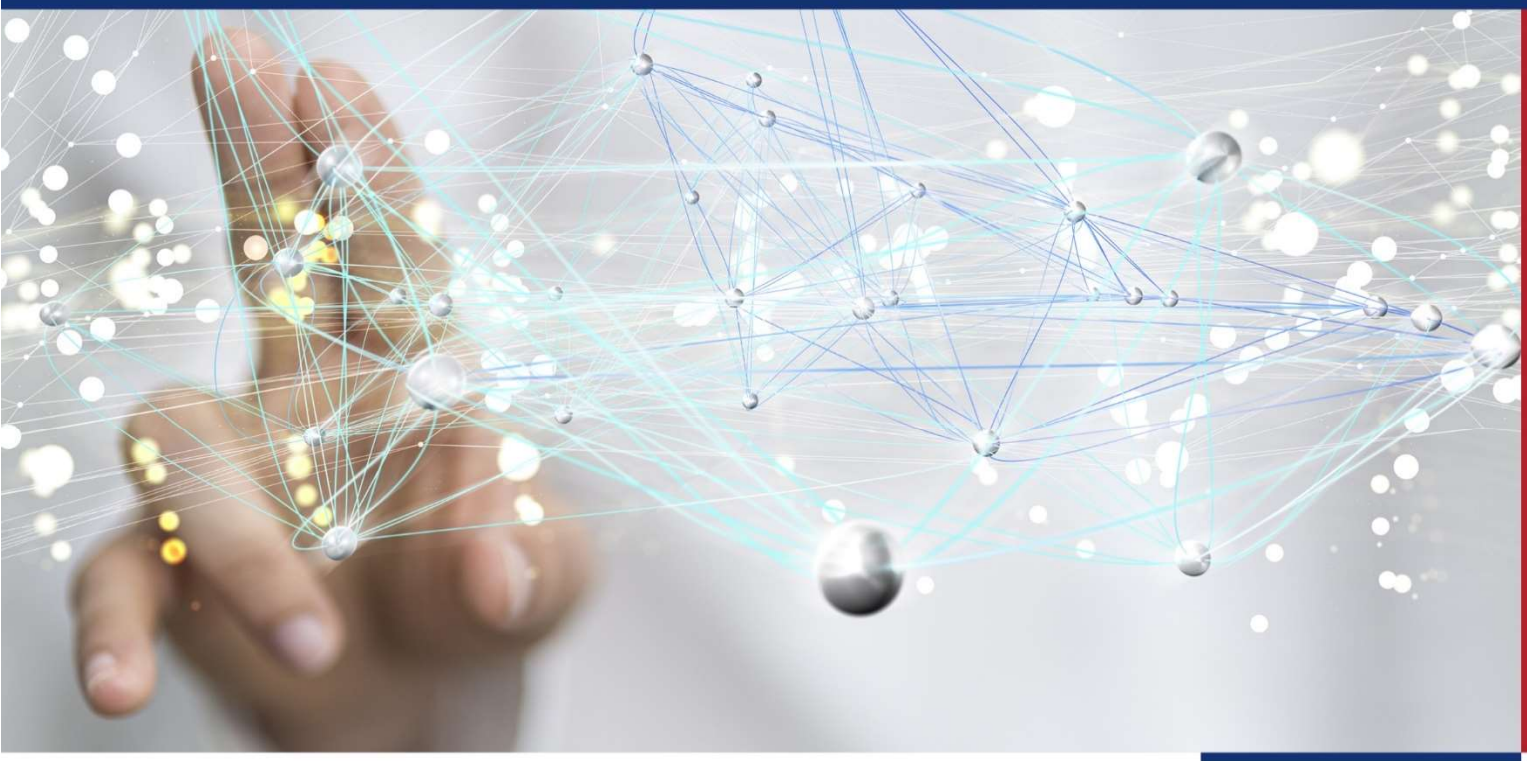
Development backgrounds: Fundamentals and methods

European Nursing care Pathways



Last edited: 2022

Dr. Pia Wieteck

Sebastian Kraus, MSc.

Sabine Kopp, BSc

Dr. Qiumei Jiang-Siebert

Isabella Schmidpeter, MPH

Thieme

RECOM

Contents

1. The evidence-based further development of ENP 3
2. The systematic search for the best possible evidence 4
2.1 Strategy and core elements of the systematic, international literature search 4
2.1.1 The research question(s) 5
2.1.2 Derivation of search terms 6
2.1.3 Inclusion and exclusion criteria 7
2.1.4 Definition of the specialist databases and other offers for research 8
2.1.5 Development of search phrases and possibilities to influence the research process 9
2.1.6 Abstract screening and full text retrieval 12
2.2 Criteria and tools for evidence assessment, evidence selection, and data extraction 13
2.2.1 Critical appraisal of publications in the ENP development process 14
2.2.2 Evidence synthesis and evaluation of individual items/intervention concepts 23
2.2.3 Evidence tables for study evaluation in the context of ENP (further) development 24
2.3 The further course of the ENP development process 34
References 35
Appendix I: Evidence table as part of the ENP development from 2022 41
Appendix II: Former Critical Appraisal Forms (2004-2021) 43
Appendix III: Protocol for the documentation of a systematic literature search 53

1. The evidence-based further development of ENP

European Nursing care Pathways (ENP) is a standardized nursing language that has monohierarchical structures and classifies and makes available nursing expertise in the form of so-called practice guidelines. The following figure shows the systematic development process of ENP with its basic steps, which has been and is being continuously improved. A new database version of ENP is made available every year in mid-year, and book publications are usually realized every three years.

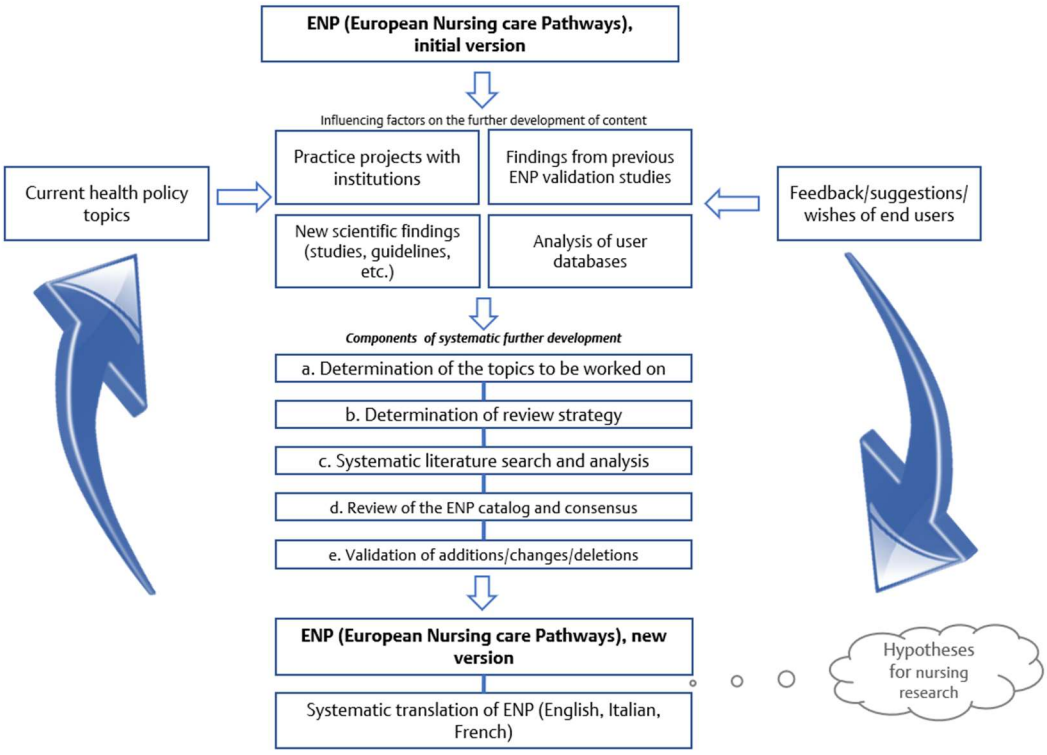


Figure 1: Diagrammatic illustration of the systematic further development process of ENP

Influenced by health policy decisions, user feedback, and new scientific evidence from nursing and neighboring health care disciplines, a decision is made annually as to which ENP practice guidelines will undergo systematic review and, if necessary, revision. The development quality of nursing diagnosis-related practice guidelines is significant for the users of the nursing classification, as it enables a high degree of certainty for evidence-based practice in the nursing decision-making process. In this context, it is therefore all the more important to present the scientific working methods and procedures for the further development of existing ENP practice guidelines and the development of new ones as transparently as possible.

2. The systematic search for the best possible evidence

The systematic further or new development of an ENP practice guideline usually takes place in a multi-stage process. Among others, the following questions are relevant:

- Is the observed phenomenon a nursing diagnostic concept?
- Can evidence-based characteristics/risk factors/etiologies, resources be found in the literature that confirm the nursing diagnosis or have a causal/influencing relationship with the nursing diagnosis?
- What are the nursing objectives in focus?
- Can the intervention concepts for resolving and/or alleviating a nursing problem of a nursing diagnosis be confirmed by the literature?
- What, among other things, is the resource usage in terms of time for a new nursing intervention and what linkages (mappings) with other systems and other instruments are required?
- ...

A core element in answering these questions is to plan and conduct an international, systematic literature research and review. Depending on the particular issue to be addressed in the development and/or update of the ENP practice guideline, a comprehensive search strategy will be developed in the first step, which will be used to search widely for relevant publications without any "blind spots". At the same time, a carefully considered search strategy should make it possible to react adequately and prepared to different circumstances in the course of the literature search, for example with regard to quality (e.g. too many search results deviating from the topic) but also quantity of the search results or number of hits (e.g. apparently too few search results if the search strategy is too narrow or too many and no longer manageable number of hits if the search strategy is too open). The following chapters address in detail the development of a search strategy, the actual conduct of the literature search, and, in particular, the individual evidence assessment of the searched publications in terms of a critical appraisal depending on the type of publication or the underlying research design.

2.1 Strategy and core elements of the systematic, international literature search

In the context of ENP further development, the classic steps of a systematic, international literature search will be followed (cf. e.g. Mayer, Raphaelis, & Kobleder, 2021; Nordhausen & Hirt, 2020) and implemented in all conscience. These can be summarized as follows:

1. narrowing the subject
2. formulation of research questions or hypotheses that are as concrete and specific as possible
3. deriving suitable search terms from the research questions and expanding them (e.g. synonyms, English translations, etc.)
4. definition of inclusion and exclusion criteria (e.g. inclusion/exclusion of certain study designs, limitation of publication periods and languages of publications, etc.).
5. defining and determining the specialist databases to be used for the research and other offers for the research
6. documentation of the search strategy as well as the entire research process
7. carrying out the literature research
8. review and evaluation of search results (abstract screening)
9. procurement and management of relevant literature
10. critical appraisal of the full texts (cf. chapter 2.2)

All steps and decisions are documented and consented at important stages in the research process within the development team. For this purpose, a protocol for the documentation of a systematic literature search was created, which can be viewed in Appendix III, and which is to be kept for each implementation of a systematic search. Depending on the number and quality of the search results, the search strategy is refined during abstract screening with the goal of achieving a manageable result that is also as comprehensive as possible. The processual course of a systematic literature search can therefore also be set back again by several steps in the course. In the case of smaller, selective questions, such as a review

of a possible addition of a characteristic to a nursing diagnosis, it may also be decided that a "small" literature search will be conducted. In this case, the protocol for systematic literature research can be neglected and the sources found can be evaluated according to the evidence tables. All individuals involved in the development process have a nursing background and a degree in at least one nursing-related course of study.

2.1.1 The research question(s)

The research question plays a critical role in the evidence search and data extraction. The clearer it is what is to be searched for, the more successful the search usually is. The starting point for the research process to create or update an ENP practice guideline and/or elements of it is therefore to formulate one or more questions as explicitly and precisely as possible.

Proven method to narrow down a topic or to concretize a question are, for example, an orienting literature research (start reading a topic with a targeted but only limited systematic approach) as well as the specification of different framework conditions or factors in which the topic of interest is embedded. Appropriate aspects may be, for example:

- the target group or the group of people of interest
- the setting
- the context
- the (nursing) action, the phenomenon, the situation (...) of interest
- the classic questions: who, when, what, how, why?
- The distinction from related subject areas

When formulating a question, it must be considered that, on the one hand, it must be answerable, but on the other hand, its structure must not already anticipate a (presumed) result. It is also essential to ensure that a question does not express any presuppositions and should always be formulated as open-ended as possible so as not to work (unconsciously) towards the presupposition. For example, the question "Which of the nursing concepts A or B has a more beneficial effect on the quality of life of people with dementia?" would be preferable to the question "Does nursing intervention A actually have a better effect than nursing intervention B on the quality of life of people with dementia?".

Particularly in the clinical context and for questions relating to nursing intervention concepts, including within ENP, the PICO scheme[1], which has been firmly established in evidence-based medicine for a long time, is useful for structuring and specifying questions (Behrens & Langer, 2004; Panfil, Ivanović, & Conrad, 2011). PICO is an acronym for the parameters "Patient / Population" - "Intervention" - "Control / Comparison" - "Outcome". According to this principle, a question should ask aspects about all these parameters:

- P (population/patient): description of the target group
- I (intervention): technology, process or measure as the subject of investigation or research
- C (control/comparison): alternative(s) with which the technology, process or measure can be compared
- O (outcome): target value What is to be achieved and how can the result be measured?

