkbaren Fehlerquellen. Gerade im Kontext der Pflegeforschung ist immer auch die methodische Frage der Eignung des Forschungsdesigns zur Beantwortung der Forschungsfrage(n) bedeutend (Mayring, 2017; Petticrew & Roberts, 2003). Mehrere Autor(inn)en empfehlen infolgedessen, die Bewertung und Einstufung der Evidenz einer Forschungsarbeit vor dem Hintergrund eben genau der jeweiligen Forschungsfrage(n) zu realisieren (Baker et al., 2010; Bartholomeyczik et al., 2008; Petticrew & Roberts, 2003). Anstelle der Einstufung in starren, hierarchischen Evidenzstufen sollte daher abhängig von den Forschungsfrage(n) und dem methodologischen Pluralismus die Nutzung einer diese Umstände berücksichtigenden Evidenzmatrix bevorzugt werden (Petticrew & Roberts, 2003). Ein anschauliches Beispiel ist aus dem Diskurs in der psychiatrischen Pflege zu entnehmen. Eine beträchtliche Menge an Wissen und Evidenz wird hier aus Einzelfallstudien abgeleitet (Stickley & Phillips, 2008). Die Übertragbarkeit starrer Evidenzhierarchien in die psychiatrische Praxis ist vor dem Hintergrund der Individualität versus einer Generalisierbarkeit kritisch zu hinterfragen und könnte dazu führen, dass vorrangig Handlungskonzepte basierend auf den mutmaßlich höchsten Evidenzstufen umgesetzt werden, jedoch im individuellen Versorgungskontext oftmals ebenso wirksame oder gar bessere Konzepte keine

Thieme RECOM

Beachtung finden. Die aufgeführten Argumente sind ein weiterer Grund, weswegen im Rahmen der ENP-Entwicklung die Evidenzgrade zu einzelnen pflegerischen Handlungskonzepten nicht explizit ausgewiesen werden.

2.2.3.3 Evidenzbewertung im Kontext der ENP-Entwicklung in Anlehnung an das GRADE-Schema

Nach dem Critical Appraisal unter Zuhilfenahme der jeweiligen Bewertungsinstrumente wird die jeweilige Publikation bzw. Forschungsarbeit gesamthaft bezüglich des Evidenzgrades eingestuft. Ziel dabei ist es, im Rahmen der Neu- bzw. Weiterentwicklungsarbeit einer ENP-Praxisleitlinie eine Gesamtschau über die Qualität der verwendeten Referenzen zu erhalten und damit die Gültigkeit dieser einzuschätzen. Der Evidenzgrad einer Referenz spiegelt das Ausmaß des Vertrauens in die Korrektheit und Zuverlässigkeit einer/mehrerer Aussagen bezüglich pflegediagnostischer Zusammenhänge bzw. Schätzungen eines Effektes bei pflegerischen Interventionskonzepten wider. Das kumulierte Ergebnis einer Gesamtbewertung wird in der ENP-Evidenztabelle eingetragen. Anspruch bei der ENP-Entwicklung ist es, wo immer möglich Referenzen heranzuziehen, die einen hohen Evidenzgrad vorweisen können, um eine hohe Zuverlässigkeit der pflegediagnostischen Aussagen und deren zugehöriger Elemente (Kennzeichen, Ursachen und Ressourcen) sowie zugeordneter Handlungskonzepte erreichen zu können. Die gesamthafte Bewertung und Einschätzung der Studie bzw. Publikation auf ihre Güte hin ist eine bereits etablierte Praxis bei der Leitlinienentwicklung (Platz, 2021). Auch finden sich bei vielen Instrumenten zum Critical Appraisal Items oder Methoden zur Gesamtbewertung einer Veröffentlichung. Einige dieser Ansätze wurden bereits bei den vorgestellten Bewertungsinstrumenten aufgeführt (vgl. Kapitel 2.2.1). Die Kriterien oder Fragen zur Gesamtbeurteilung sind allerdings unterschiedlich operationalisiert, was einen Gesamtüberblick über die Güte der zur ENP-Entwicklung genutzten Publikationen erschwert.

Vor diesem Hintergrund wird im Rahmen der ENP-Entwicklung eine einheitliche Systematik zur zusammenfassenden Evidenzbewertung einer Publikation angestrebt und umgesetzt. Nachstehendes Beispiel aus der Weiterentwicklung einer ENP-Praxisleitlinie zur Fatigue, bei der insgesamt über 50 relevante Studien gesichtet und bewertet wurden, soll den Nutzen hiervon verdeutlichen. Durch die in der Tabelle dargestellte, gebündelte Gesamtbewertung lässt sich rasch und übergreifend die Qualität der Forschungslage erfassen.

Studientyp	Evidenzbewertung in Anlehnung an GRADE	Quelle
SR	●●●●	Patterson, E., Wan, Y. W. T., & Sidani, S. (2013). Nonpharmacological nursing interventions for the management of patient fatigue: a literature review. Journal of Clinical Nursing (John Wiley & Sons, Inc.), 22(19-20), pp. 2668-2678. doi: 10.1111/jocn.12211
RCT	●●●○	Zick, S. M. et al. (2016). Investigation of 2 Types of Self-administered Acupressure for Persistent Cancer-Related Fatigue in Breast Cancer Survivors: A Randomized Clinical Trial. JAMA Oncol, 2(11), pp. 1470-1476. doi: 10.1001/jamaoncol.2016.1867
RCT	●●○○	Zick, S. M. et al. (2011). Relaxation acupressure reduces persistent cancer-related fatigue. Evid Based Complement Alternat Med, 2011. doi: 10.1155/2011/142913

Tabelle 15: Exemplarischer Auszug aus der Gesamtübersicht der verwendeten Publikationen im Rahmen einer ENP-Weiterentwicklungsarbeit

Mit Blick auf die Systematik und Methodik zur zusammenfassenden Evidenzbewertung für eine Publikation wird das sog. GRADE-Schema (Grading of Recommendations, Assessment, Development and Evaluation)

herangezogen (Kunz, Burnand, & Schünemann, 2008; Schünemann et al., 2013). Das Bewertungsurteil erfolgt anhand eines vierstufigen Punkteschemas.

Orientierung an GRADE

Im Rahmen der GRADE-Qualitätsbewertung werden die RCTs als glaubwürdiger und robuster gegenüber einem Risk of Bias betrachtet als andere Forschungsmethoden, daher startet dieses Design zunächst auf einer höheren Evidenzstufe als beispielswiese Beobachtungsstudien. Allerdings kann die Ausgangseinstufung eines bestimmten Studiendesigns durch verschiedene Faktoren sowohl eine Hoch- als auch eine Abstufung erfahren (vgl. hierzu exemplarisch die Resultate für die beiden RCTs in Tabelle 15). Die GRADE-Qualitätsbewertung ist ursprünglich auf einen Endpunkt im Rahmen der Entwicklung von Leitlinien ausgerichtet. Auf Grund dieser Tatsache wurden für die Nutzung von GRADE im Kontext der ENP-Entwicklung einige Anpassungen vorgenommen. Zum einen wird kein konkreter Endpunkt, sondern die Güte der vorliegenden Publikation bewertet. Eine ähnliche Vorgehensweise wurde bereits erfolgreich von anderen Experten genutzt (Panfil et al., 2011). Die Qualität der Evidenz wird bei GRADE in vier Stufen bzw. Graden untergliedert. Nachstehende Tabelle stellt zunächst die ins Deutsche übersetzte, inhaltlich jedoch unveränderten GRADE-Evidenzlevel aus dem offiziellen Handbuch zum Instrument vor.

Qualität der Evidenz	
Level	Definition
Hoch	Wir sind sehr zuversichtlich, dass der tatsächliche Effekt nahe an der Schätzung des Effekts liegt.
Mäßig	Wir haben mäßiges Vertrauen in die Effektschätzung: Der tatsächliche Effekt liegt wahrscheinlich in der Nähe des geschätzten Effekts, aber es besteht die Möglichkeit, dass er erheblich davon abweicht.
Niedrig	Unser Vertrauen in die Effektschätzung ist begrenzt: Der tatsächliche Effekt kann sich erheblich von der Schätzung des Effekts unterscheiden.
Sehr niedrig	Wir haben sehr wenig Vertrauen in die Effektschätzung: Der tatsächliche Effekt unterscheidet sich wahrscheinlich erheblich von der Schätzung des Effekts.

Tabelle 16: Evidenzlevel nach GRADE (Quelle: eigene Darstellung und Übersetzung nach Schünemann et al., 2013)

Für den pflegespezifischen Fokus im Rahmen der ENP-Entwicklung wurde das Wording der vier Evidenzlevel der GRADE-Evidenzstufen adaptiert, wie folgender Tabelle entnehmbar ist. Der jeweilige Evidenzlevel für eine Referenz wird mit einer vierstufigen Symbolik ausgedrückt.

Evidenzlevel	**Symbolik**	**Definition**
Hohe Qualität	++++ oder ● ● ● ●	Es ist sehr unwahrscheinlich, dass weitere Forschungsarbeiten das Vertrauen in den beobachteten pflegediagnostischen Zusammenhang und/oder beobachteten Behandlungseffekt verändern. Es ist sehr wahrscheinlich, dass der wahre Effekt nahe bei der Effektschätzung liegt.
Moderate Qualität	+++ oder ● ● ● ○	Weitere Forschung wird sich vermutlich auf das Vertrauen in den beobachteten pflegediagnostischen Zusammenhang und/oder beobachteten Behandlungseffekt auswirken. Es besteht moderate Zuversicht, dass der wahre Effekt nahe an der Effektschätzung liegt. Es besteht aber die Möglichkeit, dass dieser abweichen könnte.
Niedrige Qualität	++ oder ● ● ○ ○	Weitere Forschungen führen sehr wahrscheinlich zu Änderungen im Bereich des beobachteten pflegediagnostischen Zusammenhangs und/oder beobachteten Behandlungseffekts. Das Vertrauen in die Effektschätzung ist begrenzt und der wahre Effekt kann wesentlich abweichen.
Sehr niedrige Qualität	+ oder ● ○ ○ ○	Der beobachtete pflegediagnostische Zusammenhang und/oder Behandlungseffekt ist mit großer Unsicherheit behaftet. Es besteht wenig Vertrauen in die Effektschätzung.

Abbildung 10: Übersetzte und adaptierte Evidenzlevel zur Nutzung in der ENP-Entwicklung in Anlehnung an GRADE (Quellen: Kunz et al., 2008; Platz, 2021; Schünemann et al., 2013)

GRADE definiert, wie andere Systematiken auch eine Evidenzhierarchie in Bezug auf das jeweils gegebene Studiendesign, die als Ausgangspunkt für die folgenden Einstufungsentscheidungen im Bewertungsprozess dient. In der Entwicklergruppe von GRADE wurden Kriterien definiert, die zu einer Höherstufung oder Herabstufung des Evidenzlevels führen können. Auch hier wurden für den Kontext der Pflegeforschung punktuelle Anpassungen vorgenommen. Konkret wurden die qualitativen Forschungsdesigns abhängig von der jeweiligen Fragestellung zur Ausgangsbewertung des Evidenzlevels hinzugefügt und die Kriterien zum Herauf- oder Abstufen der Studien um die Besonderheiten qualitativer Forschungsarbeiten ergänzt.

Definierte Ausgangslage der Forschungsarbeiten in Abhängigkeit des Studiendesigns und der Fragestellung(en)

In der nachfolgenden Tabelle ist die Ausgangseinstufung für die zusammenfassende Evidenzbewertung im Rahmen der ENP-Entwicklung festgelegt und einsehbar. Diese Aufstellung ergänzt das ursprüngliche GRADE-Schema in einigen Punkten. Auf der Ebene der Pflegemaßnahmen wurde die Evidenzhierarchie sowie die Auf- bzw. Abstufungskriterien von GRADE übernommen. Für die Entwicklung von Pflegediagnosen wurden hingegen Ergänzungen vorgenommen, da es im Rahmen der ENP-Praxisleitlinienentwicklung nicht nur um Wirkungszusammenhänge von Pflegemaßnahmen geht, sondern auch um die Entwicklung. Die Ergänzungen orientieren sich an der der Diskussion über das Ranking von Evidenz im spezifischen Kontext der Pflege (Behrens & Langer, 2004; Evans, 2003) sowie der Pflegediagnostik und pflegerischen Behandlungspfaden (Grieser, Hegedüs, & Kunz, 2011). Die Ausgangsbewertung einer Studie ist abhängig vom jeweiligen Forschungsdesign sowie der Überprüfung der Wirksamkeit von Pflegemaßnahmen oder der Bestätigung einer Pflegediagnose mit Kennzeichen/Symptomen bzw. der Frage nach beeinflussenden Faktoren und Ursachenzusammenhängen unterschiedlich. Wie bereits ausgeführt, macht es aus mehreren Gründen heraus Sinn, die Evidenzhierarchien im Kontext der jeweiligen Forschungsfrage(n) festzulegen. Für die ENP-Entwicklung wurde daher die initiale Einstufung der Evidenzlevel von GRADE an den Forschungskontext von ENP adaptiert.