If nursing intervention concepts are the focus of interest in ENP development, the PICO scheme is basically applicable in an unmodified form. However, the nursing diagnostic context must always be considered, as the following examples illustrate:

[1] In German-speaking countries also synonymously referred to as PIKE scheme as an acronym for the German-language terms Patient, Intervention, Control Intervention, and Outcome.

	Question
Population / care receiver / nursing diagnostic context	Do patients with dementia in hospital have
Intervention	through the establishment of a night café
Control	compared to clinics without such an offer
Result	changed results on the Mini Mental Status Test (MMST)?

Table 1: Example 1 for a question on nursing interventions according to the PICO scheme

	Question
Population / care receiver / nursing diagnostic context	Do care recipients suffering from fatigue experience
Intervention	through complementary nursing intervention of acupressure
Control	compared to no or other complementary nursing interventions
Result	a change in subjectively perceived fatigue?

Table 2: Example 2 for a question on nursing interventions according to the PICO scheme

However, since the new or further development of ENP practice guidelines does not exclusively refer to nursing intervention concepts, but also the nursing diagnostic components in the form of the diagnosis itself as well as associated characteristics, etiologies and resources represent an essential focus, the PICO scheme can be adapted so that it can also be used as an aid for the development of questions in this context, as the following example shows:

	Question
Population / care receiver / nursing diagnostic context	Care receivers who suffer from fatigue
Observable / measurable signs and symptoms	have which measurable, observable or reported characteristics and symptoms
Result	that allow an unambiguous verification of the presence of the nursing diagnosis "...suffers from fatigue"?

Table 3: Example of a question for researching characteristics for a nursing diagnosis based on the PICO scheme

2.1.2 Derivation of search terms

A question that is formulated as concretely and precisely as possible forms an ideal starting point for deriving search terms or keywords that can be used in a database search. Usually, a question can be divided into several contextual aspects or core concepts, which should also be recorded as individual core concepts with regard to the derived search terms. An example:

“How does the use of snoezelen compare to integrative validation on the occurrence of challenging behaviors in nursing home residents with dementia?”

As core concepts and central search terms of this exemplary question can be identified and derived:

- nursing home residents with dementia (population)
- snoezelen (intervention)
- integrative validation (control intervention)
- challenging behaviors (result)

Increasing the number of key search terms for each core concept of the research question usually significantly increases the prospect of finding relevant hits and at the same time reduces the risk of “blind spots” when conducting the literature search! Relevant parameters are in particular synonyms, generic and subordinate terms as well as English translations and their synonyms. Also, considering different inflection forms. According to this procedure, all core elements/core concepts of the research question(s) should first be considered separately in order to be able to combine the resulting search terms into search phrases in a meaningful way. The creation of a matrix has proven to be a practical tool for the identification and increase of suitable search terms and keywords. In the following example, which addresses the question just presented, the fields are only exemplary and not completely filled in.

	Core concept 1 Nursing home resident	Core concept 2 Challenging behavior	Core concept 3 Dementia	Core concept 4 (Integrative) validation	Core concept 5 Snoezelen
Generic term	Nursing home, care receiver	Behavior	Cognitive impairment	Communication	Sensory stimulation, layout of surroundings
Subordinate term	--	Wandering, aggression, ...	Alzheimer's disease	Acceptance	--
Synonyms	--	Behavioral problem	--	--	--
English translations	resident, nursing home, ...	challenging behavior, defiantness, ...	dementia, Alzheimer disease, mental disorder, ...	validation, ...	Snoezelen, snoezeling

Table 4: Example of a matrix for the identification and increase of relevant search terms

2.1.3 Inclusion and exclusion criteria

The formulated inclusion and exclusion criteria for the questions are decisive for the data extraction. The aim is to clarify the formal framework of the research and focus on what is really of interest and practicable (additional filtering). Possible criteria here include: the language of the publications, the publication period, different publication/study types of interest, the focus on certain age groups, genders, settings, and/or professions, and the selection of specialist databases and other search locations to be used. An important criterion in the selection process is that all inclusion or exclusion criteria used must be justifiable and documented with the appropriate reasoning. For the example given, useful inclusion or exclusion criteria could be as follows.

Inclusion criteria	Exclusion criteria
Age group: > 65 years (Reason: presenile dementias are very rare).	Publication types: exclusion of field reports, individual case studies, etc. (Reason: only little significance regarding effectiveness expectable)
Publication period: from 1990 (Reason: neither validation nor snoezelen had been used to any significant extent before).	
Languages of publications: German and English (Reason: professional translations are very expensive and time-consuming)	

Table 5: Exemplary presentation of inclusion and exclusion criteria

The final formulation of inclusion and exclusion criteria is often only possible in the course of conducting the literature search, since in many cases only at this point do thematic or formal aspects emerge that can again raise the question of consideration or exclusion.

2.1.4 Definition of the specialist databases and other offers for research

A central aspect in the development of a search strategy is the question of “where” to search for literature with the greatest possible chance of success. In addition to classic offers such as library catalogs or bibliographies, there is an enormous and almost inexhaustible variety of tools and options available for research, especially with regard to research via the Internet. It is obvious in the context of scientific work that a (mostly unstructured) research in general search engines such as Google has a much lower significance here than, for example, the electronic research in a specialized database. The following figure summarizes the most important possibilities and offers for systematic literature research:

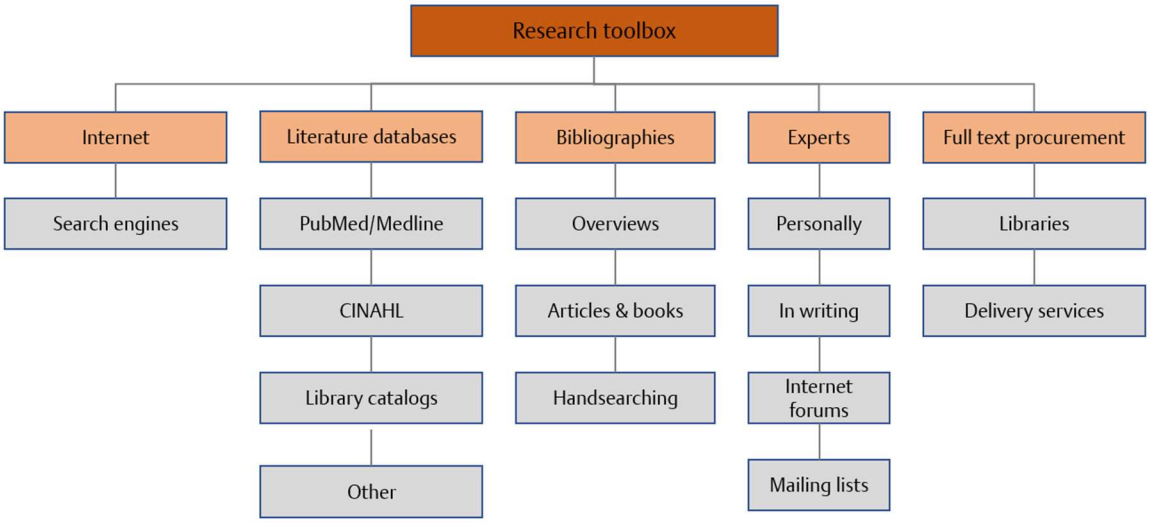


Figure 2: Essential offers and possibilities for systematic literature research (Simon, 2018)

In the context of ENP development, the most significant offers for literature research are undoubtedly electronic specialist databases such as PubMed/MEDLINE, CINAHL (Cumulative Index to Nursing and Allied Health Literature), the Cochrane Library, GeroLit or PsycNet and many more. In addition, nursing and medical guidelines from relevant associations, societies, and publishers are very significant sources of knowledge for ENP development, as they are systematically developed overviews that are compiled and reviewed in an elaborate process and provide evidence-based recommendations. They do not have a binding or obligatory character, but compile the state of knowledge on a usually very specific issue in a compact form and offer evidence-based solutions to the respective topic. Examples of relevant guideline sources include the Association of the Scientific Medical Societies (Arbeitsgemeinschaft der wissenschaftlichen medizinischen Fachgesellschaften, AWMF), the Center for Quality in Nursing (Leitlinien- und Standarddatenbank des Zentrums für Qualität in der Pflege, ZQP) guideline and standards database, the National Institute for Health and Clinical Excellence (NICE), the Registered Nurses' Association of Ontario (RNAO) Nursing Best Practice Guidelines, and the Scottish Intercollegiate Guideline Network (SIGN).

The systematic literature search in specialist databases, guideline catalogs, libraries and other structured offers can be supplemented by research in other ways if necessary. One example is “Google Scholar”, a special derivative of the widely used search engine Google, which is used for general searches for scientific documents. The so-called "snowball system", also known as footnote chasing, can be a useful addition to the research strategy. This refers to the search in the literature references of newly researched or already known specialist publications. Especially in the case of systematic overviews, this procedure can be promising in order to detect further publications relevant to one's own research question or even new search terms for further research runs.

As part of the documentation of the systematic literature research, the following points regarding the selection of specialist databases and other offers should be recorded on the research protocol:

- The selection of electronic specialist databases with short reasoning
- The selection of library catalogs/open access holdings with short reasoning.
- The selection of guideline catalogs (e.g., AWMF, etc.) with short reasoning.
- The selection of other research sites (e.g., trade journals, gray literature, Internet search engines, snowball system, etc.) with short reasoning

2.1.5 Development of search phrases and possibilities to influence the research process

An essential step in order to make the performance of the systematic literature search as targeted, manageable and complete as possible without "blind spots", but also in order to be able to react adequately to search results that are often not exactly predictable, both in terms of quality and quantity (too many hits, too few hits, most results missing the point, ...), is the generation of different search phrases. In this course, it is advisable to consider alternatives in addition to the presumed "actual" search phrases, e.g., to narrow or broaden the search runs. To develop search phrases, the following techniques or aspects should be considered:

- Field search (use of the structure of the bibliographic data in a database, e.g. search by author, by title, by abstract, ...)
- Filtering of the search, basically according to the inclusion/exclusion criteria (e.g. publication period, language, age groups, ...)
- Use of keyword searches (or MeSH terms[2]) that characterize a publication or enable a quick understanding of the contents of a publication. The keywords themselves do not necessarily have to be included in the actual text of the publication.
- Truncation and masking, means the use of placeholders to avoid entering all possible grammatical variants of a search term or word combinations. Truncation (often implemented with the * character) searches for word stems with any ending or word extension, while masking replaces a single character within a search term (often implemented with the # or ? character)
- Boolean operators, which allow the targeted linking of search terms with each other. Boolean operators are one of the most powerful features in conducting a systematic literature search, which is why their use should be double-checked rather than single-checked (common source of error). The most important operators are:
    - AND (rarely also the German UND. Shows only hits that contain all such linked search terms - "as well as").
    - OR (rarely also the German ODER. Shows hits that contain either the search term A or the search term B or the search term C (...) - "either or").
    - NOT (rarely also the German NICHT. Excludes all hits from the search results that contain the corresponding term(s) - "in no case").

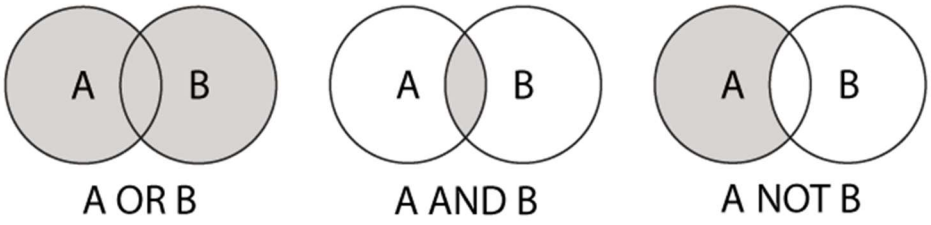


Figure 3: The Boolean operators and their effects in the literature search (source: own illustration)

- The phrase search, with which the entered search terms are searched for and considered in the specified order only. Often very helpful when searching for concepts with a fixed combination of multiple words (e.g., "caring relatives"). Often the phrase search is implemented by placing the search terms in quotation marks ("") or by activating a corresponding option in the search mask of the database (e.g. function "exact word order").

In many cases, it is purposeful to first divide the creation of search phrases into the individual core concepts of the research question for research in specialist databases and catalogs, in order to finally bring them together as part of the final search runs. For the construction of the partial search phrases as well as the final overall search phrase, it is again possible to use a matrix or several matrices in which both the respective search terms and the various search techniques such as field search, truncation and the use of Boolean operators can be brought together. The following matrices exemplify the construction of partial search phrases and the construction of an overall search phrase, respectively, incompletely on the basis of the problem discussed so far.

[2] MeSH is an acronym for "Medical Subject Headings" and describes a thesaurus for collecting journal articles and books in medicine and the life sciences. It is mainly used with the MEDLINE (PubMed) database.

Partial search phrase for core concept 1 Nursing home resident			
Linking operator (AND, OR)	Search term (if necessary truncated, masked or as a phrase)	Field search? (Titel, MeSH, ...)	Alternatives?
	Main term resident	Field search in title and abstract, if necessary	
OR	*Synonym 1 "nursing home"*	Field search in title and abstract, if necessary	
OR	*Synonym 2 "Elder care"*	Field search in title and abstract, if necessary	"elderly care", "geriatric care"
OR	*Synonym 3*		
Terms to be excluded in this core concept?			
NOT	*Hospital OR "outpatient care" OR "home care"*		

Table 6: Construction of a partial search phrase to a core concept of the research question

Partial search phrase for core concept 2 Challenging behavior			
Linking operator (AND, OR)	Search term (if necessary truncated, masked or as a phrase)	Field search? (Titel, MeSH, ...)	Alternatives?
	Main term "challenging behavior"	Field search in title and abstract, if necessary	
OR	*Synonym 1 defiantness*	Field search in title and abstract, if necessary	
OR	*Synonym 2 wandering*	Field search in title and abstract, if necessary	
OR	*Synonym 3 aggressi**		
Terms to be excluded in this core concept?			
NOT			

Table 7: Construction of a partial search phrase for another core concept of the research question

Once search phrases and possible alternatives have been created for all core concepts of the question, the task is to combine all aspects into one or more alternative overall search phrases. For this purpose, the search history in the individual specialist databases can often be used (e.g. combine "Search run for core concept 1" with the operator "AND" with the "Search run for core concept 2"). In the sense of a comprehensible documentation as well as the often given necessity to present the final search runs elsewhere, also all "total searches" as well as alternative search runs (if the search result is too broad, too narrow or thematically inappropriate) should be considered in advance and documented in tabular form as far as possible. When combining the partial search phrases into one or more overall search runs, the application of the inclusion and exclusion criteria is also generally used.

Overall search phrase - variant 1				
	Linking operator (AND, OR, NOT)		Linking operator (AND, OR, NOT)	
Partial search phrase for core concept 1	*AND*	*Partial search phrase for core concept 2*	*AND*	*Partial search phrase for core concept 3*
				Terms to be excluded from the overall search phrase?
			NOT	
Search filters to be used (inclusion/exclusion criteria)				
Publication period	Language	Study/publication designs	Age group	...

Overall search phrase - variant 2				
	Linking operator (AND, OR, NOT)		Linking operator (AND, OR, NOT)	
Partial search phrase for core concept 1	*AND*	*Partial search phrase for core concept 2*	*AND*	*Partial search phrase for core concept 3*
				Terms to be excluded from the overall search phrase?
			NOT	
Search filters to be used (inclusion/exclusion criteria)				
Publication period	Language	Study/publication designs	Age group	...

Table 8: Example of a matrix for creating search phrases to merge the searches into individual core concepts of the research question

A sensible optional or possibly supplementary alternative to the tabular presentation of the search phrase(s) may be the purely textual realization or the purely textual design of the same in the form of a database-specific syntax. Here, similar to mathematics, it is important to consider accurate bracket setting as well as the peculiarities and techniques of the respective database. Brackets express the priority in the processing of the search phrase by the database, elements inside a bracket are queried first, then the elements outside a bracket. As a rule, the nesting of several pairs of brackets is also possible. If individual terms are to be searched for in specific search fields (e.g. only in the title, only in the abstract, ...), this is expressed differently depending on the specialist database used; square brackets [] with the corresponding search field are often used for this purpose. When manually compiling search phrases using a syntax, great care and ideally verification by a second person is always required, as it is very easy for errors to sneak in this way with massive effects on the search results. For illustration, a short example of a research syntax with reference to the exemplary question:

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

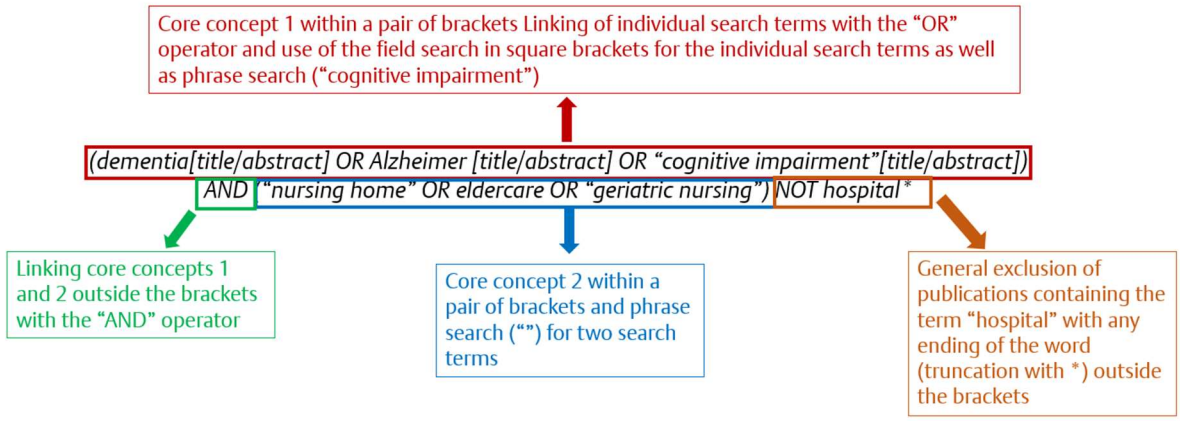


Figure 4: Textual creation of a search phrase connecting several core concepts in the form of a database-specific syntax

In order to be able to trace the research process, including the respective changes to the quantity and quality of the search results in different specialist databases (and, if necessary, by means of other search locations), it is advisable to also document the history of all searches carried out. For this purpose, all major specialist databases and catalogs offer the possibility to view and copy the search history (for an example, see screenshot) or to save it for later use. The latter usually requires the creation of a mostly free user account at the respective specialist database.

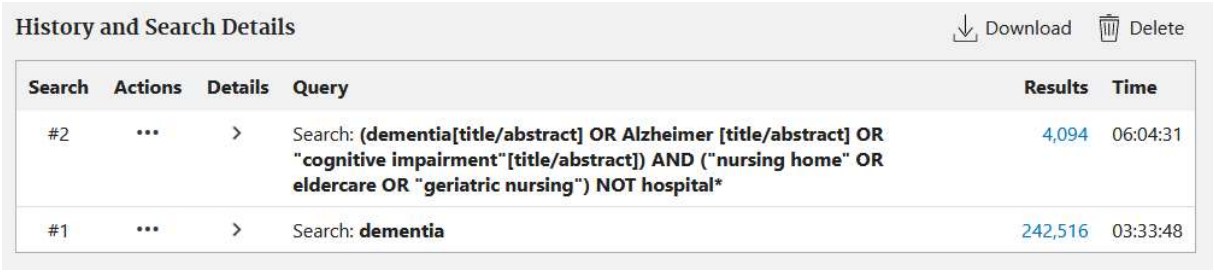


Figure 5: Example of the recording of the research process, here in the PubMed specialist database.

2.1.6 Abstract screening and full text retrieval

An appropriate user account with the relevant specialist databases can also prove to be a very helpful management tool when it comes to screening the results after successfully compiling the final search phrases and deciding which of the search results are to be obtained in full text and analyzed and evaluated in more depth in the further course. In the vast majority of cases, the method of choice for making a decision in this regard is so-called "abstract screening".

By means of this step of the systematic literature search, all search results of the final search runs are checked to what extent they fulfill the content-related and other inclusion criteria. For this purpose, the summary/abstract of the respective publication - the so-called abstract - which can usually be viewed directly in the specialist database, is studied and evaluated for all search hits according to a pattern or questionnaire to be determined in advance in accordance with the defined inclusion/exclusion criteria. All publications that meet the inclusion criteria or for which no reliable decision can be made on the basis of the abstract alone are subsequently obtained in full text if possible. All other search results that can be excluded with certainty and justification during the screening process are not taken into account any further. Whenever possible, abstract screening should be done by two people in an independent manner so that screening results can be compared afterwards and discussed if necessary if there are large discrepancies. Both exclusions and inclusions from the search results are recorded on the research protocol, along with short reasoning.

In addition to the textual representation[3] of how many hits were initially found by searching various specialist databases and other points of contact, how many hits are duplicates, and how many search results were excluded by abstract screening and the subsequent full-text analysis, it is also advisable, in line with the recommendations of the PRISMA statement (Moher et al., 2011; Ziegler & König, 2011), for example, to draw up a flow chart describing the various phases of the literature research or the phases of exclusion of search results/publications. The following figure shows an example of such a diagram.

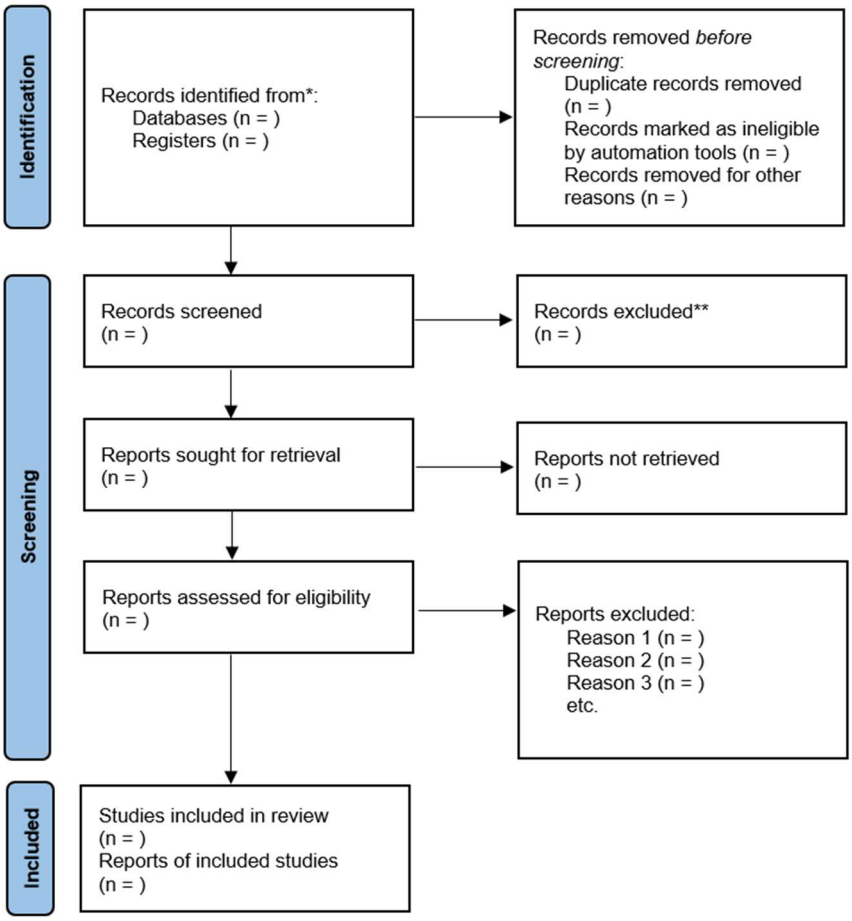


Figure 6: Flowchart illustrating the exclusion of publications in the search process (Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.)

To obtain potentially relevant full texts, document delivery services such as Subito[4] are used in addition to the available (Internet-based) access to libraries, online editions of journals, and various editors/publishers as needed. The administration and management of all full texts as well as all bibliographic data is done with the literature management program EndNote.

2.2 Criteria and tools for evidence assessment, evidence selection, and data extraction

In the (further) development of ENP practice guidelines, the procedure is often analogous to the guideline development procedure of the Association of the Scientific Medical Societies (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, AWMF). An iterative, hierarchical research process is usually chosen for the creation or revision of ENP practice guidelines (Schmucker et al., 2017). This means that preference is given to using evidence that has already been “pre-appraised” analogous to the

[3] E.g.: "Out of a total of <XX> search hits, a total of <YY> publications were rated as relevant for full-text acquisition. Of these <YY> relevant publications, <ZZ> could actually be obtained as full text during the course."

[4] Cf. https://www.subito-doc.de/

so-called 6S hierarchy (Dicenso, Bayley, & Haynes, 2009). The 6S hierarchy of pre-appraised evidence provides for sequential consideration of a) high-quality guidelines, b) aggregated evidence (systematic overviews), and c) primary literature. If the evidence of nursing intervention concepts and/or nursing diagnostic contexts cannot be supported based on high quality guidelines/overviews, the relevant primary and further literature will be consulted. Depending on the type of publication or the underlying research design, appropriate evaluation procedures (critical appraisal) are used to classify the quality. These are presented in more detail in the following chapters.

2.2.1 Critical appraisal of publications in the ENP development process

Different critical appraisal evaluation forms were used during ENP development (see Appendix II) and were in use from 2004 to 2021. These evaluation forms are now replaced by the updated variants and have determined the future further/development processes since 2021. The main reasons for updating the assessment tools used in the ENP development process are the increasingly international orientation of the newly selected approaches as well as numerous new insights in this context over time. There are a considerable number of different quality assessment tools for assessing the quality of a wide variety of research papers or publications (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 a 2010 report on a health technology assessment (HTA), 125 different instruments were already identified for intervention, observational, and diagnostic studies, as well as systematic reviews and meta-analyses, with large variances in terms of their quality (Dreier et al., 2010). The authors also describe different operationalizations of study quality in their systematic literature search. The authors recommend that the methodological quality of a research study be differentiated from the risk of systematic bias both before, during, and after the conduct and that it be evaluated separately. In addition, an explicit separation of reporting and study quality is advisable, as a mix can lead to misjudgments. The assessment of studies largely relates to internal validity, i.e. the validity of the statements made. In this regard, different study types or research designs are susceptible to different errors in terms of risk of bias. For example, to evaluate the efficacy of acupressure interventions in nursing to reduce nausea, structural and observational similarity between the intervention and control groups are very important, whereas the evaluation of systematic overviews depends largely on the quality of the individual studies selected and the inclusion or exclusion criteria chosen.