Evidenzkategorie	Vertrauenswürdigkeit des Studiendesigns bezogen auf die Wirksamkeits-nachweise von Pflegemaßnahmen	Vertrauenswürdigkeit des Studiendesigns bezogen auf das pflegediagnostische Phänomen sowie Kennzeichen	Vertrauenswürdigkeit des Studiendesigns bezogen auf beeinflussende Faktoren/Ursachen einer Pflegediagnose
Hohe Vertrauenswürdigkeit	• Systematische Übersichtsarbeiten oder Meta-Analysen mit RCTs • Multizentrische Studien • Leitlinien	• Systematische Übersichtsarbeiten, Meta-Analysen mit qualitativen und/oder quantitativen nicht randomisierten Forschungsarbeiten • Leitlinien • Multizentrische Studien	• Systematische Übersichtsarbeiten, Meta-Analysen mit qualitativen und/oder quantitativen nicht randomisierten Forschungsarbeiten • Multizentrische Studien
Moderate Vertrauenswürdigkeit	• RCTs • Kohorten-Studien	• Methoden-Triangulation • Beobachtungsstudien • Qualitative Forschungsarbeiten • Interpretative Studien*	• RCTs • Kohorten-Studien • Fall-Kontroll-Studien
Niedrige Vertrauenswürdigkeit	• Beobachtungsstudien • Quasiexperimentelle Studien ohne Randomisierung (non-	• Nicht-analytische Studien: Fallserien, Fallberichte • Nicht experimentelle deskriptive Studien	• Querschnittsstudien • Prävalenz-/Inzidenzstudien • Korrelationsstudie

Thieme RECOM

	randomized controlled trial) • Übersichtsarbeiten / Meta-Analysen von Fall-Kontroll- oder Kohortenstudien • Vorher-Nachher-Studien • Kontrollierte klinische Studien, z. B. Interventionsstudien (Cross-Over-Studien), Vergleichsstudien	• Delphi-Methode	
Sehr niedrige Vertrauenswürdigkeit	• Nicht experimentelle deskriptive Studien • Nicht-analytische Studien: Fallserien, Fallberichte • Qualitative Forschungsarbeiten • Expert(inn)enmeinung von anerkannten Autoritäten (Fachbuchebene z. B. anerkannte Pflegelehrbuchliteratur) • Delphi-Methode	• Fallserien, Fallberichte • Expert(inn)enmeinung von anerkannten Autoritäten (Fachbuchebene z. B. anerkannte Pflegelehrbuchliteratur)	• Fallserien, Fallberichte • Expert(inn)enmeinung von anerkannten Autoritäten (Fachbuchebene z. B. anerkannte Pflegelehrbuchliteratur) • Delphi-Methode

Tabelle 17: Initiale Einstufung des zusammenfassenden Evidenzlevels im Rahmen der ENP-Entwicklung in Anlehnung an GRADE

* **Interpretative Studien:** Hierunter fallen beispielsweise Dokumentationsanalysen, Narrationsanalysen, Konversationsanalysen, Inhaltsanalysen, hermeneutische Methoden, Phänomenologie oder die Grounded Theory.

Evidenzmindernde oder -erhöhende Aspekte im ENP-Forschungskontext

Nachfolgend werden die verschiedenen Faktoren vorgestellt (adaptiert und in Anlehnung an GRADE), die zu einer Verminderung oder Erhöhung bei der zusammenfassenden Bewertung des Evidenzlevels führen können. Jede Minderung oder Erhöhung des Evidenzlevels spiegelt keine starren Kategorien wider, sondern ein Kontinuum innerhalb jeder Kategorie. Ist die Beweislage in Bezug auf einen bestimmten Faktor nur moderat, hängt die Entscheidung darüber, ob eine Studie/eine Referenz über oder unter den Schwellenwert für die Minderung oder Erhöhungen der Qualität (um einen oder mehrere Faktoren) fällt, von der Beurteilung der Rater(innen) ab. Hierbei geht es auch um die Einschätzung der Bedeutung des jeweiligen Faktors auf das gesamte Ergebnis (Schünemann et al., 2013).

Faktoren, die den Evidenzgrad mindern können	
Faktor	**Folge**
Einschränkungen im Studiendesign oder in der Durchführung (hohes Risiko einer Verzerrung), z. B. Fehlende Zuordnungsverschleierung Unzureichende Verblindung Unzureichendes Follow-up Selektive Ergebnisberichterstattung Messinstrument ungeeignet zur Erhebung von Messwerten Unklare „Ein- und Ausschlusskriterien“ von Studienteilnehmer(inne)n und/oder Studien Hoher Ausfall von Teilnehmer(inne)n Verzerrung aufgrund von Confounding Keine fach- und fragenstellungsangemessene Vorgehensweise Unklare/ungenaue Interventionsbeschreibung (Hoffmann et al., 2016) Unterschiede in den Untersuchungsbedingungen Geringe Anzahl an Studienteilnehmer(inne)n, breites Konfidenzintervall (CI)	↓ 1 oder 2 Stufen

Thieme RECOM

Inkonsistenz der Ergebnisse	↓ 1 oder 2 Stufen
Indirektheit der Beweise	↓ 1 oder 2 Stufen
Ungenauigkeit z. B. Fehlende Angaben zur Entstehung von Daten, Gründe für Ausschlüsse usw.	↓ 1 oder 2 Stufen
Publikationsbias, z. B. Selektive Auswahl von Studien bei der Veröffentlichung Selektive Ergebnisberichterstattung	↓ 1 oder 2 Stufen
Fehlender Theorierahmen oder unklare wissenschaftstheoretische Verortung, fehlende Offenlegung von Vorannahmen	↓ 1 oder 2 Stufen
Faktoren, welche die Qualität der Evidenz erhöhen können	
Faktor	**Folge**
Gut beschriebene(s) und schlüssig hergeleitete(s) Studiendesign und Durchführung (niedriges Risiko einer Verzerrung) Besonders transparente, plausible und theoriegestützte Begründung für fach- und fragengestützte Entscheidungen im Forschungsdesign Plausible Auswahl der Teilnehmer(inne)n mit großen Stichproben, die eine Sättigung erwarten lässt Plausible Konstanthaltung der Versuchsbedingungen Transparente Berichterstattung von Studienabbrecher(inne)n UND geringer Anteil dieser (intention-to-treat-Analyse) mit statistischer Effektangabe Methoden-Triangulation verbunden mit homogenen Ergebnissen Fundierte theoretische Ableitung mit passendem Forschungsdesign	↑ 1 Stufe
Große Wirkungsstärke/hohe Effektstärke, z. B. Enges Konfidenzintervall bei RCTs	↑ 1 oder 2 Stufen
Alle plausiblen Störfaktoren würden den nachgewiesenen Effekt verringern oder den Effekt verstärken. Daher kann, wenn kein Effekt beobachtet wurde, dieses Risiko ausgeschlossen werden	↑ 1 Stufe

Tabelle 98: Faktoren, die den zusammenfassenden Evidenzgrad einer Studie/Publikation mindern bzw. erhöhen können (Quelle: eigene und adaptierte Darstellung in Anlehnung an Schünemann et al., 2013)

2.3 Der weitere Verlauf des ENP-Entwicklungsprozesses

Im Rahmen der ENP-Entwicklung werden grundsätzlich alle Bearbeitungsvorschläge, die im Vorfeld von einer hauptverantwortlichen Person erarbeitet werden, abschließend im Team konsentiert. Bei den Konsentierungssitzungen liegt allen teilnehmenden Personen das Protokoll zur Literaturrecherche, die Dokumente zum Critical Appraisal sowie die Evidenztabelle und Originalpublikationen vor. Stichprobenartig werden 10 % der Referenzen per Zufallsstichprobe durch eine zweite Person bewertet. Zeigen sich hier hohe Übereinstimmungen von 80 % oder mehr, werden die weiteren Studienbewertungen alleinig von der hauptverantwortlichen Person für die Bearbeitung des zuvor im Team festgelegten Themas durchgeführt. Die Literatursuche und die Bewertung der Studien werden punktuell über den gesamten Entwicklungsprozess hinweg in wöchentlichen Meetings reflektiert. Jeder erarbeitete Änderungsvorschlag zur Umsetzung in ENP wird am Ende des Bearbeitungsprozesses in einer Sitzung diskutiert und gemeinsam vereinbart. Lässt sich zu einem Aspekt nicht unmittelbar ein Konsens finden, wird die Literatur erneut zu Rate gezogen und/oder auch andere Expert(inn)en aus der Pflegewissenschaft und -praxis mit ausgewiesener Expertise befragt, um weitere Belege zu finden.

Weiterführende Informationen zum systematischen Weiterentwicklungs- und Validierungsprozess von ENP lassen sich auch in den „Wissenschaftlichen Hintergründen“ nachlesen, die in ihrer jeweils aktuellen Fassung online unter https://www.recom.eu/wiss-hintergruende abrufbar sind.

Literatur

Aromataris, E. & Munn, Z. (2020). JBI Manual for Evidence Synthesis. Retrieved from https://synthesismanual.jbi.global/ (Accessed: 13.07.2022).

Baker, A. et al. (2010). A review of grading systems for evidence-based guidelines produced by medical specialties. *Clin Med (Lond), 10*(4), pp. 358-363. doi: 10.7861/clinmedicine.10-4-358

Bartholomeyczik, S. et al. (Eds.). (2008). *Lexikon der Pflegeforschung*. München, Jena: Urban & Fischer Verlag.

Behrens, J. & Langer, G. (2004). *Evidence-based Nursing. Vertrauensbildende Entzauberung der Wissenschaft* (1. ed.). Bern [u.a.]: Huber.

Behrens, J. & Langer, G. (2016). *Evidence-based Nursing and Caring. Methoden und Ethik der Pflegepraxis und Versorgungsforschung* (4. ed.). Bern: Hogrefe.

Browman, G. P. et al. (2017). Appraisal of Guidlines for research & evaluation II. AGREE II Instrument. Update 2017. Retrieved from https://www.agreetrust.org/wp-content/uploads/2017/12/AGREE-II-Users-Manual-and-23-item-Instrument-2009-Update-2017.pdf (Accessed: 12.01.2022).

Burgers, J. S. et al. (2004). International Assessment of the Quality of Clinical Practice Guidelines in Oncology Using the Appraisal of Guidelines and Research and Evaluation Instrument. *Journal of Clinical Oncology, 22*(10), pp. 2000-2007. doi: 10.1200/jco.2004.06.157

Chen, Y. et al. (2017). Ein Instrument zur Erstellung von Leitlinienberichten: das RIGHT-Statement. *Z Evid Fortbild Qual Gesundhwes, 127-128*, pp. 3-10. doi: 10.1016/j.zefq.2017.10.008

Chiappini, E. et al. (2017). Guidelines for the symptomatic management of fever in children: systematic review of the literature and quality appraisal with AGREE II. *BMJ Open, 7*(7), pp. e015404. doi: 10.1136/bmjopen-2016-015404

Cochrane Deutschland (Ed.) et al. (2021). Manual zur Bewertung des Biasrisikos in Interventionsstudien. Retrieved from https://www.leitlinien.de/methodik/pdf/manual-bewertung-v2-0.pdf (Accessed: 10.02.2022).

Cuschieri, S. (2019a). The CONSORT statement. *Saudi J Anaesth, 13*(Suppl 1), pp. S27-s30. doi: 10.4103/sja.SJA_559_18

Cuschieri, S. (2019b). The STROBE guidelines. *Saudi J Anaesth, 13*(Suppl 1), pp. S31-s34. doi: 10.4103/sja.SJA_543_18

Datta, S., Funnell, K., & Ramuscak, N. (2012). Improving the Sample Size Calculation Process for Peel Health: A Rapid Review of the Evidence. Retrieved from https://www.peelregion.ca/health/library/pdf/Rapid-Review-Sample-Size-Report.pdf (Accessed: 15.02.2022).

Dicenso, A., Bayley, L., & Haynes, R. B. (2009). Accessing pre-appraised evidence: fine-tuning the 5S model into a 6S model. *Evid Based Nurs, 12*(4), pp. 99-101. doi: 10.1136/ebn.12.4.99-b

Dietrich, S. K. et al. (2016). Effectiveness of sleep education programs to improve sleep hygiene and/or sleep quality in college students: a systematic review. *JBI Database System Rev Implement Rep, 14*(9), pp. 108-134. doi: 10.11124/jbisrir-2016-003088

Dippmann, A. K. et al. (2013). DELBI 2.0. Aktuelle Überarbeitung des Deutschen Instruments zur methodischen Leitlinienbewertung und Validierungskonzept. Präsentation auf der 14. Jahrestagung des Deutschen Netzwerks Evidenzbasierte Medizin e.V. am. Retrieved from https://www.aezq.de/mdb/edocs/pdf/delbi/delbi-2.0-vortrag-ebm-kongress.pdf (Accessed: 08.03.2022).

Dreier, M. et al. (2010). Vergleich von Bewertungsinstrumenten für die Studienqualität von Primär- und Sekundärstudien zur Verwendung für HTA-Berichte im deutschsprachigen Raum (HTA Bericht 102). Retrieved from https://portal.dimdi.de/de/hta/hta_berichte/hta260_bericht_de.pdf (Accessed: 01.02.2022).

Durieux, N., Vandenput, S., & Pasleau, F. (2013). [OCEBM levels of evidence system]. *Rev Med Liege, 68*(12), pp. 644-649.

Evans, D. (2003). Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs, 12*(1), pp. 77-84. doi: 10.1046/j.1365-2702.2003.00662.x

Fringer, A. (2018). Triangulation, Methodenkombination und Mixed Methods Research. Retrieved from https://www.researchgate.net/profile/Andre-Fringer/publication/323959507_Triangulation_Methodenkombination_und_Mixed_Methods_Research/links/5ab4d11c0f7e9b4897c83333/Triangulation-Methodenkombination-und-Mixed-Methods-Research.pdf (Accessed: 09.02.2022).

Fringer, A. & Schrems, B. (2018). Qualitative Designs. In H. Brandenburg et al. (Eds.), *Pflegewissenschaften 2, Lehr- und Arbeitsbuch zur Einführung in die Methoden der Pflegeforschung* (3 ed.). Bern: Hogrefe Verlag.

Gastinger, I. et al. (2002). Bedeutung prospektiver multizentrischer Beobachtungsstudien für den Erkenntnisgewinn in der Chirurgie. *Der Chirurg, 73*(2), pp. 161-166.

Grieser, M., Hegedüs, A., & Kunz, S. (2011). Evidenzbasierte Pflegeinterventionen für die Pflegediagnose Angst bei depressiven Erkrankungen. *Psych Pflege, 17*, pp. 321-328.

Grypdonck, M. (2004). Eine Kritische Bewertung von Forschungsmethoden zur Herstellung von Evidenz in der Pflege. *Pflege & Gesellschaft, 19*(2), pp. 35-41.

Haile, Z. T. (2021). Critical Appraisal Tools and Reporting Guidelines. *Journal of Human Lactation*, pp. 08903344211058374. doi: 10.1177/08903344211058374

Hannes, K. (2011). Chapter 4: Critical appraisal of qualitative research. In J. Noyes et al. (Eds.), *Supplementary Guidance for Inclusion of Qualitative Research in Cochrane Systematic Reviews of Interventions*: Cochrane Collaboration Qualitative Methods Group. Retrieved from http://cqrmg.cochrane.org/supplemental-handbook-guidance (Accessed: 04.02.2022).

Harbour, R. & Miller, J. (2001). A new system for grading recommendations in evidence based guidelines. *Bmj, 323*(7308), pp. 334-336. doi: 10.1136/bmj.323.7308.334

Harrison, R. et al. (2021). Quality assessment with diverse studies (QuADS): an appraisal tool for methodological and reporting quality in systematic reviews of mixed- or multi-method studies. *BMC Health Services Research, 21*(1), pp. 144-163. doi: 10.1186/s12913-021-06122-y

Hoffmann, T. C. et al. (2016). Die TIDieR Checkliste und Anleitung – ein Instrument für eine verbesserte Interventionsbeschreibung und Replikation. *Gesundheitswesen, 78*(03), pp. 175-188.

Hong, Q. N. et al. (2018). MIXED METHODS APPRAISAL TOOL (MMAT) - VERSION 2018. User guide. Retrieved from http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/146002140/MMAT_2018_criteria-manual_2018-08-08c.pdf (Accessed: 04.02.2022).

Hopper, S. I. et al. (2019). Effectiveness of diaphragmatic breathing for reducing physiological and psychological stress in adults: a quantitative systematic review. *JBI Database System Rev Implement Rep, 17*(9), pp. 1855-1876. doi: 10.11124/jbisrir-2017-003848

Jimenez, E. et al. (2018a). Mixing and matching: using qualitative methods to improve quantitative impact evaluations (IEs) and systematic reviews (SRs) of development outcomes. *Journal of Development Effectiveness, 10*(4), pp. 400-421. doi: 10.1080/19439342.2018.1534875

Jimenez, E. et al. (2018b). Mixing and Matching: Using Qualitative Methods to Improve Quantitative Impact Evaluations (IEs) and Systematic Reviews (SRs) of Development Outcomes. CEDIL Iception Paper 5. Retrieved from https://cedilprogramme.org/wp-content/uploads/2018/08/Inception-Paper-No-5-1.pdf (Accessed: 07.02.2022).

Joanna Briggs Institute (Ed.). (2017). The Joanna Briggs Institute Critical Appraisal tools for use in JBI Systematic Reviews. Checklist for Text and Opinion. Retrieved from https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Text_and_Opinion2017_0.pdf (Accessed: 16.02.2022).

Kelle, U. (2019). Mixed Methods. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer, pp. 159-172.

Knappertsbusch, F., Langfeldt, B., & Kelle, U. (2021). Mixed-Methods and Multimethod Research. In H. Betina et al. (Eds.), *Soziologie - Sociology in the German-Speaking World*. Oldenbourg: De Gruyter, pp. 261-272.

Knippschild, S. et al. (2015). Das CONSORT-Statement zur standardisierten Berichterstattung Randomisierter Klinischer Prüfungen–Evidenz durch Transparenz. *Z Zahnärztl Impl, 31*, pp. 64-78.

Kopp, I. et al. (2008). Deutsches Instrument zur methodischen Leitlinien-Bewertung (DELBI). *Fassung 2005/2006 + Domäne 8 (2008)*. Retrieved from https://www.leitlinien.de/hintergrund/pdf/delbi-fassung-2005-2006-domaene-8-2008.pdf (Accessed: 18.07.2022).

Kunz, R., Burnand, B., & Schünemann, H. J. (2008). Das GRADE-System. *Der Internist, 49*(6), pp. 673-680. doi: 10.1007/s00108-008-2141-9

Lavin, M. A. et al. (2002). Essential differences between evidence-based nursing and evidence-based medicine. *Int J Nurs Terminol Classif, 13*(3), pp. 101-106. doi: 10.1111/j.1744-618x.2002.tb00409.x

Lockwood, C., Munn, Z., & Porritt, K. (2015). Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *13*(3), 179-187. Retrieved (Accessed.

Lühnen, J., Berger-Höger, B., & Richter, T. (2021). Systematic Reviews zur Förderung einer evidenzbasierten Praxis: Welche Standards braucht die Hochschullehre? Diskussion zur Entwicklung eines Leitfadens. *die hochschullehre. Interdisziplinäre Zeitschrift für Studium und Lehre, 7*(6), pp. 49-57.

Ma, L. L. et al. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res, 7*(1). doi: 10.1186/s40779-020-00238-8

MacDermid, J. C. et al. (2005). Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. *BMC Health Serv Res, 5*(1), pp. 18. doi: 10.1186/1472-6963-5-18

Mayer, H., Raphaelis, S., & Kobleder, A. (2021). *Literaturreviews für Gesundheitsberufe. Recherchieren - Bewerten - Erstellen* (1. ed.). Wien: Facultas.

Mayring, P. (2017). Evidenztriangulation in der Gesundheitsforschung Kombination von experimentellen, deskriptiven und inhaltsanalytischen Ansätzen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 69*(2), pp. 415–434.

McArthur, A. et al. (2015). Innovations in the systematic review of text and opinion. *International Journal of Evidence-Based Healthcare, 13*(3), pp. 188-195. doi: 10.1097/XEB.0000000000000060

Mehrholz, J. (2010). Wissenschaft erklärt: Qualitative Studiendesigns – Weiche Daten erfassen. *Ergopraxis, 8*(02), pp. 13-13.

Moher, D. et al. (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Bmj, 340*, pp. c869. doi: 10.1136/bmj.c869

Moher, D. et al. (2011). Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. *DMW - Deutsche Medizinische Wochenschrift, 136*(8), pp. 9-15. doi: 10.1055/s-0031-1272978

Moola, S. et al. (2020). Chapter 7: Systematic reviews of etiology and risk. In E. Aromataris & Z. Munn (Eds.), *JBI Manual for Evidence Synthesis*: JBI. Retrieved from https://synthesismanual.jbi.global (Accessed: 18.07.2022).

Moorley, C. & Cathala, X. (2019). How to appraise qualitative research. *Evidence Based Nursing, 22*(1), pp. 10-13. doi: 10.1136/ebnurs-2018-103044

Morse, J. M. et al. (1996). Concept analysis in nursing research: a critical appraisal. *Sch Inq Nurs Pract, 10*(3), pp. 253-277.

Munn, Z. et al. (2020). Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. *JBI Evidence Synthesis, 18*(10), pp. 2127-2133. doi: https://doi.org/10.11124/jbisrir-d-19-00099

Munthe-Kaas, H. M. et al. (2019). Systematic mapping of existing tools to appraise methodological strengths and limitations of qualitative research: first stage in the development of the CAMELOT tool. *BMC Medical Research Methodology, 19*(1), pp. 113. doi: 10.1186/s12874-019-0728-6

Niederberger, M. & Dreiack, S. (2020). Die qualitative Inhaltsanalyse in den Gesundheitswissenschaften. Ergebnisse eines systematischen Reviews einschlägiger

Fachzeitschriften. *Forum Qualitative Sozialforschung / Forum: Qualitative Social, Research and Practice in Alzheimer's disease, 21*(1).

Niederberger, M. & Peter, L. (2018). Mixed-Methods-Studien in den Gesundheitswissenschaften. Ein kritischer Überblick. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 133*, pp. 9-23. doi: https://doi.org/10.1016/j.zefq.2018.02.008

Noble, H. & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence Based Nursing, 18*(2), pp. 34. doi: 10.1136/eb-2015-102054

Nordhausen, T. & Hirt, J. (2020). RefHunter. Manual zur Literaturrecherche in Fachdatenbanken. Version 5.0 vom 13. Oktober 2020. Retrieved from https://refhunter.eu/manual/ (Accessed: 25.02.2022).

Oxford Centre for Evidence-Based Medicine. (2009). Oxford Centre for Evidence-Based Medicine: Levels of Evidence. Retrieved from https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009 (Accessed: 31.01.2022).

Paez, A. (2017). Gray literature: An important resource in systematic reviews. *J Evid Based Med, 10*(3), pp. 233-240. doi: https://doi.org/10.1111/jebm.12266

Page, M. J., McKenzie, J. E., & Higgins, J. P. T. (2018). Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open, 8*(3), pp. e019703. doi: 10.1136/bmjopen-2017-019703

Panfil, E.-M., Ivanović, N., & Conrad, C. (2011). Methodenpapier. FIT-Nursing Care. Forschungs- und IT-gestützte Expertennetzwerk-Plattform zur Unterstützung von Evidence-based Nursing. Version 1.0 Stand Juni 2011., 1-66. Retrieved from https://www.fit-care.ch/documents/8165518/0/110615_Methodenpapier_FIT%2BNC_Juni.pdf/a6d68f1c-55c9-4d09-982f-b40a374c0b2c (Accessed: 27.01.2022).

Panke-Kochinke, B. (2012). Augenscheinlich fehlgeleitet. Evidenz und Empirie. Methodische Postulate für eine qualitative Versorgungsforschung. *Pflege & Gesellschaft, 17*(1), pp. 5-21.

Petticrew, M. & Roberts, H. (2003). Evidence, hierarchies, and typologies: horses for courses. *J Epidemiol Community Health, 57*(7), pp. 527-529. doi: 10.1136/jech.57.7.527

Pieper, D. et al. (2015). Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. *J Clin Epidemiol, 68*(5), pp. 574-583. doi: 10.1016/j.jclinepi.2014.12.009

Platz , T. (2018). Leitlinien für Diagnostik und Therapie in der Neurologie. Therapie des spastischen Syndroms. Entwicklungsstufe S2k. Retrieved from https://www.dgn.org/images/red_leitlinien/LL_2019/PDFs_Download/030078_LL_Spastisches_Syndrom_2019_aktualisiert.pdf (Accessed: 11.05.2020).

Platz, T. (2021). Evidenzbasierte Leitlinienentwicklung der Deutschen Gesellschaft für Neurologie (DGN) und der Deutschen Gesellschaft für Neurorehabilitation (DGNR) - Methodik für die systematische Evidenzbasierung. *Fortschr Neurol Psychiatr, 89*(09), pp. 415-423.

Ploeg, J. (1999). Identifying the best research design to fit the question. Part 2: qualitative designs. *Evidence Based Nursing, 2*(2), pp. 36-37. doi: 10.1136/ebn.2.2.36

Pluye, P. et al. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. *International Journal of Nursing Studies, 46*(4), pp. 529-546. doi: https://doi.org/10.1016/j.ijnurstu.2009.01.009

Porritt, K., Gomersall, J., & Lockwood, C. (2014). JBI's Systematic Reviews: Study selection and critical appraisal. *Am J Nurs, 114*(6), pp. 47-52. doi: 10.1097/01.naj.0000450430.97383.64

Quasdorf, T. & Holle, D. (2018). Mixed-Methods-Forschungsdesigns. In H. Brandenburg et al. (Eds.), *Pflegewissenschaft 2: Lehr- und Arbeitsbuch zur Einführung in die Methoden der Pflegeforschung*. Bern: Hogrefe, pp. 116-123.

Roberts, J. & DiCenso, A. (1999). Identifying the best research design to fit the question. Part 1: quantitative designs. *Evidence-Based Nursing, 2*(1), pp. 4-6.

Rojahn, J. (2016). Studien richtig lesen – Teil 1: Studiendesign und Fehlerquellen. *Lege artis - Das Magazin zur ärztlichen Weiterbildung, 6*(01), pp. 16-23.

Schmucker, C. et al. (2017). Bewertung des Verzerrungsrisikos von systematischen Übersichtsarbeiten: ein Manual für die Leitlinienerstellung. *1.0*. Retrieved from https://www.cochrane.de/sites/cochrane.de/files/uploads/manual_bewertung_sr_fuer_ll.pdf (Accessed: 05.01.2022).

Schmucker, C. et al. (2016). Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien: ein Manual für die Leitlinienerstellung. *Version 1.0 vom Mai 2016*. Retrieved from http://www.cochrane.de/de/news/manual-bewertung-des-biasrisikos-risiko-systematischer-fehler-klinischen-studien-ein-manual-f%C3%BCr (Accessed: 07.03.2022).

Schünemann, H. et al. (2013). GRADE Handbook - Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. *Updated October 2013*. Retrieved from https://gdt.gradepro.org/app/handbook/handbook.html (Accessed: 12.01.2022).

Shea, B. J. et al. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ, 358*(j4008).

Simon, M. (2018). Literaturrecherche. In H. Brandenburg, E.-M. Panfil & H. Mayer (Eds.), *Pflegewissenschaft 2. Lehr- und Arbeitsbuch zur Einführung in die Methoden der Pflegeforschung* (3. ed.). Bern: Hogrefe, pp. 47-71.

Stickley, T. & Phillips, C. (2008). Einzelfallstudie und evidenzbasierte Praxis im Vergleich. *PPH, 14*(01), pp. 20-23.

Stratil, J. M. et al. (2021). Non-pharmacological measures implemented in the setting of long-term care facilities to prevent SARS-CoV-2 infections and their consequences: a rapid review. *Cochrane Database Syst Rev, 9*(9), pp. Cd015085. doi: 10.1002/14651858.CD015085.pub2

Taylor, E. & Hignett, S. (2014). Evaluating Evidence: Defining Levels and Quality Using Critical Appraisal Mixed Methods Tools. *HERD: Health Environments Research & Design Journal, 7*(3), pp. 144-151. doi: 10.1177/193758671400700310

Thangatorai, R., Lim, F. C., & Nalliah, S. (2018). Cervical pessary in the prevention of preterm births in multiple pregnancies with a short cervix: PRISMA compliant systematic review and meta-analysis. *J Matern Fetal Neonatal Med, 31*(12), pp. 1638-1645. doi: 10.1080/14767058.2017.1319930

The AGREE Collaboration (Ed.). (2001). Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument – Deutschsprachige Version. Retrieved from https://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_Instrument_German.pdf (Accessed: 08.03.2022).