The type and scope of studies used in the new or further development of ENP practice guidelines depend on the questions being addressed and the research papers available. If, for example, evidence is specifically searched for a causal relationship to a nursing diagnosis, the question is usually formulated very “sharply & narrowly”. The research results for very narrowly formulated questions are usually less extensive; conversely, more abstract questions often lead to a significantly larger number of search results. In the course of the iterative research process for the further or new development of an ENP practice guideline, it may therefore be necessary to conduct an additional literature search with a focus on a specific intervention concept. All publications on which an item in ENP is based have previously been assessed for quality. The complex process of critical appraisal of each publication used is realized with the appropriate assessment tool depending on the type of reference (see the evidence tables in Appendix I). Depending on the particular question being addressed in ENP development, different types of studies or research designs may be considered. For example, randomized controlled trials (RCTs) are highly relevant for nursing intervention concepts, since they have the lowest risk of bias if conducted adequately. Ideally, systematic overviews and/or meta-analyses are also available that use RCTs as primary studies. For the other questions, such as research on characteristics and/or etiology related to a nursing diagnosis or even to experiences of a care receiver, other types of studies, such as the different types of qualitative study designs, will be significant. As already mentioned, a wide variety of critical appraisal tools are available for most types of studies and publications. The following list summarizes the selection of tools used for ENP development before explaining in more detail each approach and, where appropriate, their adaptations for internal use.

2.2.1.1 Systematic reviews (SR) & meta-analyses – rating system according to AMSTAR 2 (Assessment of Multiple SysTemAtic Reviews)

2.2.1.2 Observational studies (case-control, cohort, cross-sectional) – JBI (Joanna Briggs Institute) Critical Appraisal Tools

2.2.1.3 Randomized controlled trials (RCTs) – rating system based on the Cochrane recommendations according to Platz (2021)

2.2.1.4 Qualitative research papers – Critical Appraisal Checklist for Qualitative Research: rating system of the JBI for qualitative research (e.g. for grounded theory, content analysis ...)

2.2.1.5 Guidelines and clinical practice guidelines – rating system based on the AGREE-II Instrument

2.2.1.6 Mixed methods research papers – rating system (excerpts) of the MMAT (Mixed Methods Appraisal Tool)

2.2.1.7 Reference books/textbooks – own implementation of the evaluation criteria of the library of Cornell University in Ithaca, New York

2.2.1.8 Gray literature and websites – evaluation criteria of the AACODS checklist (Authority, Accuracy, Coverage, Objectivity, Date, Significance)

The literature relevant to each ENP practice guideline is systematically assessed according to the specified quality criteria depending on the study design or publication type. The use of checklists supports the evaluation process in the ENP development team. Afterwards, the quality of the evidence is summarized according to a classification scheme (level of evidence). The description of the structure of the evidence tables and their use follows in section 2.2.3. The various evaluation criteria used from the different instruments are first presented in depth.

2.2.1.1 Systematic reviews (SR) & meta-analyses – rating system according to AMSTAR 2

In medicine, systematic reviews are often used as the basis for determining external evidence when selecting appropriate therapies or diagnostic procedures. The methods and procedures used in a systematic review have long been established. This is also true for a wide range of tools to assess the quality of SRs. There are numerous assessment methods for classifying the quality of systematic overviews (Zeng et al., 2015). Schmucker and colleagues (2017) list 20 different instruments in their manual on "Assessing the Risk of Bias in Systematic Overviews" alone. The AMSTAR (Assessment of Multiple SysTemAtic Reviews) for assessing systematic overviews based on randomized controlled trials (RCTs) is classified as a validated and most widely used tool, even outside the scope of medicine (Lühnen, Berger-Höger, & Richter, 2021; Schmucker et al., 2017). Similarly, so-called "reporting statements" for systematic overviews have been developed, which provide various criteria for reporting or publishing a systematic review. A frequently used example of this is, for example, the PRISMA statement, "Preferred Reporting Items for Systematic reviews and Meta-Analyses" (Moher et al., 2011; Thangatorai, Lim, & Nalliah, 2018; Ziegler & König, 2011).

The quality of AMSTAR, developed by a Canadian research group, has been proven in studies. In comparison, the measurement properties of a revised version of the instrument (R-AMSTAR) published in the course have not yet been sufficiently investigated (Pieper et al., 2015). AMSTAR is also available as an online application (Shea et al., 2017). The 11 guiding questions of the original instrument are rated "yes", "no", "unclear" or "not applicable". An overall assessment is not provided. The revised version (R-AMSTAR), on the other hand, provides score points between 1 and 4 for the individual guiding questions with the aim of achieving a better quantification of the overall quality of the publication (Schmucker et al., 2017). In addition to the revised version, another updated version exists, called AMSTAR 2. The focus of this further development was to extend the use of AMSTAR to include overviews that include non-randomized studies and/or meta-analyses (Shea et al., 2017). Since non-randomized studies are often found in nursing and, as a result, overviews also include such studies, it seems reasonable and purposeful to use AMSTAR 2 in this context. Compared to the original version, AMSTAR 2 comprises a total of 16 items, which are to be rated as "yes", "partially fulfilled", "no" or "not applicable", depending on the individual assessment criterion. The checklist with the concrete explanations for the interpretation can be downloaded free of charge from the Internet[5] (Shea et al., 2017). AMSTAR 2 has not yet been comprehensively validated. Previous tests examined interrater reliability using kappa statistics, most of which is found to be of acceptable quality. The usability of AMSTAR 2 was also evaluated, according to which the duration of the application requires between 15 and 32 minutes. This can be justified as 10 questions were retained from the original instrument and the development was based on broad expert consensus and user feedback (Shea et al., 2017). While the first version of AMSTAR is available in a

[5] cf. https://amstar.ca/Amstar_Checklist.php (accessed 02.03.2022)

translated German-language version (Schmucker et al., 2017), no official German version exists for AMSTAR 2 so far.

To assist in the application of AMSTAR 2, explanations and interpretation aids exist for all 16 assessment criteria, which are used in the literature assessment as part of ENP development.

The assignment of score points to quantify and rank the overall quality of a systematic overview or meta-analysis is not provided for in AMSTAR 2, as it was in the original version. Instead, a summary assessment based on the judgment of the person applying AMSTAR 2 is proposed. A grid for such classification is proposed by Shea and colleagues (2017) (see table below), but it is noted that other methods can be used.

Assessment of the overall confidence in the results of the evaluation of a SR/meta-analysis	
High	Zero or one non-critical weakness: The systematic review provides an accurate and comprehensive summary of the results of the available studies that address the question of interest
Moderate	More than one non-critical weakness: The systematic review has more than one weakness, but no critical flaws. It may provide an accurate summary of the results of the available studies that were included in the review.
Low	One critical flaw with or without non-critical weaknesses: The review has a critical flaw and may not provide an accurate and comprehensive summary of the available studies that address the question of interest.
Critically low	More than one critical flaw with or without non-critical weaknesses: The review has more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies.

Table 9: Summary assessment of the overall confidence in the results of a systematic overview/meta-analysis(Source: own representation according to Shea et al., 2017, p. 6)

For the summary assessment of a study or publication, a grading procedure based on GRADE (Grading of Recommendations, Assessment, Development and Evaluation) is used in the context of ENP development, which has many similarities with this scheme for overall assessment (cf. chapter 2.2.3).

2.2.1.2 Observational studies (case-control, cohort, cross-sectional) – Tools of the Joanna Briggs Institute (JBI) for Critical Appraisal

Observational studies have very different objectives. It ranges from the discovery of new correlations to the confirmation of already known findings. The three most common study designs used in the context of different objectives (case-control, cohort, or cross-sectional studies) can be assessed by the widely used STROBE statement (Strengthening the Reporting of Observational Studies in Epidemiology). The development of the STROBE Statement was based on previous experience in developing observational study reporting tools. In particular, the CONSORT Statement (Consolidated Standards of Reporting Trials), which was developed primarily for reporting RCTs, should be mentioned here (Cuschieri, 2019a; Moher et al., 2010). By considering empirical evidence and theoretical considerations, a group of researchers developed the STROBE Statement. The instrument includes a total of 22 items, some of which are specific to the case-control, cohort, and cross-sectional research designs, respectively. With the STROBE Statement, authors are offered a supporting tool for reporting research results in the best possible way. Although the criteria of the STROBE statement can also be used to evaluate and interpret research papers and assess their quality (Cuschieri, 2019b; Vandenbroucke et al., 2014; von Elm et al., 2014), this is contrary to the actual purpose of the application. This misappropriation is widely viewed critically (Haile, 2021). The STROBE Statement in the German version can be used for orientation and further inspiration or in the context of reporting an ENP development work.

Numerous specific instruments have been developed for the critical appraisal of observational studies, many of which also consider specific features of a particular study design. Dreier and colleagues (2010) report a variety of 30 instruments for critical appraisal and reporting of observational studies, a number that has increased since then. Examples include the checklists of the Scottish Intercollegiate Guidelines Network (SIGN), the ROBINS-I tool (Risk of Bias In Non-randomized Studies of Interventions), the Newcastle-Ottawa Scale (NOS), and the AXIS tool (Appraisal tool for Cross-Sectional Studies) (Cochrane Deutschland (Ed.) et al., 2021; Ma et al., 2020). For example, the ROBINS-I tool is used for evidence synthesis within the framework of guideline development by the Association of the Scientific Medical

Societies (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, AWMF) (Cochrane Deutschland (Ed.) et al., 2021; Ma et al., 2020). It is a multilevel, complex tool for assessing the systematic risk of bias in the results of nonrandomized intervention studies. Like many other instruments, ROBINS-I is subject to various limitations (Page, McKenzie, & Higgins, 2018). The tool is limited to quantitative studies examining the benefits or harms of an intervention. These include, for example, cohort studies, case-control studies, or before-and-after studies.

After reviewing various available tools for evaluating observational studies, it was decided to use the checklists of the Adelaide (Australia)-based Joanna Biggs Institute (JBI) in the context of ENP development ((Ma et al., 2020). These are an accepted method and comprehensively cover the required spectrum in the field of observational studies (Munn et al., 2020). In addition, the JBI tools are widely used and established in nursing and medicine, as well as other healthcare professions (Dietrich et al., 2016; Hopper et al., 2019; Porritt, Gomersall, & Lockwood, 2014; Stratil et al., 2021; Williams, Boylan, & Nunan, 2020). Explanations or interpretive aids for each of the JBI's Critical Appraisals are also provided.

Observational studies can basically be divided into analytical and descriptive studies. Descriptive studies evaluate the data collected but do not examine causal relationships. Cross-sectional studies and longitudinal studies are descriptive observational studies. In contrast, analytical studies are dedicated to the search for the cause-effect principle and/or for correlations or specific risk factors (exposition). Analytic methods include case-control studies and cohort studies.

Case-control studies can be set up retrospectively or prospectively. In the study, individuals with a specific disease or the presence of a specific nursing-related phenomenon are compared with non-diseased individuals or those without the presence of the nursing-related phenomenon. Critical appraisal of case-control studies in the context of ENP development is conducted using the JBI's Critical Appraisal Checklist for Case Control Studies (Moola et al., 2020).

In a cohort study, participants are selected according to defined characteristics and the sample is observed over a longer period of time. For example, risk factors of the test persons (e.g. obesity, diabetes) and whether or how certain events (e.g. pressure ulcers) occur in comparison to unexposed test persons are recorded. For this type of research design, the following critical appraisal is available from JBI: “Checklist for Cohort Studies” (Moola et al., 2020).

In cross-sectional studies, clinical or nursing findings or parameters are collected and described once at a defined point in time within a sample in the sense of a “snapshot”. The JBI assessment tool for evaluating studies with this research design is the “Checklist for Analytical Cross Sectional Studies” (Moola et al., 2020).

For the evaluation of publications based on a quasi-experimental research design (all experimental studies in which the groups were formed non-randomly and thus without randomization), a corresponding checklist of the JBI is also used, the “Checklist for Quasi-Experimental Studies”.

All JBI Critical Appraisal Tools, including interpretation aids, are available and can be accessed via a separate website: https://jbi.global/critical-appraisal-tools (accessed: 13.07.2022).

2.2.1.3 Randomized controlled trials (RCTs) – rating system based on the Cochrane recommendations according to Platz (2021)

A variety of different tools for both reporting and critical appraisal are also available for randomized controlled trials (RCTs) (Zeng et al., 2015). A widely used example of RCTs reporting is the CONSORT Statement (Consolidated Standards of Reporting Trials) (Cuschieri, 2019a; Moher et al., 2010). The instrument, which comprises a total of 25 items, is a further development of the previous SORT statement (Standards of Reporting Trials), which was frequently criticized for a number of weaknesses. Because RCTs are often considered the gold standard for evaluating intervention concepts in medicine and sometimes in nursing, and because biased results can occur even with this research design if there are methodological weaknesses, good reporting of study results is significant. In the context of ENP development, the CONSORT statement can therefore be used as a complementary tool to identify good reporting. In addition, it is recommended to study the explanations and interpretation of the questions and items of the CONSORT statement (cf. Knippschild et al., 2015; Moher et al., 2010).