The AGREE Next Steps Consortium (Ed.). (2013). Appraisal of Guidelines for Research & Evaluation II. AGREE II Instrument. Deutsche Version 2014. Retrieved from https://www.agreetrust.org/wp-content/uploads/2014/03/AGREE_II_German-Version.pdf (Accessed: 08.03.2022).

Thole, H. et al. (2007). Kritische Bewertung von Leitlinien. In R. Kunz et al. (Eds.), *Lehrbuch Evidenz-basierte Medizin in Klinik und Praxis* (2. ed.). Köln: Deutscher Ärzte-Verlag, pp. 177-187.

Tyndall, J. (2008). How long can you go? Toward a hierarchy of grey literature. Retrieved from https://dspace.flinders.edu.au/xmlui/bitstream/handle/2328/3326/Tyndall.pdf?sequence=1&isAllowed=y (Accessed: 24.02.2022).

Tyndall, J. (2010). AAOCDS Checklist. Retrieved from https://dspace.flinders.edu.au/xmlui/bitstream/handle/2328/3326/AACODS_Checklist.pdf?sequence=4&isAllowed=y (Accessed: 15.02.2022).

Vandenbroucke, J. P. et al. (2014). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg, 12*(12), pp. 1500-1524. doi: 10.1016/j.ijsu.2014.07.014

Vaska, M. et al. (2019). Exploring all that is Grey in the Health Sciences: What is Grey Literature and How to use it for Comprehensive Knowledge Synthesis. *Journal of National Heart Foundation of Bangladesh, 8*(1), pp. 14-19.

von Elm, E. et al. (2014). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg, 12*(12), pp. 1495-1499. doi: 10.1016/j.ijsu.2014.07.013

Werner, R. N. et al. (2016). The quality of European dermatological guidelines: critical appraisal of the quality of EDF guidelines using the AGREE II instrument. *J Eur Acad Dermatol Venereol, 30*(3), pp. 395-403. doi: 10.1111/jdv.13358

Williams, V., Boylan, A. M., & Nunan, D. (2019). Qualitative research as evidence: expanding the paradigm for evidence-based healthcare. *BMJ Evid Based Med, 24*(5), pp. 168-169. doi: 10.1136/bmjebm-2018-111131

Williams, V., Boylan, A. M., & Nunan, D. (2020). Critical appraisal of qualitative research: necessity, partialities and the issue of bias. *BMJ Evid Based Med, 25*(1), pp. 9-11. doi: 10.1136/bmjebm-2018-111132

Woods, S., Phillips, K., & Dudash, A. (2020). Grey literature citations in top nursing journals: a bibliometric study. *Journal of the Medical Library Association, 108*(2), pp. 262-269. doi: 10.5195/jmla.2020.760

Zeng, X. et al. (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med, 8*(1), pp. 2-10. doi: 10.1111/jebm.12141

Ziegler, A. & König, I. R. (2011). Leitlinien für Forschungsberichte: Deutschsprachige Übersetzungen von CONSORT 2010, PRISMA und STARD. *DMW - Deutsche Medizinische Wochenschrift, 136*(08), pp. 357-358. doi: 10.1055/s-0031-1272535

Anlage I: Exemplarisch befüllte Evidenztabelle im Rahmen der ENP-Weiterentwicklung seit 2021

Beispiel für eine noch unvollständig ausgefüllte Evidenztabelle im Rahmen der ENP-Weiterentwicklung:

KOMMENTARE ZU DEN EINZELNEN SPALTEN BEACHTEN (Hinweise, Bedeutung, Umfang, etc.)!!!

Ref.Nr.	Autor(en)	Jahr	Studientyp	Population, TN-Zahlen, ggf. Studienanzahl	Evidenzbewertung (Rater 1: Wieteck)	Evidenzbewertung (Rater 2: Kraus)	Fragestellung, Methodik, Outcomemessung und Hauptergebnisse	Schlussfolgerung und ENP-Relevanz	ENP-Item-Typ	Quelle (formatiert aus EndNote)
1	Miranzadeh, S	2014	RCT		●●●○	●●●○	Scharfqarben-Destilat 50/50 in Routinemundwasser reduziert Mukositis/Stomatitis bei Chemotherapie; 4 x täql. mit 15 ml spülen		Intervention	Miranzadeh, S. et al. (2014). A New mouthwash for Chemothera
2	Ferreira, A. S.	2022	Review		●●○○	●●●○	Vor allem Naturprodukte		Intervention	Ferreira, A. S. et al. (2022). Natural Products for the Prevention a
3	Charalambou	2018	RCT		●●●○	●●●○	Thyminahoniq 20 ml & 100 ml qereiniqtes Wasser; statistische Siqnifikanz		Intervention	Charalambous, M. et al. (2018). The effect of the use of thyme h
4	Khanjani Pour	2019	singel-blind clinical trial		●●○○	●●○○	Die Prävalenz verschiedener Mukositisgrade in den Studiengruppen war am Ende der dritten (p = 0,002) und vierten (p < 0,001) Woche signifikant. Der		Intervention_Ho	Khanjani Pour-Fard-Pachekenari, A. et al. (2019). The effect of a
5	Bogner, S. (20	2022	monozentrische, retrospektive Analys		●●○○	●●○○	Inzidenz OM nach Stammzellentransplantation		Inzidenz	Bogner, S. (2022). Orale Mukositis: Inzidenz und Schweregrad b
6	Daugėlaitė, G.	2019	systematic Review		●●●○	●●○○	Ergebnisse: Die Ergebnisse der in dieser systematischen Literaturrecherche verwendeten Studien zeigten, dass Lasertherapie, Kryotherapie, professionelle		Interventionen	Daugėlaitė, G. et al. (2019). Prevention and Treatment of Chemo
7	Liao, Y. C. et a	2021	RCT		●●●○	●●●○	Grüner Tee hat positive Auswirkungen		Mundspüllösun	Liao, Y. C. et al. (2021). Effectiveness of green tea mouthwash fo
8	Bornemann-C	2013	Systemtic Review		●●○○		Lokale Schmerztherapie		Mundspüllösun	Bornemann-Cimenti, H. et al. (2013). [Topical pain therapy in o
9	Rosemann, E.	2022	retrospektive Datenauswertung		●●○○		Inzidenz OM bei 5-Fluorouracil und möglichen Einflussfaktoren wie Zahnstatus; Mundhygiene..; Dissertation		Inzidenz, Risikof	Rosemann, E. (2022). Orale Mucositis bei gastrointestinalen Tu
10	Elad, S. et al.	2020	Leitlinie Aktualisierung		●●●○		Evidenz für Kryotherapie; LLLT; Mundpflege; Mundspülungen; Honig, Patientenedukation, Anlagesei soie natürliche Wirkstoffe		Interventionen	Elad, S. et al. (2020). MASCC/ISOO clinical practice guidelines fo
11	Colella, G. et a	2023	Review		●●○○		Entzündungshemmende Medikamente (einschließlich Benzydamin), Wachstumsfaktoren und Zytokine (einschließlich Palifermin), Kryotherapie, Laser-		Interventionen	Colella, G. et al. (2023). Interventions for the Prevention of Oral
12	Jicman Stan, D	2022	systematic Review		●●○○		Ergebnisse : Zahlreiche Studien haben die Vorteile von Honig bei oraler Mukositis hervorgehoben. Durch seine schmerzstillende, entzündungshemmend		Intervention Bie	Jicman Stan, D. et al. (2022). Oral Mucositis Induced by Chemor
13	Hinkelmann,	2019	systematic Review		●●○○		Im Rahmen einer Bachleorarbeit		Interventionen c	Hinkelmann, A. (2019). Die Prävention der tumortherapieinduzi
14	Hosseinjani, H	2017	RCT		●●●○		EPO-Mundwasserqruppe und der Kontrollqruppe war siqnifikant unterschiedlich (27,5 % vs. 77,5 %, p < 0,001). Der Mittelwert ± SD von zwei anderen OM		Erythropoetin-M	Hosseinjani, H. et al. (2017). The efficacy of erythropoietin mou
15	Hasheminasal	2020	double-blind, randomized, controlled		●●○○		Im Vergleich zum Placebo reduzierte die pflanzliche Verbindung den Grad der Mukositis, die Schwere der Schmerzen und den Grad der Xerostomie deutl		Plantago ovata-l	Hasheminasab, F. S. et al. (2020). Effects of a Plantago ovata-ba
16	Pulito, C. et al	2020	Review		●●○○		Überblick über die präventieven Maßnahmen, bestätigt bisheriges		Interventionen	Pulito, C. et al. (2020). Oral mucositis: the hidden side of cancer
17	Fidan, Ö. & Ar	2021	Studie mit Regressionsanalyse zu Risil		●●○○		Population: risiko einer OM bei hematologischen Patienten; 187 Patienten waren in der Stichprobe einbezogen; Folgende Risikofaktoren wurden Untersu		Risiko-Assessme	Fidan, Ö. & Arslan, S. (2021). Development and Validation of the
18	Raza, A. et al.	2022	systemati Review &Metaanalyse		●●●○		Datenbanken: Cochrane-, EMBASE-, PubMed- und Web of Science-Datenbanken wurden verwendet, um nach randomisierten kontrollierten Studien (RCT		orale Antioxidar	Raza, A. et al. (2022). Efficacy of oral and topical antioxidants in
19	de Menêses, A	2020	systematic review and meta-analysis.		●●●○		6 Datenbanken; Population: oM bei Krebspatienten mit Chemotherapie oder Strahlentherapie; Risk of Bias Tool der Cochrane Collaboration, Evidenzbew		orale Aufnahme	de Menêses, A. G. et al. (2020). Effects of oral supplementation i
20	Rupe, C. et al.	2022	Beobachtungsstudie		●●○○		Ziel dieser Studie ist es, einen Zusammenhang zwischen Candida spp. orale Kolonisierung vor der Strahlentherapie (RT) und (i) die Entwicklung einer sch		Zusammenhang	Rupe, C. et al. (2022). Oral Candida spp. Colonisation Is a Risk Fa
21	Leppla, L. et al	2016	Pilot-RCT		●●○○		Oral care self-management support protocol (OrCaSS); OrCaSS ist eine vielversprechende Intervention zur Verzögerung und Reduzierung von OM.		Assessment	Leppla, L. et al. (2016). An oral care self-management support p
22	Wong, S. P. et	2023	Psychometrische evaluation		●●○○		Oral Mucositis Daily Questionnaire; Interne Konsistenz und Reproduzierbarkeit wurden durch den Cronbach-Alpha- bzw. Intraklassen-Korrelationskoeffizi		Assessment	Wong, S. P. et al. (2023). Psychometric evaluation of Oral Muco

Anlage II: Checklisten zum Critical Appraisal verschiedener Publikations- und Studientypen seit 2021

Critical Appraisal von systematischen Übersichtsarbeiten gemäß AMSTAR 2:

Thieme RECOM

Critical Appraisal von systematischen Übersichtsarbeiten und Meta-Analysen gemäß AMSTAR-2[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von systematischen Übersichtsarbeiten und Meta-Analysen gemäß AMSTAR-2	Ja	Teil-weise	nein	Nicht anwendbar
1. Enthalten die Forschungsfragen und Einschlusskriterien des SR / der Meta-Analyse die Komponenten des PICO-Schemas?	☐		☐	
2. Wird ausdrücklich beschrieben, dass die Methodik und Techniken zur Durchführung des SR / der Meta-Analyse im Vorfeld (a priori) festgelegt wurden und werden alle Abweichungen begründet?	☐	☐	☐	
3. Wurde die Auswahl der Studiendesigns für die Aufnahme in die Übersichtsarbeit erklärt?	☐		☐	
4. Wurde eine umfassende und systematische Strategie zur Literaturrecherche entwickelt und angewendet?	☐	☐	☐	
5. Wurde die Studienauswahl von mindestens zwei der Autoren unabhängig voneinander ausgeführt?	☐		☐	
6. Wurde die Datenextraktion von mindestens zwei der Autoren unabhängig voneinander ausgeführt?	☐		☐	
7. Wurde eine Liste der ausgeschlossenen Studien bereitgestellt und die Ausschlüsse begründet?	☐	☐	☐	
8. Wurden die eingeschlossenen Studien ausreichend detailliert beschrieben?	☐	☐	☐	
9. Wurde eine angemessene Technik zur Bewertung des systematischen Verzerrungsrisikos (Risk of Bias) in den einzelnen Studien, die in das SR / die Meta-Analyse aufgenommen wurden, verwendet?	☐	☐	☐	
10. Wurde über die Finanzierungsquellen der in das SR / die Meta-Analyse einbezogenen Studien berichtet?	☐		☐	
11. Im Falle einer Meta-Analyse: Wurden geeignete Methoden für die statistische Zusammenführung der Ergebnisse verwendet?	☐		☐	☐
12. Im Falle einer Meta-Analyse: Wurden mögliche Auswirkungen eines systematischen Verzerrungsrisikos (Risk of Bias) der einzelnen Studien auf die Ergebnisse der Meta-Analyse oder einer anderen Evidenzsynthese bewertet?	☐		☐	☐
13. Wurde bei der Interpretation/Diskussion der Ergebnisse des SR / der Meta-Analyse das Risiko für eine systematische Verzerrung (Risk of Bias) in den einzelnen Studien berücksichtigt?	☐		☐	
14. Wurde für eine ggf. in den Ergebnissen beobachtete Heterogenität eine angemessene Erklärung angegeben sowie eine Diskussion geführt?	☐		☐	
15. Falls eine quantitative Synthese durchgeführt wurde, wurde eine angemessene Analyse des Publikationsbias (Verzerrung durch kleine Studien) durchgeführt und mögliche Auswirkungen auf die Ergebnisse des SR / der Meta-Analyse diskutiert?	☐		☐	☐
16. Wurde über mögliche Interessenkonflikte der Autor*innen berichtet, einschließlich einer eventuellen Finanzierung, die sie für die Durchführung des SR / der Meta-Analyse erhalten haben?	☐		☐	