Besides checklists for reporting, there are also numerous instruments for the critical appraisal of randomized controlled trials. The Cochrane Manual for assessing the systematic risk of bias of randomized trials, for example, is widely used in the preparation of guidelines. It is now available in a revised second version from 2021 (Cochrane Deutschland (Ed.) et al., 2021; Schmucker et al., 2016). In principle, the instrument can be used for individually randomized studies with parallel groups as well as for cluster-randomized studies with parallel groups or individually randomized cross-over studies (Cochrane Deutschland (Ed.) et al., 2021). Based on the guidance of the Cochrane recommendation on risk of bias, Platz (2021) derived a Critical Appraisal for RCTs. Against the background of the alignment of the assessment criteria with the Cochrane recommendations and the already proven use in the context of guideline developments of the Association of the Scientific Medical Societies (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften e. V., AWMF), this instrument is used for the assessment of RCTs in the context of ENP development.

Platz proposes to rate the study overall in terms of validity after evaluating the 14 questions, very good validity (++), good validity (+), poor validity (-), low validity (--) (Platz, 2021). The overall evaluation of a used publication is also implemented in a similar form in the context of the ENP new/further development consistently for all research designs and publication types, but based on a different foundation in the form of the GRADE system (cf. chapter 2.2.3).

2.2.1.4 Qualitative research papers – rating system based on JBI Critical Appraisal Checklist for Qualitative Research

Qualitative research designs are becoming increasingly important in evidence-based nursing research. This is due to the necessity of describing and understanding complex facts, experiences, perspectives and lifeworlds of care receivers, their relatives and professionals in the context of various health and disease-related phenomena. Through qualitative research, exploratory insights can be gained into the underlying questions of “how” or “why” (Fringer & Schrems, 2018; Williams, Boylan, & Nunan, 2019). Depending on the research question and objective, a wide range of different research designs is available in qualitative research, for example phenomenological, ethnographic or the grounded theory method (Behrens & Langer, 2016). In the context of ENP development, it is essential to examine and evaluate qualitative studies with regard to their quality and suitability in the context of various practice guidelines. In this context, core criteria or quality criteria of qualitative research designs must be considered in order to assess the credibility of a study and its research results (Moorley & Cathala, 2019; Noble & Smith, 2015; Williams et al., 2020). Numerous instruments or checklists have been developed for the critical evaluation of studies, for which there is no international consensus on the best suitability with regard to the evaluation of qualitative studies (Hannes, 2011; Munthe-Kaas et al., 2019). An internationally established critical appraisal tool is the “Critical Appraisal Checklist for Qualitative Research” developed by staff members of the Joanna Briggs Institute, which was approved by the Institute's Scientific Committee following a peer review process (Lockwood, Munn, & Porritt, 2015; Williams et al., 2020). This checklist can be used to assess the totality of qualitative studies and does not take into account possible specifics of different qualitative research approaches. The critical appraisal tool including interpretation aids is available and accessible via the JBI's own website: https://jbi.global/critical-appraisal-tools (Stand 13.07.2022).

The need for the development of method-specific and scientifically studied instruments for the assessment of qualitative studies is part of the current discourse of evidence-based health research (Williams et al., 2020). For this reason, the critical assessment of qualitative studies in the course of ENP development is currently not method-specific according to the respective qualitative study design, but progress on this will be continuously monitored in order to be able to make possible additions to the assessment tools.

2.2.1.5 Guidelines and clinical practice guidelines – rating system based on the AGREE-II Instrument

The methodological quality of the guideline development process is often not transparent, and the statements and recommendations it contains cannot be clearly interpreted. Against this background, a number of different tools for both critical appraisal and reporting have unsurprisingly also been developed for the evaluation of guidelines and clinical practice guidelines. An example of a guideline reporting tool is the RIGHT checklist (Reporting Items for practice Guidelines in HealThcare), which was developed by a multidisciplinary, international team with policy makers as part of a four-step methodological approach to improve the quality of guideline publication (Chen et al., 2017).

An established example that is used in the critical appraisal of guidelines is the AGREE tool (Appraisal of Guidelines for Research and Evaluation), an internationally used and validated method (Burgers et al.,

2004; Chiappini et al., 2017; MacDermid et al., 2005; Werner et al., 2016). AGREE was published in a further developed version (AGREE II) in 2009 (Browman et al., 2017). Among other things, this revised version replaces the former yes/no response options for the individual evaluation criteria with a seven-point rating scale ranging from *"does not agree at all"* (1) to *"completely agree"* (7). In addition, a calculation basis was developed to compare the evaluated guidelines with each other in terms of their quality. All changes compared to the first version of AGREE as well as interpretation aids and explanations for the six domains and 23 items of AGREE II are described in the user manual (Browman et al., 2017). For German-speaking countries, there is both a translated original version (The AGREE Collaboration (Ed.), 2001), and an adapted and supplemented version of the original AGREE instrument for guideline assessment, called DELBI (Deutsches Instrument zur methodischen Leitlinien-Bewertung) (Kopp et al., 2008; Thole et al., 2007). An additional seventh domain was added to the DELBI when it was initially published in 2005, and an additional eighth domain was added later (2008). Compared with the yes/no assessment of the first AGREE, the DELBI offers an assessment of the individual items on a scale from 1 ("does not apply at all") to 4 ("fully applies"). The two added domains focus on the one hand on the applicability of the guideline to be assessed in the German health care system (domain 7) and on the other hand on the methodological accuracy of guideline development when using existing guidelines (domain 8); the items of the other domains essentially correspond to the AGREE items. The experience of the Association of Scientific Medical Societies (Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften, AWMF), the Medical Center for Quality in Medicine (Ärztliches Zentrum für Qualität in der Medizin, ÄZQ) and their partners, the developers of AGREE and the international network for guideline development were involved in the development of DELBI (Kopp et al., 2008).

In addition, the more recent AGREE II is also available in a systematically translated German version, which corresponds to the English-language original in terms of content and application (The AGREE Next Steps Consortium (Ed.), 2013). Last but not least, a new version of DELBI seems to have been initiated, called DELBI 2.0, which in turn is based on a form of AGREE II adapted for German-speaking countries. Apparently, however, DELBI 2.0 has not managed to leave the status of a consultation version (Dippmann et al., 2013) and has not been officially published to date[6]. Unfortunately, a search for a reason for this or an outline of the connections remained unsuccessful. This also seems unfortunate and confusing because the current version (as of March 2022) shows a validity of the first DELBI only until mid-2015:

"The German Instrument for Methodological Guideline Evaluation (DELBI) was published on June 7, 2005 (version 2005/2006), supplemented on September 22, 2008 (version 2005/2006 + domain 8), and is valid until the next revision or until mid-2013 at the latest. A full update process has been initiated accordingly. After review by the editors, the validity was initially extended until mid-2015" (Kopp et al., 2008, S. 4).

Against the background described above and in particular given the uncertain status of a current version of the DELBI instrument, a slightly adapted form of the German version of AGREE II is therefore used for the critical appraisal of guidelines and/or clinical practice guidelines in the context of ENP development.

As with many of the assessment instruments presented so far, there are also interpretation aids and explanatory texts for the individual 23 items for AGREE II, which are presented in detail in the instructions for use (The AGREE Next Steps Consortium (Ed.), 2013).

According to the manual, all items of AGREE II are rated on a seven-point Likert scale (1 does not apply at all to 7 fully applies). The assessment procedure also specifies that a so-called domain score is calculated for each of the six domains, which are then considered independently of each other and are not to be combined into an overall quality score for the guideline being assessed. For this purpose, the individual values for each item of a domain are added and multiplied by the number of raters and then expressed relatively (as a percentage) as a share of the maximum possible score of the respective domain (again multiplied by the number of raters). This approach is intended to allow comparison of guidelines and to support decision-making regarding the recommendation of a guideline. A cut-off for a "good" or "bad" domain value was deliberately not chosen by the AGREE consortium; this decision should be made by the rater in the respective context of the guideline (The AGREE Next Steps Consortium (Ed.), 2013). The overall evaluation of the respective guideline follows the individual evaluation of the six domains or 23 criteria with two additional, concluding items. On the one hand, the overall quality of the guideline is assessed on a seven-point scale (1: lowest possible quality to 7: highest possible quality) and, on the other hand, a

[6] Even under the official DELBI website, only the 2008 version with the addition of the eighth domain can be found; all references to DELBI 2.0 can no longer be found (cf. https://www.leitlinien.de/hintergrund/leitliniengrundlagen#delbi – last access on 08.03.2022)

recommendation is made by the rater with regard to the application of the guideline (yes - yes with the following changes - no).

This evaluation scheme should also remain possible in the context of the ENP development, which is why the detailed evaluation score of an item should also always be noted. Since a methodology based on GRADE is used within the framework of the ENP further development for the summary evidence assessment of a publication (cf. chapter 2.2.3) and, in addition, the interrater reliability is calculated in the form of the percentage agreement rate when a publication is assessed by several raters, the two summary items for the overall assessment from AGREE II are not included. In addition, for the internal use of AGREE II, with a view to determining interrater reliability, the seven response levels in the evaluation of the individual criteria are additionally combined into a total of three: 1-3 (tending not to apply), 4 (neutral), and 5-7 (tending to apply).

2.2.1.6 Mixed methods research papers – rating system (excerpts) of the MMAT (Mixed Methods Appraisal Tool)

The term mixed methods is usually understood to mean the combination of quantitative as well as qualitative research methods in a combined study design. In this respect, mixed methods research goes much further than other approaches that combine the two research strands, such as triangulation or the combination of methods (Fringer, 2018; Kelle, 2019). The essential feature of mixed methods is the systematic and targeted integration and synthesis of qualitative as well as quantitative elements in both the step of data collection and that of data analysis as well as the reporting of the results of the same research project.

While the combination of quantitative and qualitative research methods within one and the same empirical study has been common practice in disciplines such as social or educational research for a long time, mixed methods have played an increasingly important role in the health sciences and also in nursing science in particular only since the recent past (Kelle, 2019; Knappertsbusch, Langfeldt, & Kelle, 2021; Niederberger & Peter, 2018). This approach seems to be of high importance especially for nursing (research), which operates in a complex system of interdisciplinary and cross-sectoral health care, since the care of people in need of nursing care follows a broad spectrum of nursing diagnostic perspectives as well as intervention concepts. It is not uncommon for nursing care contexts to exhibit a high degree of complexity, and various nursing phenomena often influence one another. Mono-methodological research approaches, which try to address this complexity exclusively with quantitative or qualitative methods, often do not sufficiently meet the requirements (Quasdorf & Holle, 2018). The controversy that has been observed for many years between representatives of the quantitative and qualitative methodological tradition, often with mutual agreement on the validity and scientificity of the respective procedure, makes the collection of data on the complex nursing care process even more difficult (Kelle, 2019; Quasdorf & Holle, 2018). According to Niederberger & Peter (2018), mixed-methods studies represent a promising approach, as their insight potential is both in the recording of causal relationships and in the intersubjective understanding of diverse (also new) phenomena, and they also allow a subject-oriented and evidence-based approach to (nursing) practice.

The increasing importance and prevalence of the mixed-methods approach in health and nursing research is accompanied by an increasing need for appropriate tools to evaluate related research. The search for appropriate critical appraisal tools for mix methods studies revealed a number of different approaches. One example is the “Quality Assessment with Diverse Studies” (QuADS), (Harrison et al., 2021), which is quite broad in terms of the articles to be assessed and contains criteria for both qualitative and quantitative work. For the assessment of studies that do not follow a mono-method approach only, but choose a combination of methods or the mixed-methods approach, it is recommended to apply all criteria of the instrument. However, since the QuADS does not contain explicit evaluation criteria for the central feature of mixed-methods research, more precisely the systematic integration and synthesis of quantitative as well as qualitative elements in several or all relevant phases of the research process, it has not been shortlisted for use in the context of further development of ENP.

In contrast, the “Mixed Methods Impact Evaluation Tool” (MMIE) and the “Mixed Methods Systematic Reviews Appraisal Tool” (MMSR) are very specific within the field of mixed methods research (Jimenez et

al., 2018a, 2018b). These tools specifically focus on mixed methods work to evaluate the impact/effects of interventions or systematic reviews that take a mixed methods approach to produce an evidence synthesis for a particular question. While both instruments take into account the special integrative factor of quantitative as well as qualitative aspects of mixed-methods research through specific evaluation criteria, the quality of the MMIE or MMSR has not yet been sufficiently examined. This circumstance, as well as the overly detailed focus on specific mixed-methods research designs for ENP further development, were essential to the decision not to pursue these two assessment tools either at this time.

Finally, the "Mixed Methods Appraisal Tool" (MMAT) was chosen for the critical appraisal of mixed methods studies in the context of the further development of ENP practice guidelines (Hong et al., 2018; Pluye et al., 2009; Taylor & Hignett, 2014). Originally developed in 2006, the current revision of the MMAT was released in 2018. It assesses both the quantitative and qualitative aspects of a mixed-methods study using five different categories, each with five evaluation criteria. These relate to the quality of the following research designs: qualitative research, RCTs, nonrandomized quantitative-oriented studies, quantitative-descriptive studies, and explicit mixed-methods studies. These five categories are preceded by two general evaluation criteria for all study designs. All evaluation criteria are rated "Yes" / "No" / "Not assessable", and in addition for each evaluation criterion the possibility for comments is given. It is not advisable to calculate an overall score from the evaluation of individual criteria. Instead, it is recommended that the assessments made of the individual criteria be described in more detail. The MMAT is primarily used for empirical studies, but not for theoretical research or reviews. The instrument should be used independently by at least two people.