[1] **Shea, B. J. et al.** (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ: British Medical Journal, 358*(j4008). doi: https://doi.org/10.1136/bmj.j4008

Erklärungen und Interpretationshilfen zu den Bewertungskategorie von AMSTAR-2
1. Enthalten die Forschungsfragen und Einschlusskriterien des SR / der Meta-Analyse die Komponenten des PICO-Schemas? • Für Übersichtsarbeiten im Kontext von Pflegediagnosen kann hier auch ein „ja“ vergeben werden, wenn die in Kapitel 2.1.1 beschriebenen Elemente enthalten sind. • Für ein „ja“ sind die entsprechenden Elemente (Population, Intervention, Kontrollgruppe, Outcome ODER Pflegediagnose, beobachtbare/messbare Kennzeichen, Ergebnis) vorhanden. „Nein“ trifft zu, wenn Elemente fehlen.
2. Wird ausdrücklich beschrieben, dass die Methodik und Techniken zur Durchführung des SR / der Meta-Analyse im Vorfeld (a priori) festgelegt wurden und werden alle Abweichungen begründet? • Es ist erforderlich, dass sich die Autoren der Übersichtsarbeit auf ein Protokoll oder auf vorab definiere/a priori veröffentlichte Forschungsziele und ein im Vorfeld erdachtes methodisches Vorgehen beziehen. • Folgende Elemente müssen für ein „teilweise erfüllt“ angegeben sein: Forschungsfrage, Suchstrategie, Ein-Ausschlusskriterien sowie eine Bewertung des Risikos für eine systematische Verzerrung (Bias). • Für ein „ja“ sind darüber hinaus Angaben über Gründe für eine mögliche Heterogenität sowie Gründe für ein Abweichen vom geplanten Vorgehen erforderlich, bei Metaanalysen zudem Angaben über die genutzten Syntheseverfahren.
3. Wurde die Auswahl der Studiendesigns für die Aufnahme in die Übersichtsarbeit erklärt? Für ein „ja“ ist eine der drei folgenden Bedingungen erforderlich: • Eine Erklärung, warum ausschließlich randomisiert kontrollierte Studien (RCTs) einbezogen wurden • Oder eine Erklärung, warum ausschließlich nicht-randomisierte Interventionsstudien (NRSI) einbezogen wurden • Oder eine Erklärung, warum sowohl RCTs als auch NRSI einbezogen wurden
4. Wurde eine umfassende und systematische Strategie zur Literaturrecherche entwickelt und angewendet? Für ein „teilweise erfüllt“ müssen folgende Kriterien erfüllt sein: • Die Durchführung der Literaturrecherche erfolgte in mindestens 2 für die Fragestellung relevanten Fachdatenbanken • Die verwendeten Suchbegriffe und/oder Suchphrasen werden angegeben • Die Ein- und Ausschlusskriterien (z.B. Publikationssprache) mit Begründung Für ein „ja“ müssen darüber hinaus Angaben zu folgenden Kriterien gemacht werden: • Wurden die Literaturverzeichnisse der eingeschlossenen Studien durchsucht? • Wurden spezifische Studienregister durchsucht? • Wurden Fachexperten zum jeweiligen klinischen Thema involviert oder konsultiert? • Wurde, falls relevant, nach grauer Literatur gesucht? • Wurde die Literaturrecherche innerhalb von 24 Monaten durchgeführt?
5. Wurden die Studienauswahl von mindestens zwei der Autoren unabhängig voneinander ausgeführt? Zur Beurteilung mit „ja“ ist eines der beiden Kriterien zu erfüllen: • Haben sich mindestens zwei Gutachter unabhängig voneinander auf die Auswahl der in Frage kommenden Studien geeinigt und erzielten einen Konsens darüber, welche Studien einbezogen werden sollten • Oder haben zwei Gutachter eine Stichprobe geeigneter Studien ausgewählt und haben dabei eine gute Übereinstimmungsrate von mindestens 80% erzielt, während die restlichen Studien von einem Gutachter ausgewählt wurden?
6. Wurde die Datenextraktion von mindestens zwei der Autoren unabhängig voneinander ausgeführt? Zur Beurteilung mit „ja“ ist eines der beiden Kriterien zu erfüllen: • Haben mindestens zwei Gutachter einen Konsens darüber erzielt, welche Daten aus den eingeschlossenen Studien extrahiert werden sollten? • Oder haben zwei Gutachter Daten aus einer Stichprobe in Frage kommender Studien extrahiert und erzielten dabei eine gute Übereinstimmung von mindestens 80%, während die Daten aus den restlichen Studien von einem Gutachter extrahiert wurden?

7. Wurde eine Liste der ausgeschlossenen Studien bereitgestellt und die Ausschlüsse begründet? Dieses Kriterium ist „teilweise erfüllt“, wenn eine Liste aller potenziell relevanten Studien zur Verfügung gestellt wird, die im Volltext gelesen, aber von der Überprüfung ausgeschlossen wurden. Für eine Beurteilung mit „ja“ müssen darüber hinaus individuelle Begründungen für den jeweiligen Ausschluss angegeben sein.
8. Wurden die eingeschlossenen Studien ausreichend detailliert beschrieben? Dieses Kriterium ist „teilweise erfüllt“, wenn folgende Aspekte der eingeschlossenen Studien beschrieben wurden: • Population, Intervention(en), Kontrollen/Kontrollgruppen, Outcome(s), Forschungsdesign Eine Bewertung mit „ja“ erfordert darüber hinaus, dass folgende Aspekte der eingeschlossenen Studien beschrieben wurden: • Detailbeschreibung der Population, Detailbeschreibung der Intervention(en) ggf. mit Dosierung, Detailbeschreibung der Kontrolle/Kontrollgruppen ggf. mit Dosierung, Studiensetting, Zeitrahmen für Nachbeobachtung(en)/Follow-Up Die Bewertungsaspekte sind bei eingeschlossenen nicht-randomisierten Studien und/oder pflegediagnostischen Studien entsprechend anzupassen!
9. Wurde eine angemessene Technik zur Bewertung des systematischen Verzerrungsrisikos (Risk of Bias) in den einzelnen Studien, die in das SR / die Meta-Analyse aufgenommen wurden, verwendet? Für randomisiert kontrollierte Studien (RCTs): • Für RCTs ist dieses Kriterium ist „teilweise erfüllt“, wenn das Risiko einer Verzerrung durch eine nicht-verdeckte Zuteilung und eine fehlende Verblindung von Patienten und Untersuchung bei der Bewertung der Ergebnisse geprüft wurde • Ein „ja“ ist dann gegeben, wenn darüber hinaus auch das Risiko einer Verzerrung durch eine nicht-zufällige Zuteilung der Probanden in Kontroll- bzw. Interventionsgruppe sowie durch eine (Teil)Auswahl der publizierten Ergebnisse aus mehreren Messungen oder einer Analyse von ausgewählten (Teil)Ergebnisses geprüft wurde Für nicht-randomisierte Interventionsstudien (NRSI): • Für NRSI ist dieses Kriterium ist „teilweise erfüllt“, wenn das Risiko einer Verzerrung durch Confounder (Störgrößen) sowie eines Selektionsbias (ungleiche Zusammensetzung der Vergleichsgruppen) geprüft wurde • Ein „ja“ ist dann gegeben, wenn darüber hinaus auch das Risiko einer Verzerrung durch die verwendeten Methoden zur Ermittlung von Exposition und Ergebnissen sowie durch eine (Teil)Auswahl der publizierten Ergebnisse aus mehreren Messungen oder einer Analyse von ausgewählten (Teil)Ergebnisses geprüft wurde
10. Wurde über die Finanzierungsquellen der in das SR / die Meta-Analyse einbezogenen Studien berichtet? Zur Beantwortung mit „ja“ Es müssen Angaben zu den Finanzierungsquellen der einzelnen in die Überprüfung einbezogenen Studien gemacht werden. Hinweis: Die Angabe, dass die Autor*innen nach diesen Informationen gesucht haben, diese aber von den Verfassern der einzelnen Studien nicht angegeben wurden, gilt ebenfalls als Nachweis.

11. Im Falle einer Meta-Analyse: Wurden geeignete Methoden für die statistische Zusammenführung der Ergebnisse verwendet? Für randomisiert kontrollierte Studien (RCTs) ist dies mit „ja" zu beantworten, wenn alle drei Bedingungen erfüllt sind: • Die Kombination der Daten in einer Meta-Analyse wird begründet • Und es wurde eine geeignete, gewichtete Technik zur Kombination der Studienergebnisse verwendet und eine ggf. vorhandene Heterogenität bereinigt • Und die Ursachen für eine eventuelle Heterogenität untersucht wurden Für nicht-randomisierte Interventionsstudien (NRSI) ist dies mit „ja" zu beantworten, wenn alle vier Bedingungen erfüllt sind: • Die Kombination der Daten in einer Meta-Analyse wird begründet • Und es wurde eine geeignete, gewichtete Technik zur Kombination der Studienergebnisse verwendet und eine ggf. vorhandene Heterogenität bereinigt • Und es wurden statistisch bereinigte Effektschätzungen anstelle von Rohdaten aus NRSI zusammengeführt (oder die Kombination von Rohdaten wurde begründet, wenn bereinigte Effektschätzungen nicht verfügbar waren) • Und es wurden separate Effektschätzungen für RCTs und NRSI angegeben, insofern beide Studientypen eingezogen wurden
12. Im Falle einer Meta-Analyse: Wurden mögliche Auswirkungen eines systematischen Verzerrungsrisikos (Risk of Bias) der einzelnen Stunden auf die Ergebnisse der Meta-Analyse oder einer anderen Evidenzsynthese bewertet? Für ein „ja" ist eine der beiden folgenden Bedingungen erforderlich: • Es wurden nur randomisiert kontrollierte Studien (RCTs) mit einem niedrigen systematischen Verzerrungsrisiko eingeschlossen • Oder es wurden Analysen der möglichen Auswirkungen eines systematischen Verzerrungseffekts auf die zusammenfassenden Effektschätzungen durchgeführt, wenn diese auf RCTs bzw. NRSI mit unterschiedler Gefahr für eine systematische Verzerrung basieren
13. Wurde bei der Interpretation/Diskussion der Ergebnisse des SR / der Meta-Analyse das Risiko für eine systematische Verzerrung (Risk of Bias) in den einzelnen Studien berücksichtigt? Für ein „ja" ist eine der beiden folgenden Bedingungen erforderlich: • Es wurden nur randomisiert kontrollierte Studien (RCTs) mit einem niedrigen systematischen Verzerrungsrisiko eingeschlossen • Oder, wenn RCTS mit mäßigen oder hohem Verzerrungsrisiko und/oder NRSI eingeschlossen waren, wurden die mutmaßlichen Auswirkungen eines systematischen Verzerrungseffekts auf die Ergebnisse diskutiert?
14. Wurde für eine ggf. in den Ergebnissen der beobachteten Heterogenität eine angemessene Erklärung angegeben sowie eine Diskussion geführt? Für ein „ja" ist eine der beiden folgenden Bedingungen erforderlich: • Es gab keine signifikante Heterogenität in den Ergebnissen • Oder, wenn eine Heterogenität vorlag, wurden die Ursachen versucht zu ergründen und die Auswirkungen auf die Ergebnisse diskutiert?
15. Falls eine quantitative Synthese durchgeführt wurde, wurde eine angemessene Analyse des Publikationsbias (Verzerrung durch kleine Studien) durchgeführt und mögliche Auswirkungen auf die Ergebnisse des SR / der Meta-Analyse diskutiert? Zur Beantwortung mit „ja" müssen grafische oder statistische Tests auf ein mögliches Publikationsbias durchgeführt und die Wahrscheinlichkeit sowie das Ausmaß der möglichen Auswirkungen diskutiert worden sein.
16. Wurde über mögliche Interessenkonflikte der Autor*innen berichtet, einschließlich einer eventuellen Finanzierung, die sie für die Durchführung des SR / der Meta-Analyse erhalten haben? Zur Beantwortung mit „ja" ist eine der beiden folgenden Bedingungen erforderlich: • Es wurden keine Interessenkonflikte angegeben • Oder Finanzierungsquellen wurden dargelegt und eine Beschreibung des Umgangs mit potenziellen Interessenkonflikten ist erfolgt

Critical Appraisal von Beobachtungsstudien (Querschnittstudien, Fall-Kontroll-Studien, Kohortenstudien, quasi-experimentelle Studien)

Thieme RECOM

Critical Appraisal von analytischen Querschnittstudien gemäß JBI[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von analytischen Querschnittstudien (JBI)	Ja	Nein	unklar	Nicht anwendbar
Waren die Kriterien für die Aufnahme in die Stichprobe klar definiert?	☐	☐	☐	☐
Wurden die Studienteilnehmer und das Setting der Studie detailliert beschrieben?	☐	☐	☐	☐
Wurde die Exposition auf eine valide und reliable Weise gemessen?	☐	☐	☐	☐
Wurden objektive und standardisierte Kriterien für die Messung des Zustands / der Befunde / der Parameter verwendet?	☐	☐	☐	☐
Wurden störende Faktoren/Einflüsse identifiziert?	☐	☐	☐	☐
Wurden Strategien zum Umgang mit Störfaktoren angegeben?	☐	☐	☐	☐
Wurden die Outcomes auf eine valide und reliable Art und Weise gemessen?	☐	☐	☐	☐
Wurden geeignete statistische Analyseverfahren gewählt und angewendet?	☐	☐	☐	☐

[1] **Joanna Briggs Institute (Ed.).** (2017). The Joanna Briggs Institute Critical Appraisal tools for use in JBI Systematic Reviews. Checklist for Analytical Cross Sectional Studies. Retrieved from https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Analytical_Cross_Sectional_Studies2017_0.pdf (Accessed: 16.02.2022).