The MMAT is applied in a two-step process. Prior to the respective evaluation of the mentioned categories, the appropriate methodological study design elements for the given mixed-methods research work are first selected by means of a flowchart.

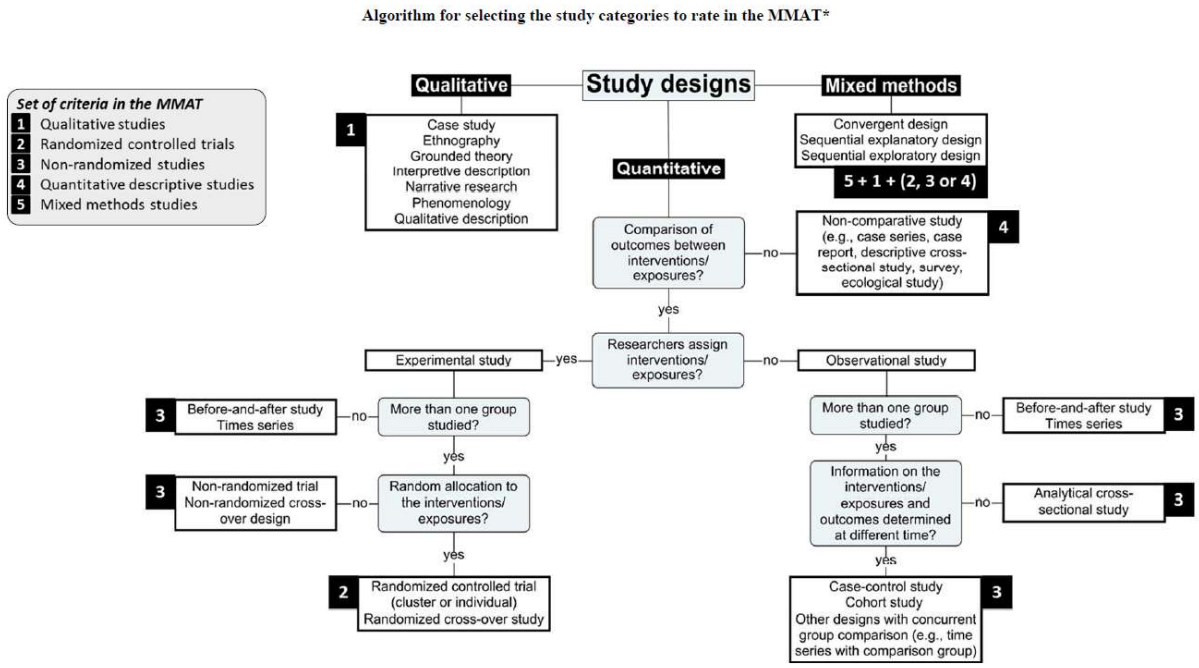


Figure 7: Flowchart for the selection of methodological study design elements for the critical appraisal of a mixed methods research paper according to the MMAT (Quelle: Hong et al., 2018, S. 8)

If, for example, a quantitative strand in the form of a randomized-controlled (sub)study is pursued alongside a qualitative approach, categories 1, 2, and 5 of the MMAT would have to be used as evaluation criteria in order to evaluate both the quantitative and the qualitative elements individually, on the one hand, and to critically reflect on these special features of the mixed-methods approach, on the other. In addition, explanations or interpretation aids are provided for all categories and evaluation criteria, which explain the meaning and intention of the respective criterion in more detail. In addition to the English-language original, the MMAT is also available in French, Brazilian, Turkish and Chinese; an official German

translation does not yet exist. With regard to the quality of the MMAT, the aspects of usefulness, reliability, efficiency, content validity, and interrater reliability have been tested in individual studies so far[7].

For the critical appraisal of mixed-methods research in the context of the further development of ENP, the MMAT is only used in excerpts; the original English-language version of the entire instrument can be found in the associated publication (Hong et al., 2018). More specifically, the fifth category of the instrument, which explicitly addresses the specifics of this integrative research design, is used for evaluation. The reason for this is can be seen in the fact that for the assessment of the other categories of the MMAT (qualitative research, RCTs, non-randomized quantitatively oriented studies, quantitative-descriptive studies), in some cases much more differentiated and better tested instruments are already available and have been selected, which have been presented in detail in the previous chapters. The combination of the fifth category of the MMAT with the more specific assessment grids for quantitative or qualitative research designs gives hope for an even more meaningful assessment of a mixed-methods research paper.

2.2.1.7 Reference books/textbooks - own implementation of the evaluation criteria of the library of Cornell University in Ithaca, New York

In the context of ENP development, reference and textbooks from nursing and neighboring professions are also used when necessary, especially when the research situation on a particular topic proves to be still largely or completely unfounded. When subject-specific book publications or book chapters are the publication form of studies or research papers, or when a literature-based review of a topic has been prepared, the evaluation criteria of the respective research design already described are used. In all other cases, the need for a specific evaluation of the publication type reference book/textbook or chapters results from these publications itself. However, the search for appropriate assessment tools from the context of nursing or health care for this specific purpose revealed few results in the form of two different approaches. This is the “Checklist for Text and Opinion” from the Joanna Briggs Institute's Critical Appraisal series (Joanna Briggs Institute (Ed.), 2017). The checklist includes a total of six criteria, rated as “Yes”, “No”, “Unclear” or “Not Applicable”. Finally, an overall assessment is made on the inclusion or exclusion of the respective book publication or the need to obtain further information before making a decision. Although the “Checklist for Text and Opinion” is subject to a controlled development process and has been repeatedly evaluated and, by definition, is applicable to text- and opinion-based evidence derived from expert opinion, consensus, commentary, assumptions, and assertions published in journals, magazines, monographs, and reports (Aromataris & Munn, 2020; McArthur et al., 2015), the decision to evaluate reference books and textbooks was made against this instrument. The main reason for this is to be seen in the very different nature, in the vast majority of cases, of content conveyed in reference books and textbooks, which do not attempt to convey opinions, consensus, or commentary, but rather a body of knowledge (often non-scientifically) edited for practice and education on one or more topics. Accordingly, the items of the “Checklist for Text and Opinion” do not consistently prove suitable for the evaluation of reference books and textbooks in the classical sense[8].

With this in mind, an instrument was chosen that was constructed from a general set of criteria for evaluating information sources at the Cornell University Library in Ithaca, New York[9] and was developed specifically according to the occasion for evaluating books/book chapters (Datta, Funnell, & Ramuscak, 2012). Unfortunately, according to the state of research, this instrument has no systematic development history and no quality tests. However, since reference books/textbooks or chapters thereof that cannot be evaluated with one of the other, specific instruments from the previous chapters are to be placed on a very low evidence level anyway independent of the evaluation result, the targeting of the evaluation criteria for the context of refence book/textbook was rated as more significant than the scientific development status of the instrument. This is also the reason why the implementation of the evaluation criteria presented in Datta and colleagues (2012) was adapted to the needs of ENP development and reduced in scope. For

[7] An overview of the associated publications is published online at http://mixedmethodsappraisaltoolpublic.pbworks.com/w/page/127425458/Publications%20on%20the%20MMAT (accessed 23.02.2022).

[8] As an example of such inaccuracy of fit, item 3 is to be mentioned: “Are the interests of the relevant population the central focus of the opinion?”

[9] Cf. https://guides.library.cornell.edu/critically_analyzing/home (accessed 23.02.2022)

internal use, the items have also been translated from the English original into German to the best of our ability. The result is a total of ten evaluation criteria to be rated “yes”, “partly” or “no” resulting in a summary overall rating (strong, moderate, weak).

2.2.1.8 Gray literature and websites – evaluation criteria of the AACODS checklist (Authority, Accuracy, Coverage, Objectivity, Date, Significance)

In the context of ENP development, information from websites and gray literature, which cannot be evaluated by means of one of the instruments presented so far, plays only a minor or supporting role. In this context, “gray literature” refers to all publications published in print and/or electronic form that have not been published by a commercial publishing house. Typically, this includes academic theses/qualification papers, congress reports, company publications and the like, but also generic websites (Paez, 2017; Woods, Phillips, & Dudash, 2020). If no other, more specific instrument can be used to assess the information provided, the so-called “AACODS checklist” for the assessment of gray literature will be used in the following to address all remaining publication types and publications that cannot be assessed with the instruments presented so far. The AACODS checklist was published in 2010 at Flinders University (Adelaide, Australia) (Tyndall, 2008, 2010). The AACODS consists of six broad categories, the abbreviations of which also result in the name of the instrument in the form of an acronym:

- Authority
- Accuracy
- Coverage
- Objectivity
- Date
- Significance

The AACODS is now regularly used to assess gray literature, also in the context of producing knowledge syntheses in the health context (Vaska et al., 2019).

2.2.2 Evidence synthesis and evaluation of individual items/intervention concepts

Evidence synthesis on individual intervention concepts of a nursing diagnosis is currently identified in ENP during the literature review but is not reported in the publications. The intervention concepts of a practice guideline represent the professionally meaningful options for nursing action on a nursing diagnosis; a prioritization of the intervention concepts and/or disclosure of the evidence base found for the individual intervention concepts is not currently published. However, the synthesized level of evidence on individual intervention concepts is kept in the internal ENP development documentation. The background is that in most cases the research situation in nursing is still too limited to identify the (external) evidence base for the intervention concepts in the context of nursing diagnoses. Numerous nursing intervention concepts are still at the “best practice” level. On the other hand, reporting the strength of evidence on the intervention concepts would create the risk that, if necessary, the decision-making process of the nurse in the consideration process between their internal and the external evidence would be influenced unilaterally. This bears the risk that the choice of nursing interventions appropriate to the individual care situation of a person in need of nursing care may deviate from the “most evident” choice for a variety of reasons. Intervention concepts not confirmed in the studies or professional publications are consistently not included in ENP or linked to an ENP nursing diagnosis, even if corresponding interventions are (still) used in nursing practice or a request for inclusion is submitted to ENP development from nursing practice. Evidence synthesis and evaluation of individual items/intervention concepts will ensure that they are “best practice” or evidence-based intervention concepts/items that have demonstrated appropriate validity (at least a low level of evidence). Also, in the evidence synthesis of characteristics, etiologies to describe a nursing diagnostic statement, the strength of evidence is not reported on the individual items. ENP development focuses on the most complete illustration possible of correlations confirmed in the literature and/or the confirmed occurrence of corresponding items. This serves the goal of achieving the necessary “completeness” for illustrating a nursing context, also in order to be able to use this data basis for nursing research in the further course. A major challenge in this context is finding the appropriate level of granularity (level of detail or abstraction), also to avoid overlapping content.

In summary: generally, no nursing interventions and/or items in the characteristics/risk factors/etiologies section will be included for which there is evidence of negative effect size and/or significance in the context of the respective ENP practice guideline. Also, no items are included which could not be proven with corresponding reliable references.

2.2.3 Evidence tables for study evaluation in the context of ENP (further) development

So-called “evidence tables” are used to document the results of the critical appraisal process in ENP development. In this context, the evidence tables are an important method for illustrating the quality of the publications used for ENP development. It is significant for both development team members and interested external individuals to be able to view concise evidence summaries, which also helps to promote transparency in the ENP development process. The new evidence table has been developed by the ENP development team following the guideline development process of the German Society of Neurology (Deutsche Gesellschaft für Neurologie, DGN) and the German Society for Neurorehabilitation (Deutsche Gesellschaft für Neurorehabilitation, DGNR) (Platz, 2021). As already explained in 2.2.1, depending on the study design, the different evaluation criteria for a publication are assessed and the results are documented in abbreviated form in the evidence table. The overall consideration of the assessed questions is also documented in the evidence table, as are key findings and information about the study itself. The basic structure of the evidence table is presented below.

2.2.3.1 Structure of the evidence table

The header indicates the study design of the respective publication and, consequently, which instrument was used for the critical appraisal. The name of the study is inserted using the literature management program EndNote and also the name of the person who performed the evaluation of the study.

Appendix 1: Evidence table AMSTAR-2, systematic review	Name of the study: Amarsheda, S. & Bhise, A.R. (2021)	Name of the reviewer: Pia Wieteck

Figure 8: Header of the evidence table

In the context of ENP development on a particular topic, practice guideline or nursing diagnosis, each publication used is assigned a work number referencing the original publication. This number is entered in column 1 in the evidence table so that references to ENP change documentation can be quickly established. The following is an example from the ENP practice guideline revision on “fatigue”.

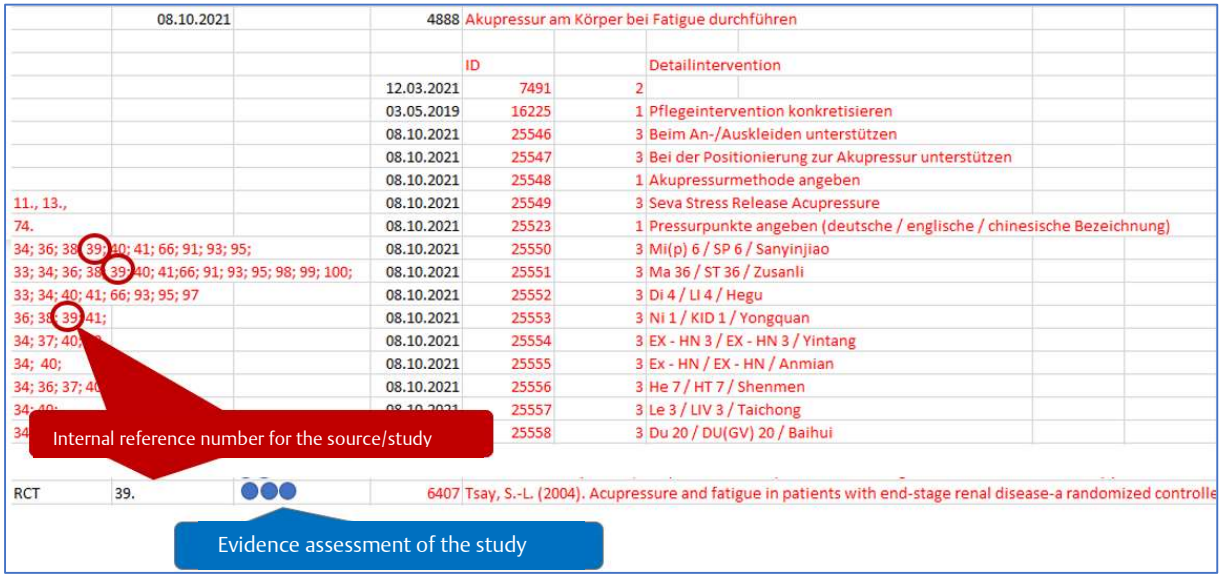


Figure 9: Sample excerpt from ENP change documentation for the development of an intervention for acupressure on the body for fatigue

Column 2 contains the central bibliographic information on the respective publication in the form of author(s) and year of publication as well as the overall assessment of the level of evidence related to the evaluated study (cf. chapter 2.2.3.2).

Column 3 should contain concise information on the main characteristics of the study, e.g. research design, in the case of reviews the number of included individual studies or information on group or participant numbers.

Column 4 is intended for concise information about the research path to the evaluated publication, such as the search period and in which specialist database (or by which other means) the publication was found. Column 4 is not about describing the search algorithm in detail, but about indicating the qualitative assessment.

Column 5 should include a bulleted description of the population(s) in the study and their key characteristics, such as age, gender distribution, or disease duration or severity.

Column 6 is for information that addresses the research focus of the paper and describes the focus in a few words, e.g., regarding the research question(s).

Column 7 indicates how results/outcomes were recorded and measured within the research papers (e.g., by means of assessments, measurements, video recordings with content analyses, etc.) and, if applicable, how often and at what intervals this was done (follow-up).

Column 8 describes the main results of the study. Depending on the type of study design and subject, the effect sizes of a randomized controlled trial related to the intervention in the intervention group and control group should be reported here, for example, in the case of a cross-sectional study, statistical information on the occurrence of a specific risk factor, etc.

Column 9 lists the cumulative responses and ratings for each critical appraisal's evaluation criteria. Depending on which assessment tool was used in advance, this can be numbers, characters or yes/no statements.

Finally, in column 10, the key conclusions are formulated in such a way that the recommendation for inclusion or modification in ENP can be derived.

Evidence table: Name of the study: _____ Name of the reviewer: _____

Ref. nr.	Author, year, level of evidence	Study type, number of studies, number of participants	Search date, searched databases, search algorithm	Population (e.g. age, gender distribution, disease duration, severity)	☐ Intervention and control measures ☐ Nursing diagnosis ☐ Nursing diagnostic phenomena	Outcome measurement Follow-up period (Observation period of a study)	Key findings Effect sizes Key features of the nursing diagnostic concept Causal relationships, risk factors... Risk of bias	Validity assessment Checklist: Click or tap here to enter text. (F x-F y)	Conclusion (based on PICO; results, effect estimate, benefit-harm ratio and acceptability, subgroup analyses; methodological weaknesses / risk of bias, imprecision, heterogeneity; relevance of the results for clinical practice, relevant main points and conclusions)
	Author(s). year GRADE - level of evidence ○ ○ ○ ○	• .. •			• ..	*How and with what results was the outcome measured, e.g. assessment values or measured values, etc.?* •..	*Central results with details of the effect sizes* •..	F1: yes F2: no F3: no F4: yes F5: no F6: no F7: no F8: no F9: yes F10: no F14: yes	*Conclusions from the main results and the evidence assessment of the study - recommendation for the inclusion of items in ENP* • ..
1	2	3	4	5	6	7	8	9	10

Figure 10: Structure of the evidence table in the context of ENP development

It is important to consider the significance and impact of individual questions/criteria on the reliability of the nursing-relevant statements when transferring the individual results from the evaluation systems into an overall assessment of the level of evidence and reliability of the respective publication. If, for example, critical areas of the questions that have a high probability of influencing the validity of the statements have been classified as “not fulfilled”, this has a more serious significance than if a question has been assessed as “not fulfilled”, which in the overall view has only a limited significance for the validity of the study statements/results.

2.2.3.2 Introductory reflections on grading procedures in the context of evidence assessment

Over the past decades, a number of different evidence ranking systems have emerged that have also repeatedly faced various criticisms, depending on the area of application. In the context of nursing

research in particular, it is well known that randomized controlled trials, which are often regarded as the evident gold standard in medicine, are not necessarily and automatically to be regarded as being of higher quality in terms of evidence content than, for example, observational studies (Behrens & Langer, 2004; Evans, 2003; Grypdonck, 2004; Lavin et al., 2002). For some nursing-related questions, the "best" evidence is not verified via RCTs or RCTs are not feasible for ethical reasons (Roberts & DiCenso, 1999). Increasingly, it is recognized that both observational studies and case series can contribute well to evidence generation for interventions, and qualitative and mixed-methods research also provide valuable evidence for the development of, for example, nursing diagnoses. Which study design can provide the best answer to concrete questions in the context of ENP development always depends on the context of the respective question (Petticrew & Roberts, 2003). Especially in the case of nursing diagnostic phenomena, such as suffering in the context of grief or loneliness, qualitative studies from phenomenology or observational studies are suitable to derive characteristics from the experience world of the affected persons for the description of a nursing diagnostic concept (Panke-Kochinke, 2012; Ploeg, 1999). In summary, observational studies, cross-sectional studies, qualitative studies, conceptual analyses, and the like are more suitable for clarifying nursing diagnostic phenomena, whereas randomized controlled trials (RCTs) should be preferred for research on appropriate nursing treatment concepts. If the focus is on questions about causal relationships or effects of, for example, environmental influences on our health, cohort studies or case-control studies may also be well suited. Depending on the research question, systematic reviews and meta-analyses summarize multiple research papers to answer a research question. Results of meta-analyses and systematic overviews usually have the highest level of evidence if the primary studies included are well selected and of good quality. The following overview presents the most essential study designs and their central features without claiming to be exhaustive.

Study designs	
Perspective/features	Description, special features
Time perspective	Retrospective: for example, retrospective evaluation of data from patient records, hospital information systems, nursing care process documentation, etc. Prospective: data collection continuously from the start of the study over a certain period of time, defined method and subsequent evaluation
Observation period	Cross-sectional study: "snapshot", e.g. one-time measurement or survey Longitudinal study: observation over time or multiple measurements with time intervals
Control group	Without control group (e.g. case series, observational studies, surveys, document analyses) With control group (e.g. RCTs)
Triangulation	Various methods are used in the research work to bring together different perspectives. These can be qualitative and quantitative study designs.
Analytical studies are dedicated to the search for cause-and-effect principles, such as correlations between certain risk factors for a nursing phenomenon.	Cohort study: observes participants according to defined characteristics in a selected sample over a longer period of time. Case-control study: compares care receivers with and without a specific nursing phenomenon with regard to, for example, contributing factors, risk factors, characteristics, etc.
Descriptive studies are purely describing studies; data are presented and analyzed without attempting to explain causes or causal relationships.	Observational study: surveys a nursing-relevant event/finding/occurrence at a specific point in time; no proof of causality. The observational study is a type of epidemiological study design and can be designed in the form of a cohort, case-control or cross-sectional study, for example (Cuschieri, 2019b). Epidemiological study: investigates e.g. risk factors and their distribution in the population, usually observational study Individual case studies; case reports; case series Correlation studies: Suitable for forming hypotheses. Data is used to check whether, for example, there is a connection with certain observed factors and, for example, a nursing diagnosis. Cross-sectional study e.g. prevalence/incidence studies: For example, the occurrence of adverse events in a facility or the occurrence of nursing phenomena in a population
Experimental studies include a range of research designs in which researchers manipulate one or more variables and control and measure	Randomized controlled intervention study (RCT): additional random allocation of subjects to the groups; (double) blinded randomized controlled: additional blinding of subjects (and practitioners) Quasi-experimental study: Have an experimental design, but lack essential characteristics such as random distribution of the test subjects or the control group. For example, there is no random selection process for the study participants; at least

any change in other variables. The extent to which a certain factor (variable) influences a situation/behavior or condition is examined.	one independent variable is actively investigated by the experimental set-up of the study design. Crossover trial intervention study/cross-over study: the intervention and control groups are switched in the middle of the study Intervention study non-randomized controlled: subjects are assigned to different interventions controlled: study determines intervention, control group available Before-and-after study: The participants in a study are examined before and after an intervention, for example. Multicenter study: these are clinical studies that are conducted simultaneously in several institutions with a multicenter study design involving several research teams.
Qualitative study designs	Qualitative content analysis: based on the analysis of written data, e.g. interviews, documentaries, video material with spoken word, etc. Interpretative studies, e.g. communication analyses, narrative analyses, documentation analyses Grounded theory: qualitative research method with the special feature that data collection and data analysis alternate. The researcher identifies, refines and integrates categories from the data material in order to derive and/or confirm a theory. Delphi method: this is a procedure to obtain forecasts from expert opinions and/or to clarify consensus and dissent between expert opinions. A highly structured group communication process is used to develop a solution, a consensus for complex problems/questions from the individual contributions. Phenomenological study design: this involves research designs that describe specific care-related phenomena such as grief, loneliness and hopelessness (Mehrholz, 2010). Phenomenological studies are particularly important in the context of nursing diagnostics. These studies also belong to the descriptive approaches. Concept analysis: very helpful, for example, to describe and summarize nursing phenomena in nursing practice more precisely
Summary of individual studies	Review/overview: evaluation of the relevant current literature with the aim of summarizing similar studies with the same research question. In a “systematic” review according to a set standard, studies with the same endpoint and at a high level, e.g. RCTs, are ideally evaluated against each other in order to be able to make an evidence statement. Meta-analysis: Systematic review that summarizes the results of individual studies using statistical methods. The included primary studies should be as similar as possible in terms of their study population, and all studies should ideally have the same primary endpoint. Multicenter studies: Clinical studies that are conducted simultaneously at several institutions and whose scientific significance is increased by the participation of different investigators.

Table 10: Overview of the most important study designs and their key features(Source: own expanded presentation based on Rojahn, 2016, S. 17). Publications used for expansion: (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).

Different study entail harbor different risks of systematic error or systematic bias (risk of bias). Since there can be very different sources of error in the generation of knowledge, it is not possible to establish a one-dimensional ranking of study types with regard to the level of evidence[10]. A systematic risk of bias is, for example, the distortion of study results due to an inappropriate selection method of study participants (selection bias). RCTs attempt to prevent this potential bias by randomizing participants and blinding both the participants and the people conducting the study. If there were only this one error, the ranking (evidence hierarchy) could be solved one-dimensionally. Among the many other possibilities of systematic error that can lead to invalid results in a study are, for example, the inappropriate selection of measurement instruments or a lack of a theoretical framework for the research work. It is therefore not surprising that even in medicine the classic “hierarchy of evidence” is disputed (Petticrew & Roberts, 2003). Nevertheless, the evidence hierarchy is a frequently used tool to derive and assess the quality and reliability of clinical decision-making based on the study designs. The following shows the frequently used evidence hierarchy, which can be found in numerous publications, particularly medical publications, with some adaptations.

[10] Level of evidence (degree of evidence/evidence level): scientific proof that indicates the degree of generalizability or validity of research results according to a grid of criteria (Bartholomeyczik et al., 2008).

Classical evidence hierarchy in medicine	
1	Systematic review and meta-analysis
2	Randomized controlled trials with definitive results
3	Randomized controlled trials with inconclusive results
4	Cohort studies
5	Case-control studies
6	Cross-sectional studies
7	Individual case studies

Table 11: Classical evidence hierarchy in medicine (source: own presentation based on Petticrew & Roberts, 2003)

When comparing the different evidence hierarchies, similarities can be recognized, and it can also be seen that qualitative research designs are not represented in the vast majority of cases. In the evidence hierarchy of the Scottish Intercollegiate Guidelines Network (SIGN), a qualitative implementation level of the study is already included.