Thieme RECOM

Critical Appraisal von Fall-Kontroll-Studien gemäß JBI[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von Fall-Kontroll-Studien gemäß JBI	Ja	Nein	unklar	Nicht anwendbar
1. Waren die Gruppen vergleichbar, abgesehen vom Vorhandensein der Krankheit / des pflegerelevanten Phänomens bei der Fallgruppe oder dem Nichtvorhandensein der Krankheit / des pflegerelevanten Phänomens bei der Kontrollgruppe?	☐	☐	☐	☐
2. Wurden Fallgruppe und Kontrollgruppe in angemessener Weise verglichen?	☐	☐	☐	☐
3. Wurden für die Identifizierung von Personen der Fallgruppe und der Kontrollgruppen dieselben Kriterien verwendet?	☐	☐	☐	☐
4. Wurde die Messung der Exposition auf standardisierte, valide und reliable Weise gemessen?	☐	☐	☐	☐
5. Wurde die Messung der Exposition auf selbe Weise gemessen?	☐	☐	☐	☐
6. Wurden störende Faktoren/Einflüsse identifiziert?	☐	☐	☐	☐
7. Wurden Strategien zum Umgang mit Störfaktoren angegeben?	☐	☐	☐	☐
8. Wurden die Ergebnisse auf standardisierte, valide und reliable Weise für die Fallgruppe und die Kontrollgruppe bewertet?	☐	☐	☐	☐
9. War der interessierende Expositions-/Beobachtungsszeitraum ausreichend lange, um aussagekräftig zu sein?	☐	☐	☐	☐
10. Wurden geeignete statistische Analyseverfahren gewählt und angewendet?	☐	☐	☐	☐

[1] **Joanna Briggs Institute (Ed.)**. (2017). The Joanna Briggs Institute Critical Appraisal tools for use in JBI Systematic Reviews. Checklist for Case Control Studies. Retrieved from https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Case_Control_Studies2017_0.pdf (Accessed: 16.02.2022).

Critical Appraisal von Kohortenstudien gemäß JBI[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von Kohortenstudien gemäß JBI	Ja	Nein	unklar	Nicht anwendbar
1. Waren beide Gruppen ähnlich und wurden sie aus derselben Population rekrutiert?	☐	☐	☐	☐
2. Wurden die Exposition in ähnlicher Weise durchgeführt, um die Personen sowohl der exponierten als auch der nicht exponierten Gruppen zuzuordnen?	☐	☐	☐	☐
3. Wurde die Exposition auf eine valide und reliable Weise gemessen?	☐	☐	☐	☐
4. Wurden störende Faktoren/Einflüsse identifiziert?	☐	☐	☐	☐
5. Wurden Strategien zum Umgang mit Störfaktoren angegeben?	☐	☐	☐	☐
6. Waren die Gruppen/Teilnehmer zu Beginn der Studie (oder zum Zeitpunkt der Exposition) frei von den zu untersuchenden Outcomes?	☐	☐	☐	☐
7. Wurden die Outcomes auf eine valide und reliable Art und Weise gemessen?	☐	☐	☐	☐
8. Wurde die Nachbeobachtungszeit angegeben und war sie lang genug, um Ergebnisse zu erzielen?	☐	☐	☐	☐
9. Wurde das Follow-up vollständig durchgeführt, und wenn nicht, wurden die Gründe für den Verlust des Follow-up beschrieben und untersucht?	☐	☐	☐	☐
10. Wurden Strategien zum Umgang mit unvollständigen Nachuntersuchungen angewandt?	☐	☐	☐	☐
11. Wurden geeignete statistische Analyseverfahren gewählt und angewendet?	☐	☐	☐	☐

[1] **Joanna Briggs Institute (Ed.)**. (2017). The Joanna Briggs Institute Critical Appraisal tools for use in JBI Systematic Reviews. Checklist for Cohort Studies. Retrieved from https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Cohort_Studies2017_0.pdf (Accessed: 16.02.2022).

Thieme RECOM

Critical Appraisal von quasi-experimentellen Studien gemäß JBI[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von quasi-experimentellen Studien gemäß JBI	Ja	Nein	unklar	Nicht anwendbar
1. Wird in der Studie deutlich, was die "Ursache" und was die "Wirkung" ist (d. h. es gibt keine Verwirrung darüber, welche Variable zuerst kommt)?	☐	☐	☐	☐
2. Waren die Teilnehmer der Vergleichsgruppen hinsichtlich ihrer Merkmale (z.B. Alter, Schweregrad der Erkrankung etc.) ähnlich? 3. HINWEIS: Bei Prä-Test/Post-Test-Studien mit einer einzigen Gruppe, bei denen die Patienten in allen Prä-Post-Vergleichen dieselben sind (dieselbe Gruppe), sollte die Antwort auf diese Frage „ja" lauten	☐	☐	☐	☐
4. Erhielten die Teilnehmer der Vergleichsgruppen abgesehen von der interessierenden Exposition/Intervention eine ähnliche Behandlung/Betreuung?	☐	☐	☐	☐
5. Gab es eine Kontrollgruppe?	☐	☐	☐	☐
6. Wurden mehrere Messungen des Outcomes vor und nach der Intervention/Exposition durchgeführt?	☐	☐	☐	☐
7. Wurde die Nachbeobachtung (follow-up) vollständig durchgeführt, und falls nicht, wurden die Unterschiede zwischen den Vergleichsgruppen in Bezug auf ihre Nachbeobachtung angemessen beschrieben und analysiert?	☐	☐	☐	☐
8. Wurden die Outcomes aufseiten der Teilnehmer der Vergleichsgruppen auf dieselbe Weise gemessen?	☐	☐	☐	☐
9. Wurden die Outcomes auf eine valide und reliable Art und Weise gemessen?	☐	☐	☐	☐
10. Wurden geeignete statistische Analyseverfahren gewählt und angewendet?	☐	☐	☐	☐

[1] **Joanna Briggs Institute (Ed.)**. (2017). The Joanna Briggs Institute Critical Appraisal tools for use in JBI Systematic Reviews. Checklist for Quasi-Experimental Studies (non-randomized experimental studies). Retrieved from https://jbi.global/sites/default/files/2019-05/JBI_Quasi-Experimental_Appraisal_Tool2017_0.pdf (Accessed: 16.02.2022).

Critical Appraisal von randomisiert kontrollierten Studien (RCTs):

Thieme RECOM

Critical Appraisal von randomisiert kontrollierten Studien (RCTs) gemäß Platz (2021)[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von randomisiert kontrollierten Studien (RCTs) gemäß Platz (2021)	Ja	Nein	unklar	Nicht anwendbar
1. Wurden die Ein- und Ausschlusskriterien klar definiert? (externe Validität)	☐	☐	☐	☐
2. Gibt es eine klare Definition und ein adäquates Assessment der Zielgrößen (Outcomes) der Studie (klinische Relevanz)?	☐	☐	☐	☐
3. Wurden Nebenwirkungen und Akzeptabilität der Interventionen berichtet?	☐	☐	☐	☐
4. Fand eine adäquate „Follow-up"-Untersuchung statt (Erfassung von Langzeiteffekten)?	☐	☐	☐	☐
5. Wird eine klare Definition und Beschreibung der experimentellen und der Kontrollbedingungen gegeben (Studienkontrast)?	☐	☐	☐	☐
6. Werden die Teilnehmer randomisiert zugeteilt (Selektionsbias)?	☐	☐	☐	☐
7. Wurde die Zuordnung unter Verschluss gehalten („allocation concealment", Selektionsbias)?	☐	☐	☐	☐
8. Waren die Experimental- und Kontrollgruppen zu Beginn der Studie vergleichbar (Selektionsbias)?	☐	☐	☐	☐
9. Waren die Patienten sowie das Personal während der Behandlung verblindet, und waren außerhalb der untersuchten Effekte die Behandlungen in den randomisierten Gruppen vergleichbar (Durchführungsbias)?	☐	☐	☐	☐
10. Wurden die Studienergebnisse verblindet erhoben („detection bias")?	☐	☐	☐	☐
11. War die Berichterstellung nicht selektiv („reporting bias")?	☐	☐	☐	☐
12. Waren die Ergebnisdaten (fast) komplett („attrition bias")?	☐	☐	☐	☐
13. Wurde eine „Intention-to-treat"-Analyse berichtet („attrition bias")?	☐	☐	☐	☐
14. Unterstützen die Ergebnisse die berichteten Schlussfolgerungen ausreichend?	☐	☐	☐	☐

[1] **Platz, T.** (2021). Evidence-based Practice Guidelines for the German Society for Neurology (DGN) and the German Society for Neurorehabilitation (DGNR): Methods for systematic evidence-to-decision process (in German). *Fortschritte der Neurologie - Psychiatrie, 89*(09), pp. 415-423. doi: https://doi.org/10.1055/a-1309-1856

Critical Appraisal von qualitativen Forschungsarbeiten:

Thieme RECOM

Critical Appraisal von qualitativer Forschung gemäß JBI[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von qualitativer Forschung gemäß JBI	Ja	Nein	unklar	Nicht anwendbar
1. Besteht eine Kongruenz zwischen der angegebenen philosophischen Perspektive und der Forschungsmethodik?	☐	☐	☐	☐
2. Besteht Kongruenz zwischen der Forschungsmethodik und der Forschungsfrage oder den Forschungszielen?	☐	☐	☐	☐
3. Besteht eine Kongruenz zwischen der Forschungsmethodik und den zur Datenerhebung verwendeten Methoden?	☐	☐	☐	☐
4. Gibt es eine Übereinstimmung zwischen der Forschungsmethodik und der Darstellung und Analyse der Daten?	☐	☐	☐	☐
5. Besteht Kongruenz zwischen der Forschungsmethodik und der Interpretation der Ergebnisse?	☐	☐	☐	☐
6. Gibt es eine Erklärung, die den Forscher kulturell oder theoretisch verortet?	☐	☐	☐	☐
7. Wird der Einfluss des Forschers auf die Forschung und umgekehrt angesprochen?	☐	☐	☐	☐
8. Werden die Teilnehmenden und ihre Meinungen angemessen wiedergegeben?	☐	☐	☐	☐
9. Ist die Forschung nach den aktuell geltenden Kriterien ethisch vertretbar oder gibt es bei neueren Studien einen Nachweis über die ethische Genehmigung durch eine anerkannte Einrichtung?	☐	☐	☐	☐
10. Sind die im Forschungsbericht gezogenen Schlussfolgerungen aus der Analyse oder Interpretation der Daten abgeleitet?	☐	☐	☐	☐

[1] **Joanna Briggs Institute (Ed.)**. (2017). The Joanna Briggs Institute Critical Appraisal tools for use in JBI Systematic Reviews. Checklist for Qualitative Research. Retrieved from https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Qualitative_Research2017_0.pdf (Accessed: 16.02.2022).

Critical Appraisal von Leitlinien und Clinical Practice Guidelines – AGREE II:

Thieme RECOM

Critical Appraisal von Leitlinien und Clinical Practice Guidelines gemäß AGREE II[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von Leitlinien und Clinical Practice Guidelines gemäß AGREE II	1-3	4	5-7	Bemerkungen/ Detail-bewertung
Domäne 1: Geltungsbereich und Zweck				
1. Das / die Gesamtziel(e) der Leitlinie ist / sind eindeutig beschrieben.	☐	☐	☐	
2. Die in der Leitlinie behandelte(n) gesundheitsrelevante(n) Frage(n) ist (sind) eindeutig beschrieben.	☐	☐	☐	
3. Die Zielpopulation (z. B. Patienten, Bevölkerung) der Leitlinie ist eindeutig beschrieben.	☐	☐	☐	
Domäne 2: Beteiligung von Interessengruppen				
4. Die Entwicklergruppe der Leitlinie schließt Mitglieder aller relevanten Berufsgruppen ein.	☐	☐	☐	
5. Die Ansichten und Präferenzen der Zielpopulation (z. B. Patienten, Bevölkerung) wurden ermittelt.	☐	☐	☐	
6. Die Anwenderzielgruppe(n) der Leitlinie ist (sind) eindeutig beschrieben.	☐	☐	☐	
Domäne 3: Genauigkeit der Leitlinienentwicklung				
7. Es wurde systematisch nach Evidenz gesucht.	☐	☐	☐	
8. Die Kriterien für die Auswahl der Evidenz sind eindeutig beschrieben.	☐	☐	☐	
9. Die Stärken und die Schwächen der Evidenz sind eindeutig beschrieben.	☐	☐	☐	
10. Das methodische Vorgehen bei der Formulierung der Empfehlungen ist eindeutig beschrieben.	☐	☐	☐	
11. Der gesundheitliche Nutzen, Nebenwirkungen und Risiken wurden bei der Formulierung der Empfehlungen berücksichtigt.	☐	☐	☐	
12. Die zugrunde liegende Evidenz kann den Empfehlungen eindeutig zugeordnet werden.	☐	☐	☐	
13. Die Leitlinie wurde vor ihrer Veröffentlichung durch externe Experten begutachtet.	☐	☐	☐	
14. Es existiert ein Verfahren zur Aktualisierung der Leitlinie.	☐	☐	☐	
Domäne 4: Klarheit der Gestaltung				
15. Die Empfehlungen der Leitlinie sind spezifisch und eindeutig.	☐	☐	☐	
16. Die unterschiedlichen Alternativen für die Behandlung der Erkrankung oder des Gesundheitsproblems sind eindeutig dargestellt.	☐	☐	☐	
17. Die Schlüsselempfehlungen der Leitlinie sind einfach zu finden.	☐	☐	☐	

[1] **The AGREE Next Steps Consortium (Ed.).** (2013). Appraisal of Guidelines for Research & Evaluation II. AGREE II Instrument. Deutsche Version 2014. Retrieved from https://www.agreetrust.org/wp-content/uploads/2014/03/AGREE_II_German-Version.pdf (Accessed: 08.03.2022).

Domäne 5: Anwendbarkeit				
18. Mögliche förderliche und hinderliche Faktoren für die Anwendung der Leitlinie werden beschrieben.	☐	☐	☐	
19. Die Leitlinie macht Vorschläge und / oder benennt Instrumente, die die Anwendung der Leitlinienempfehlungen unterstützen.	☐	☐	☐	
20. Die möglichen finanziellen Auswirkungen der Leitlinienempfehlungen wurden berücksichtigt.	☐	☐	☐	
21. Die Leitlinie nennt Messgrößen für die Bewertung der Prozess- und / oder Ergebnisqualität der Leitlinie.	☐	☐	☐	
Domäne 6: Redaktionelle Unabhängigkeit				
22. Die finanzierende Organisation hat keinen Einfluss auf die Inhalte der Leitlinie genommen.	☐	☐	☐	
23. Interessenkonflikte der Mitglieder der Entwicklergruppe der Leitlinie wurden dokumentiert und bei der Leitlinienerstellung berücksichtigt.	☐	☐	☐	

Critical Appraisal von Mixed-Methods-Forschungsarbeiten:

Thieme RECOM

Critical Appraisal von Mixed-Methods-Studien gemäß Kapitel 5 des MMAT[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von Mixed Methods Forschungsarbeiten	ja	nein	nicht beurteilbar	Kommentar
Auszug aus dem Mixed Methods Appraisal Tool (MMAT) Version 2018. Kategorie 5: Spezifische Bewertungskriterien für Mixed Methods Forschung				
5.1 Gibt es eine angemessene Begründung für die Verwendung eines Mixed-Methods-Designs, um die Forschungsfrage zu beantworten?	☐	☐	☐	
5.2 Sind die verschiedenen Komponenten der Studie effektiv und systematisch integriert worden, um die Forschungsfrage zu beantworten?	☐	☐	☐	
5.3 Werden die Ergebnisse der systematischen Integration von qualitativen und quantitativen Komponenten angemessen interpretiert?	☐	☐	☐	
5.4 Werden Divergenzen und Inkonsistenzen zwischen quantitativen und qualitativen Ergebnissen angemessen adressiert und behandelt?	☐	☐	☐	
5.5 Halten sich die verschiedenen Komponenten der Studie an die Qualitätskriterien der jeweiligen Methodentradition?	☐	☐	☐	

(Erläuterungen zu den Fragen vergleiche Folgeseite!)

[1] **Hong, Q. N. et al.** (2018). MIXED METHODS APPRAISAL TOOL (MMAT) - VERSION 2018. User guide. Retrieved from http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/146002140/MMAT_2018_criteria-manual_2018-08-08c.pdf (Accessed: 04.02.2022).

Erklärungen und Interpretationshilfen zu Kategorie 5 des Mixed Methods Appraisal Tool (MMAT) Version 2018
Zu 5.1: Die Gründe für die Durchführung einer Mixed Methods Studie sollten klar dargelegt sein. Es können mehrere Gründe angeführt werden. Etwa, um qualitative Erkenntnisse mit quantitativen Ergebnissen zu ergänzen/bestärken oder darauf aufzubauen – und umgekehrt. Weitere Gründe können beispielsweise ein umfassendes und vollständiges Verständnis eines Phänomens oder die Entwicklung und Testung von Instrumenten sein.
Zu 5.2: Die systematische Integration verschiedener Forschungsdesigns ist eine Kernkomponente der Mixed-Methods-Forschung und wird definiert als „explizite Verzahnung der quantitativen und qualitativen Elemente innerhalb einer Mixed-Methods-Studie". Bewertet werden sollten die bereitgestellten Informationen darüber, wie quantitative und qualitative Phasen, Ergebnisse und Daten integriert wurden. Zum Beispiel, wie die mit beiden Forschungsmethoden gesammelten Daten zu einem Gesamtbild zusammengeführt wurden (z.B. gemeinsame Darstellungen) und wann die Integration erfolgte (z.B., während der Datenerhebung und -analyse oder/und während der Interpretation der qualitativen und quantitativen Ergebnisse).
Zu 5.3: Dieses Kriterium steht im Zusammenhang mit der sog. Metainterferenz, die als Gesamtinterpretation definiert ist, die aus der Integration quantitativer und qualitativer Ergebnisse resultiert. Die Metainterferenz entsteht bei der Interpretation der Ergebnisse aus der Integration der qualitativen und quantitativen Komponenten und zeigt den Mehrwert der Durchführung einer Mixed-Methods-Studie anstelle von zwei getrennten Studien.
Zu 5.4: Bei der Integration der Ergebnisse aus den qualitativen und quantitativen Komponenten können Divergenzen und Inkonsistenzen (auch Konflikte, Widersprüche, Diskordanzen, Diskrepanzen und Dissonanzen genannt) festgestellt werden. Es reicht nicht aus, nur über die Divergenzen zu berichten - sie müssen erklärt werden. Es gibt verschiedene Ansätze zum Umgang mit Divergenzen: • Reconciliation, wenn die Daten neu analysiert und plausibel interpretiert werden können • „Bracketing", dass, wenn die Ergebnisse nicht miteinander vereinbar sind, Extremfälle liefert, um das Bild eines statistischen Konfidenzintervalls zu bestimmen • Initiierung neuer Perspektiven durch die Datenanalyse, ein neues Forschungsprojekt oder eine neue Datenerhebung auf der Grundlage einer neuen Forschungsfrage • Ausschluss, d. h. Weglassen eines Teils der Daten, z.B. wenn die Ergebnisse unvollständig sind oder wenn einige Daten in Bezug auf die Validität nicht ausreichend robust sind Dieses Kriterium ist mit „Ja" zu bewerten, wenn keine Divergenz vorliegt.
Zu 5.5: Die Qualität der qualitativen und der quantitativen Komponente sollte einzeln bewertet werden, um sicherzustellen, dass keine wesentlichen Gefahren für die Vertrauenswürdigkeit bestehen. Zur Bewertung werden die Kriterien für die qualitative Komponente und die entsprechenden Kriterien für die quantitative Komponente herangezogen (Achtung: In der hier beschriebenen Umsetzung die entsprechenden Instrumente zur Bewertung der entsprechenden Forschungsdesigns). Die Qualität beider Komponenten sollte hoch sein, damit die Mixed-Methods-Studie als qualitativ hochwertig angesehen werden kann. Es gilt die Prämisse, dass die Gesamtqualität einer Mixed-Methods-Studie nicht höher sein darf als die Qualität ihrer schwächsten Komponente. Wenn zum Beispiel die quantitative Komponente mit hoher Qualität und die qualitative Komponente mit niedriger Qualität eingestuft wird, resultiert auch die Gesamtbewertung für dieses Kriterium in „niedriger" Qualität.

Critical Appraisal von Fach- und Lehrbüchern:

Thieme RECOM

Critical Appraisal von Fach-/Lehrbüchern und Buchkapiteln[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von Fach-/Lehrbüchern und Buchkapiteln	ja	teils	nein
I. Initiale Bewertung			
1. Autor(en) Ist das Buch/Kapitel zu einem Thema geschrieben, das zum Fachgebiet des Autors gehört? Wurde der Name des Autors bereits in anderen Quellen oder Bibliografien zitiert? Gibt es eine Erklärung über etwaige Interessenkonflikte für jeden Autor?	☐	☐	☐
2. Erscheinungsdatum der Publikation Wann wurde die Publikation veröffentlicht? Ist die Quelle aktuell oder veraltet mit Bezug auf das Thema?	☐	☐	☐
3. Ausgabe / Revision Handelt es sich um eine Erstausgabe dieser Veröffentlichung? Falls nein, Erläutert der Autor, wie und wer frühere Ausgaben überarbeitet hat oder welches Verfahren bei der Überarbeitung neuerer Ausgaben angewandt wurde? (z. B. im Vorwort oder an anderer Stelle im Text)	☐	☐	☐
4. Verlag / Herausgeber Handelt es sich um einen seriösen Verlag/Herausgeber?	☐	☐	☐
II. Inhaltliche Analyse			
5. Angemessenheit für das Zielpublikum Richtet sich die Veröffentlichung an ein spezialisiertes oder ein allgemeines Publikum? Ist diese Quelle zu einfach, zu technisch, zu fortgeschritten oder genau richtig für Ihre Bedürfnisse?	☐	☐	☐
6. Literaturverzeichnis / Bibliografie Sind Literaturangaben enthalten? Sind die zitierten Referenzen von hoher Qualität? Sind die zitierten Referenzen aktuell? Enthalten die Referenzen Arbeiten von Experten auf dem Gebiet oder zum Thema, das Sie interessiert?	☐	☐	☐
7. Objektivität / Argumentation Handelt es sich bei den enthaltenen Informationen um Meinungen oder Fakten Ist der Standpunkt des Autors objektiv und unparteiisch? Ist die verwendete Sprache frei von Emotionen und Vorurteilen? Scheinen die Informationen stichhaltig und gut recherchiert zu sein, oder sind sie fragwürdig und nicht durch Beweise untermauert? Stimmen die vorgebrachten Ideen und Argumente mehr oder weniger mit anderen Publikationen überein?	☐	☐	☐

[1] **Cornell University Library (Ed.)**. (2021). Critically Analyzing Information Sources: Critical Appraisal and Analysis. Retrieved from https://guides.library.cornell.edu/critically_analyzing (Accessed: 15.02.2022).

8. Umfang / Erfassungsbereich Handelt es sich um eine Primärquelle oder Sekundärliteratur? Wenn primär, aktualisiert die Arbeit andere Erkenntnisse/Arbeiten oder untermauert sie andere Veröffentlichungen bzw. fügt neue Informationen hinzu?	☐	☐	☐
9. Nützlichkeit / Praktikabilität Bietet der Inhalt der Veröffentlichung eine vollständige Beschreibung des Themas? Weist der Autor / weisen die Autoren auf offene Fragen hin, die nicht behandelt werden? Besteht ein Gleichgewicht zwischen Theorie und Praxis?	☐	☐	☐
10. Schreibstil Ist die Veröffentlichung logisch aufgebaut? Sind die wichtigsten Punkte klar dargestellt? Wiederholen sich die Argumente des Autors?	☐	☐	☐
III. Gesamtbewertung			
11. Gesamtbewertung Wie würden Sie die Veröffentlichung auf der Grundlage der vorstehenden Beurteilung zusammenfassend bewerten?	stark ☐	mäßig ☐	schwach ☐

Critical Appraisal von grauer Literatur und Webseiten:

Thieme RECOM

Critical Appraisal von grauer Literatur und Webseiten gemäß AACODS[1]

Publikation/ Studie: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 1: Klicken oder tippen Sie hier, um Text einzugeben.

Reviewer 2: Klicken oder tippen Sie hier, um Text einzugeben.

Critical Appraisal von grauer Literatur und Webseiten (AACODS)	ja	nein	nicht beurteilbar
Autorität / Kompetenz	☐	☐	☐
Einzelner Autor / mehrere individuelle Autoren: • Ist der / sind die Autoren mit einer angesehenen Organisation assoziiert? • Besteh eine angemessene berufliche Qualifikation oder umfangreiche Erfahrung? • Wurden andere Arbeiten in diesem Bereich veröffentlicht? • Handelt es sich um einen anerkannten Experten, auf den in anderen Quellen referenziert wird? • Wird der Text von anderen zitiert? (z.B. Google Scholar für eine schnelle Überprüfung) • Im Falle von Abschlussarbeiten: Handelt es sich um einen Hochschulabschluss unter "fachkundiger" Aufsicht? Gruppe oder Organisation: • Ist die Gruppe/Organisation seriös? (z. B. W.H.O.) • Ist die Gruppe/Organisation eine Autorität auf dem Gebiet? In allen Fällen: • Verfügt der Text über ein ausführliches Literaturverzeichnis bzw. eine Bibliografie?			
Genauigkeit	☐	☐	☐
• Hat die Veröffentlichung ein klar formuliertes Ziel oder einen Auftrag? Falls ja, wurde das Ziel erfüllt? • Wurde vorab definierte Methodik benannt? Falls ja, wurde sie befolgt? • Wurde die Veröffentlichung einem peer-review-Verfahren unterzogen? • Wird die Veröffentlichung durch maßgebliche, dokumentierte Referenzen oder glaubwürdige Quellen gestützt? • Ist die Veröffentlichung repräsentativ für den gegebenen Themenbereich? Falls Nein, verfolgt sie es eine gültige Gegenargumentation? • Ist jede Datenerhebung explizit und für die Forschung geeignet? • Wenn es sich um eine Sekundärveröffentlichung, wird auf die Primärliteratur verwiesen. • Handelt es sich um eine genaue, unvoreingenommene Interpretation oder Analyse?			
Umfang/ Reichweite	☐	☐	☐
• Sind etwaige Grenzen der beschriebenen Inhalte/Ergebnisse deutlich angegeben bzw. ist die Reichweite der Veröffentlichung klar beschrieben?			
Objektivität	☐	☐	☐
Es ist wichtig, Verzerrungen (Bias) und/oder Voreingenommenheit zu erkennen, vor allem wenn sie nicht beschrieben werden oder nicht erkannt / wahrgenommen wurden. • Eine Meinung, egal ob sie von einem Experten stammt oder nicht, ist immer noch eine Meinung: Ist der Standpunkt des Autors klar? • Wirkt das Werk in seiner Darstellung ausgewogen?			

[1] **Tyndall, J.** (2010). AAOCDS Checklist. Retrieved from https://dspace.flinders.edu.au/xmlui/bitstream/handle/2328/3326/AACODS_Checklist.pdf?sequence=4&isAllowed=y (Accessed: 15.02.2022).

Datum / Veröffentlichungszeitraum	☐	☐	☐
Damit die Veröffentlichung für die eigene Recherche von Nutzen ist, muss sie innerhalb eines Veröffentlichungszeitraums liegen, der eine zum heutigen Zeitpunkt gegebene Relevanz bestätigt. • Weist die Veröffentlichung ein eindeutiges Datum in Bezug auf die präsentierten Inhalte auf? (Das Fehlen eines leicht erkennbaren Datums ist meist als signifikantes Problem zu werten) • Gibt es einen triftigen Grund für das Fehlen eines Datums, wenn es nicht angegeben ist, aber genau ermittelt werden kann? • Überprüfen Sie das Literaturverzeichnis: Wurde wichtiges zeitgenössisches Material aufgenommen?			
Bedeutung / Signifikanz	☐	☐	☐
• Ist die Veröffentlichung aussagekräftig und bedeutungsvoll? (dies umfasst Machbarkeit, Nutzen und Relevanz) • Fügt sie einen Kontext oder neue Erkenntnisse hinzu oder bereichert sie die Forschung oder fügt sie ihr etwas Einzigartiges hinzu? • Bestärkt oder widerlegt sie eine bestehende Position zum Thema? • Hat die Veröffentlichung das Potenzial für Auswirkungen/Konsequenzen? (im Sinne einer Beeinflussung der Arbeit oder des Verhaltens anderer)			

Anlage III: Protokoll zur Dokumentation einer systematischen Literaturrecherche

Anlass und Kontext der systematischen Literaturrecherche:

-

Durchführende Person(en):

-

Recherchezeitraum:

-

Thieme RECOM

1. Die Fragestellung(en) und deren Konkretisierung

Thema:

-

Gewählter Weg zur Eingrenzung (z.B. orientierende Recherche, Anwendung des PICO-Schemas, etc.) und zentrale Erkenntnisse hieraus:

-

Konkrete Fragestellung(en):

-

Thieme RECOM

2. Ableitung von Suchbegriffen

Die Vermehrung von Suchbegriffen erhöht die Aussicht auf relevante Treffer! Relevante Stellschrauben: Synonyme, Ober- und Unterbegriffe, englische Übersetzungen und deren Synonyme, Beachtung verschiedener Flexionsformen.

Alle Kernelemente/Kernkonzepte der Fragestellung(en) sollten nach diesem Vorgehen zunächst für sich betrachtet werden, um die resultierenden Suchbegriffe im Anschluss sinnvoll zu Suchphrasen verbinden zu können.

Als Hilfskonstrukt eignet sich ggf. die folgende Matrix.

	Kernkonzept 1 Pflegeheim-bewohner(in)	Kernkonzept 2 Herausfordern des Verhalten	Kernkonzept 3 Demenz	Kernkonzept 4 (Integrative) Validation	Kernkonzept 5 Snoezelen
Oberbegriff					
Unterbegriff					
Synonyme					
englische Übersetzung					

Thieme RECOM

3. Definieren von Ein- und Ausschlusskriterien

Ziel dabei: Klärung des formalen Rahmens der Recherche und Fokussierung auf das, was wirklich von Interesse und praktikabel ist (zusätzliche Filterung).

Mögliche Kriterien hierbei sind beispielsweise:

- Sprache der Publikationen
- Erscheinungszeitraum
- Publikations- / Studientypen
- Altersgruppen, Geschlecht u.Ä.
- Datenbanken / Suchorte
- Settings, Professionen,
- ...

Wichtig: Alle Ein- bzw. Ausschlusskriterien müssen begründbar sein!!

Einschlusskriterien	Ausschlusskriterien

Thieme RECOM

4. Definition der Fachdatenbanken und weiterer Anlaufstellen zur Recherche

In Abhängigkeit des zu bearbeitenden Themas ist auch die Frage danach, „wo“ nach Literatur recherchiert wird, zu beantworten.

Abbildung 11 : Wesentliche Anlaufstellen und Möglichkeiten zur systematischen Literaturrecherche (Quelle: eigene Darstellung in Anlehnung an Simon, 2018)

Auswahl elektronischer Fachdatenbanken mit kurzer Begründung:

-

Auswahl von Bibliothekskatalogen / Freihandaufstellungen mit kurzer Begründung:

-

Ggf.: Auswahl von Leitlinienkatalogen (z.B. AWMF etc.) mit kurzer Begründung:

-

Auswahl weiterer Rechercheorte (z.B. Fachzeitschriften, graue Literatur, Suchmaschinen im Internet, Schneeballsystem, etc.) mit kurzer Begründung:

-

5. Entwicklung von Suchphrasen und Alternativen

Ein wesentlicher Schritt, um die Durchführung der systematischen Literaturrecherche möglichst zielgerichtet, handhabbar und „vollständig ohne „blinde Flecken“ zu gestalten, aber auch, um auf oft nicht exakt vorhersehbare Suchergebnisse sowohl hinsichtlich Qualität als auch hinsichtlich Quantität adäquat reagieren zu können (zu viele Treffer, zu wenige Treffer, die meisten Ergebnisse am Thema vorbei, ...), ist die Generierung von verschiedenen Suchphrasen. In diesem Zuge ist es ratsam, neben den mutmaßlich „eigentlichen“ Suchphrasen auch Alternativen mitzudenken, um die Suchläufe z.B. einzuengen oder zu erweitern. Zur Entwicklung von Suchphrasen sollten folgende Techniken bzw. Aspekte berücksichtigt werden:

- **Feldsuche** (Nutzung des gegliederten Aufbaus der bibliographischen Angaben in einer Datenbank, z.B. Suche nach Autor, nach Titel, nach Abstract, ...)
- **Filterung der Suche**, im Wesentlichen nach den Ein- / Ausschlusskriterien (z.B. Veröffentlichungszeitraum, Sprache, Altersgruppen, ...)
- Nutzung der **Schlagwortsuche (bzw. MeSH-Terms)**, die eine Publikation charakterisieren bzw. die schnelle Erschließung der Inhalte einer Publikation ermöglichen. Die Schlagworte selbst müssen dabei nicht zwingend im eigentlichen Text der Publikation enthalten sein.
- **Trunkierung und Maskierung**, d.h. die Nutzung von Platzhaltern, um die Eingabe aller möglichen grammatikalischen Varianten eines Suchbegriffs bzw. von Wortkombinationen zu vermeiden. Die Trunkierung (häufig mit dem Zeichen *) sucht nach Wortstämmen mit beliebiger Endung bzw. Worterweiterung, die Maskierung ersetzt hingegen ein einziges Zeichen innerhalb eines Suchbegriffs (häufig mit dem Zeichen # oder ? umgesetzt)
- **Boolesche Operatoren**, welche die zielgerichtete Verknüpfung der Suchbegriffe miteinander ermöglichen. Die Booleschen Operatoren sind mit die mächtigste Funktion bei der Durchführung einer systematischen Literaturrecherche, weswegen deren Einsatz lieber doppelt als nur einfach geprüft werden sollte (schnelle Fehlerquelle). Die wichtigsten Operatoren sind:
    - ✓ **AND** (selten auch das deutsche UND. Zeigt nur Treffer, die alle derart verknüpften Suchbegriffe enthalten – „sowohl als auch“)
    - ✓ **OR** (selten auch das deutsche ODER. Zeigt Treffer, die entweder den Suchbegriff A, oder den Suchbegriff B oder den Suchbegriff C (...) enthalten – „entweder oder“)
    - ✓ **NOT** (selten auch das deutsche NICHT. Schließt alle Treffer aus den Suchergebnissen aus, die den oder die entsprechenden Begriffe enthalten – „auf keinen Fall“).

- **Phrasensuche**, mit der die eingegebenen Suchbegriffe in ausschließlich der angegebenen Reihenfolge gesucht und berücksichtigt werden. Oft sehr hilfreich bei der Suche nach Konzepten mit einer feststehenden Kombination aus mehreren Wörtern (z.B. „pflegende Angehörige“). Oft wird die Phrasensuche durch das Setzen der Suchbegriffe in Anführungszeichen („“) umgesetzt oder durch die Aktivierung einer entsprechenden Option in der Suchmaske der Datenbank (z.B. Funktion „exakte Wortfolge“).

Thieme RECOM

Zu nutzende Suchphrasen und mögliche Alternativen:

Teilsuchphrase zu Kernkonzept 1			
Verknüpfender Operator (AND, OR)	Suchbegriff (ggf. trunkiert, maskiert oder als Phrase)	Feldsuche? (Title, MeSH, ...)	Alternativen?
	Hauptbegriff		
OR	Synonym 1		
OR	Synonym 2		
OR	Synonym 3		
Bei diesem Kernkonzept auszuschließende Begriffe?			
NOT			

Teilsuchphrase zu Kernkonzept 2			
Verknüpfender Operator (AND, OR)	Suchbegriff (ggf. trunkiert, maskiert oder als Phrase)	Feldsuche? (Title, MeSH, ...)	Alternativen?
	Hauptbegriff		
OR	Synonym 1		
OR	Synonym 2		
OR	Synonym 3		
Bei diesem Kernkonzept auszuschließende Begriffe?			
NOT			
Teilsuchphrase zu Kernkonzept 3			
Verknüpfender Operator (AND, OR)	Suchbegriff (ggf. trunkiert, maskiert oder als Phrase)	Feldsuche? (Title, MeSH, ...)	Alternativen?
	Hauptbegriff		
OR	Synonym 1		
OR	Synonym 2		
OR	Synonym 3		
Bei diesem Kernkonzept auszuschließende Begriffe?			
NOT			

Verknüpfung der Teilsuchphrasen zu den einzelnen Kernkonzepten der Forschungsfrage(n) zu einer Gesamt-Suchphrase und Erarbeitung möglicher Alternativen, falls Suchergebnis zu eng, zu breit oder thematisch unangemessen:

Gesamtsuchphrase – Variante 1				
	Verknüpfender Operator (AND, OR, NOT)		Verknüpfender Operator (AND, OR, NOT)	
Teilsuchphrase zu Kernkonzept 1	AND	Teilsuchphrase zu Kernkonzept 2	AND	Teilsuchphrase zu Kernkonzept 3
Bei der Gesamtsuchphrase auszuschließende Begriffe?				
			NOT	
Zu nutzende Suchfilter (Ein- / Ausschlusskriterien)				
Publikationszeitraum	Sprache	Studien- / Publikationsdesigns	Altersgruppe	...

Gesamtsuchphrase – Variante 2				
	Verknüpfender Operator (AND, OR, NOT)		Verknüpfender Operator (AND, OR, NOT)	
Teilsuchphrase zu Kernkonzept 1	AND	Teilsuchphrase zu Kernkonzept 2	AND	Teilsuchphrase zu Kernkonzept 3
Bei der Gesamtsuchphrase auszuschließende Begriffe?				
			NOT	
Zu nutzende Suchfilter (Ein- / Ausschlusskriterien)				
Publikationszeitraum	Sprache	Studien- / Publikationsdesigns	Altersgruppe	...

Eine sinnvolle optionale Alternative zur tabellarischen Darstellung der Suchphrase(n) ist ggf. die rein textliche Umsetzung bzw. der rein textliche Entwurf derselben. Hierbei ist wichtig, ähnlich wie in der Mathematik eine akkurate Klammersetzung zu berücksichtigen. Klammern drücken dabei den Vorrang in der Bearbeitung der Suchphrase durch die Datenbank aus, Elemente innerhalb einer Klammer werden zunächst abgefragt, anschließend die Elemente außerhalb einer Klammer. In der Regel ist auch die Verschachtelung mehrerer Klammerpaare möglich. Sollen einzelne Begriffe in spezifischen Suchfeldern (z.B. nur im Titel, nur im Abstract, ...) gesucht werden, so wird dies je nach verwendeter Fachdatenbank unterschiedlich ausgedrückt, oft werden hierzu eckige Klammern [] mit dem entsprechenden Suchfeld genutzt.

Zur Illustration ein kurzes Beispiel:

*(dementia[title/abstract] OR Alzheimer [title/abstract] OR "cognitive impairment"[title/abstract]) AND ("nursing home" OR eldercare OR "geriatric nursing") NOT hospital**

Thieme RECOM

Suchphrase(n) in Textform:

-

6. Dokumentation der Resultate einzelner Suchläufe im Verlauf und der final genutzten Suchphrase(n)

Zur Nachvollziehbarkeit des Rechercheverlaufs inklusive der jeweiligen Änderungen an Quantität und Qualität der Suchergebnisse in verschiedenen Fachdatenbanken (und ggf. mittels anderer Suchorte) empfiehlt es sich, die Historie aller durchgeführten Suchläufe zu dokumentieren. Hierzu bieten alle größeren Fachdatenbanken und Kataloge die Möglichkeit, den Sucherverlauf einzusehen und zu kopieren (beispielhaft vgl. Screenshot)

History and Search Details — Download — Delete

Search	Actions	Details	Query	Results	Time
#2	•••	>	Search: **(dementia[title/abstract] OR Alzheimer [title/abstract] OR "cognitive impairment"[title/abstract]) AND ("nursing home" OR eldercare OR "geriatric nursing") NOT hospital***	4,094	06:04:31
#1	•••	>	Search: **dementia**	242,516	03:33:48

Suchverlauf am <Datum> in der Datenbank <Name>:

-

Suchverlauf am <Datum> in der Datenbank <Name>:

-

Suchverlauf am <Datum> in der Datenbank <Name>:

-

Finale Suchphrase(n) vom <Datum>, mittels der das Abstract-Screening gestartet wurde:

-

Thieme RECOM

7. Dokumentation der Ergebnisse des Abstract-Screenings und der Volltextbeschaffung:

Abstract-Screening des finalen Suchlaufs in der Datenbank <Name> am <Datum>:

- Von insgesamt <XX> Suchtreffern wurden insgesamt <YY> Publikationen als relevant für die Anschaffung im Volltext bewertet.
- Von diesen <YY> relevanten Publikationen konnten im Verlauf <ZZ> tatsächlich als Volltext beschafft werden.

Abstract-Screening des finalen Suchlaufs in der Datenbank <Name> am <Datum>:

- Von insgesamt <XX> Suchtreffern wurden insgesamt <YY> Publikationen als relevant für die Anschaffung im Volltext bewertet.
- Von diesen <YY> relevanten Publikationen konnten im Verlauf <ZZ> tatsächlich als Volltext beschafft werden.

Abstract-Screening des finalen Suchlaufs in der Datenbank <Name> am <Datum>:

- Von insgesamt <XX> Suchtreffern wurden insgesamt <YY> Publikationen als relevant für die Anschaffung im Volltext bewertet.
- Von diesen <YY> relevanten Publikationen konnten im Verlauf <ZZ> tatsächlich als Volltext beschafft werden.