Evidence hierarchy according to the SIGN grading system	
1++	High-quality meta-analyses, systematic reviews of RCTs or RCTs with a very low risk of bias
1+	Well-conducted meta-analyses, systematic reviews or RCTs with low risk of bias
1-	Meta-analyses, systematic reviews or RCTs with a high risk of bias
2++	High-quality systematic reviews of case-control or cohort studies High-quality case-control or cohort studies with a very low risk of confounding or bias and a high probability that the association is causal
2+	Well-conducted case-control or cohort studies with a low risk of confounding or bias and a moderate probability that the association is causal
2-	Case-control or cohort studies with a high risk of confounding or bias and a significant risk that the association is not causal
3	Non-analytical studies, e.g. case reports, case series
4	Expert opinion

Table 12: Evidence hierarchy according to the SIGN grading system (source: own presentation based on Harbour & Miller, 2001)

The Oxford Centre for Evidence-Based Medicine suggests that the evidence hierarchy should not be based on the overall study design, but on the underlying research question (Durieux, Vandenput, & Pasleau, 2013; Oxford Centre for Evidence-Based Medicine, 2009). In this context, there is now also a growing consensus that evidence statements in the context of intervention research do not have to be fundamentally and exclusively based on RCTs in order to achieve a higher level of evidence (Behrens & Langer, 2004; Gastinger et al., 2002; Petticrew & Roberts, 2003). There is even scientific evidence that RCTs do not always occupy the top rung of the methodological ladder of evidence and that observational studies occupy the lower rungs (Petticrew & Roberts, 2003). In addition, there is a growing consensus that quantitative studies only make sense and can only generate real benefits for the respective target groups if they are embedded in qualitative research and findings (Grypdonck, 2004). There have already been several attempts to transfer nursing research into its own evidence hierarchy. For example, Lavin et al. (2002) have developed a nursing-specific classification and the need to consider the underlying research question(s) based on the differences between evidence-based nursing and evidence-based medicine. Another approach to defining a hierarchy of evidence in the context of nursing research comes from Evans (2003). He looks at three dimensions: the effectiveness, appropriateness and feasibility of interventions in the healthcare system. Here, for example, observational studies are also suggested as “good” in the same way as RCTs.

Level of evidence	Effectiveness	Appropriateness	Feasibility
Excellent	Systematic overviews Multicenter studies	Systematic overviews Multicenter studies	Systematic overviews Multicenter studies
Good	RCT Observational studies	RCT Observational studies Interpretative studies*	RCT Observational studies Interpretative studies*
Moderate	Non-controlled study designs with pronounced results	Descriptive studies Focus groups	Descriptive studies Intervention studies Before-and-after studies

	Before-and-after studies Non-randomized controlled studies		Focus groups
Low/poor	Descriptive studies Case series Expert opinion Studies of poor methodological quality	Expert opinion Case series Studies of poor methodological quality	Expert opinion Case series Studies of poor methodological quality

Table 13: Evidence hierarchy according to Evans for evaluations of interventions in the healthcare system (source: own presentation based on Evans, 2003)

* Interpretative studies: documentation analysis, narrative analysis, conversation analysis, content analysis, hermeneutic methods, phenomenology, grounded theory

When dealing with the topic, it quickly becomes clear that, in addition to the considerations outlined so far, the quality of implementation of a published study is also of great importance for assessing its quality (Behrens & Langer, 2016). For example, an RCT or a systematic review cannot be classified as high quality as such. The following table provides a rough overview of common sources of systematic error in studies that can influence and reduce the assessment of the evidence.

Frequent systematic sources of error when conducting studies	
Source of error	Explanation
Number of cases too small	If the samples are too small, it may not be possible to detect existing differences.
Imprecise question or lack of theoretical derivation	Inaccurate questions are asked or there is a lack of theoretical embedding in a framework. Solution: theoretical framework, possibly PIKE question scheme
Missing measured values	Problem with meta-analyses, for example, in which estimated values may be used to make further calculations possible; measurements are not possible due to the inhomogeneous study situation.
No appropriate selection of measuring instruments	Intervention studies in particular are about determining the outcome/change. If inaccurate or unsuitable instruments are used to measure the effects, it is possible that something different may be measured than originally intended. Solution: if necessary, qualitative research in order to develop suitable measuring instruments
Errors, confounder (confounder)	Parameter or risk factor that is associated with both the exposure/intervention and the target variable (e.g. disease). E.g.: age, gender, nicotine consumption, additional illnesses Solution: in experimental study design e.g. randomization, in data analysis e.g. stratification
Fluctuations in the course of the disease	Improvement or worsening in the course of the disease, independent of the intervention, influenced by unrecorded variables/influences e.g. spontaneous improvement of colds, progression of diabetes, episodes of multiple sclerosis or rheumatism Solution: control group
Selection bias	Systematic differences in the selection of subjects or the composition of the groups, e.g. men vs. women, patients with mild vs. severe symptoms Solution: randomization, concealed assignment
Performance bias	Systematic differences between the groups in terms of e.g. care, attention, etc. Solution: blinding of subjects and practitioners
Recall bias	Recall bias, e.g. in retrospective studies: subjects whose memory of a possible exposure is inaccurate are more likely to report it when they are ill than when they are not. Other examples include unfavorable, incomplete data in document analyses.
Observer ascertainment bias	Systematic differences in the evaluation because the group assignment is not (sufficiently) blinded.

Detection bias	Systematic differences between the groups in the determination of the outcome, e.g: hormone treatment in women leads to more visits to the doctor (e.g. due to bleeding), as a result of which more uterine carcinomas are found - although hormones do not promote carcinomas.
Deviations from the curriculum (attrition bias, drop out, loss to follow up)	Systematic differences between the groups in the event of deviations from the study protocol, e.g. if more subjects in the intervention group drop out of the study due to side effects than in the placebo group. Solution: intention-to-treat analysis (ITT)
Publication/reporting bias	Systematic differences between published and unpublished results because, for example, significant, desirable and spectacular results are published more frequently, faster and at a higher level than others. Solution: use of the relevant guidelines for publication and transparency of the research process

Table 14: Frequent systematic sources of error when conducting studies (source: own presentation based on Grypdonck, 2004; Rojahn, 2016)

In the discussions so far, it has become clear that the classification of a study or its publication in an evidence hierarchy is a complex construct and should not be reduced to the study design alone. In addition, the question of appropriateness of the study design for the credibility of the evidence must be critically scrutinized, especially in light of the numerous conceivable sources of error. Particularly in the context of nursing research, the methodological question of the suitability of the research design for answering the research question(s) is always important (Mayring, 2017; Petticrew & Roberts, 2003). As a result, several authors recommend that the evaluation and classification of the evidence of a research paper should be carried out against the background of precisely the respective research question(s) (Baker et al., 2010; Bartholomeyczik et al., 2008; Petticrew & Roberts, 2003). Instead of classifying evidence into rigid, hierarchical levels of evidence, the use of an evidence matrix that takes these circumstances into account should therefore be preferred, depending on the research question(s) and methodological pluralism (Petticrew & Roberts, 2003). An illustrative example can be taken from the discourse in psychiatric care. A considerable amount of knowledge and evidence is derived here from individual case studies (Stickley & Phillips, 2008). The transferability of rigid evidence hierarchies to psychiatric practice must be critically questioned against the background of individuality versus generalizability and could lead to the implementation of intervention concepts based primarily on the presumably highest levels of evidence, but often equally effective or even better concepts are not considered in the individual care context. These arguments are another reason why the levels of evidence for individual nursing intervention concepts are not explicitly specified in the ENP development process.

2.2.3.3 Evidence assessment in the context of ENP development based on the GRADE scheme

After the critical appraisal with the aid of the respective assessment tools, the respective publication or research work is classified overall with regard to the level of evidence. The aim is to obtain an overall view of the quality of the references used as part of the new or further development of an ENP practice guideline and thus to assess their validity. The level of evidence of a reference reflects the degree of confidence in the correctness and reliability of (a) statement(s) regarding nursing diagnostic correlations or estimates of an effect in nursing intervention concepts. The cumulative result of an overall assessment is entered in the ENP evidence table. The aim of ENP development is to use references with a high level of evidence wherever possible in order to achieve a high degree of reliability of the nursing diagnostic statements and their associated elements (characteristics, etiologies and resources) as well as the associated intervention concepts. The overall evaluation and assessment of the quality of the study or publication is already an established practice in guideline development (Platz, 2021). Many critical appraisal tools also contain items or methods for the overall assessment of a publication. Some of these approaches have already been listed in the assessment tools presented (see chapter 2.2.1). However, the criteria or questions for the overall assessment are operationalized differently, which makes it difficult to gain an overall overview of the quality of the publications used for ENP development.

Against this background, a uniform system for summarizing the evidence assessment of a publication is being sought and implemented as part of the ENP development. The following example from the further development of an ENP practice guideline on fatigue, in which a total of over 50 relevant studies were

reviewed and evaluated, is intended to illustrate the benefits of this. The combined overall assessment shown in the table allows the quality of the research situation to be assessed quickly and comprehensively.

Study type	Evidence assessment based on GRADE	Source
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

Table 15: Exemplary excerpt from the total overview of publications used in the context of ENP further development work

The GRADE (Grading of Recommendations, Assessment, Development and Evaluation) scheme is used with regard to the system and methodology for summarizing the evidence assessment for a publication (Kunz, Burnand, Schünemann, 2008; Schünemann et al., 2013). The assessment is based on a four-level scoring system.

Orientation to GRADE

In the GRADE quality assessment, RCTs are considered more credible and more robust to risk of bias than other research methods, so this design initially has a higher level of evidence than, for example, observational studies. However, the initial classification of a particular study design can be upgraded or downgraded by various factors (see the results for the two RCTs in Table 15 as an example). The GRADE quality assessment is originally aimed at an endpoint in the development of guidelines. Due to this fact, some adjustments were made for the use of GRADE in the context of ENP development. On the one hand, no specific endpoint is assessed, but rather the quality of the present publication. A similar approach has already been successfully used by other experts (Panfil et al., 2011). GRADE divides the quality of evidence into four levels or grades. The following figure shows the unchanged GRADE evidence levels from the official instrument manual.

Table 5.1: Quality of Evidence Grades	
Grade	Definition
High	We are very confident that the true effect lies close to that of the estimate of the effect.
Moderate	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect.
Very Low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Figure 11: Evidence level according to GRADE (Schünemann et al., 2013)

The wording of the four evidence levels of the GRADE evidence levels was translated and adapted for the nursing-specific focus in the context of ENP development, as can be seen in the following table. The respective evidence level for a reference is expressed with a four-level symbolism.

Evidence level	Symbol	Definition
High quality	++++ or ⊕⊕⊕⊕	It is very unlikely that further research will change the confidence in the observed nursing diagnostic relationship and/or observed treatment effect. It is very likely that the true effect lies close to that of the estimate of the effect.
Moderate quality	+++ or ⊕⊕⊕◯	Further research will likely have an impact on the confidence in the observed nursing diagnostic relationship and/or observed treatment effect. There is moderate confidence that the true effect is likely to be close to the estimate of the effect. But there is a possibility that it is substantially different.
Low quality	++ or ⊕⊕◯◯	Further research will most likely lead to changes in the observed nursing diagnostic relationship and/or the observed treatment effect. The confidence in the effect estimate is limited and the true effect may be substantially different.
Very low quality	+ or ⊕◯◯◯	The observed nursing diagnostic relationship and/or treatment effect is subject to high uncertainty. There is very little confidence in the effect estimate.

Figure 12: Translated and adapted evidence levels for use in ENP development based on GRADE (Quellen: Kunz et al., 2008; Platz, 2021; Schünemann et al., 2013)

Like other systems, GRADE also defines a hierarchy of evidence in relation to the given study design, which serves as the starting point for the subsequent classification decisions in the assessment process. The GRADE development group defined criteria that can lead to an upgrading or downgrading of the evidence level. Here, too, selective adjustments were made for the context of nursing research. Specifically, the qualitative research designs were added to the initial assessment of the evidence level depending on the respective research question and the criteria for upgrading or downgrading the studies were added by the special features of qualitative research.

Defined starting point of the research work depending on the study design and the research question(s)

In the following table, the initial classification for the summarized evidence assessment in the context of the ENP development is defined and can be viewed. This list supplements the original GRADE scheme in some respects. At the level of the nursing interventions, the evidence hierarchy and the GRADE grading criteria were adopted. However, additions were made for the development of nursing diagnoses, as the ENP practice guideline development is not only concerned with the interdependencies of nursing interventions, but also with development. The additions are based on the discussion about the ranking of evidence in the specific context of nursing (Behrens & Langer, 2004; Evans, 2003) and nursing diagnostics and nursing pathways (Grieser, Hegedüs, & Kunz, 2011). The initial evaluation of a study varies depending on the respective research design and the examination of the effectiveness of nursing interventions or the confirmation of a nursing diagnosis with characteristics/symptoms or the question of influencing factors and causal relationships. As already explained, it makes sense to define the evidence hierarchies in the context of the respective research question(s) for several reasons. For the ENP development, the initial classification of the levels of evidence of GRADE were therefore adapted to the research context of ENP.

Evidence category	Reliability of the study design in relation to the evidence of efficacy of nursing interventions	Reliability of the study design in relation to the nursing diagnostic phenomenon as well as characteristics	Reliability of the study design in relation to influencing factors/etiologies of a nursing diagnosis
High level of trustworthiness	○ Systematic overviews or meta-analyses with RCTs ○ Multicenter studies ○ Guidelines	○ Systematic overviews, meta-analyses with qualitative and/or quantitative non-randomized research studies ○ Guidelines ○ Multicenter studies	○ Systematic overviews, meta-analyses with qualitative and/or quantitative non-randomized research studies ○ Multicenter studies
Moderate trustworthiness	○ RCTs ○ Cohort studies	○ Method triangulation ○ Observational studies ○ Qualitative research work ○ Interpretative studies*	○ RCTs ○ Cohort studies ○ Case-control studies
Low level of trustworthiness	○ Observational studies ○ Quasi-experimental studies without randomization (non-randomized controlled trial) ○ Overviews/meta-analyses of case-control or cohort studies ○ Before-and-after studies ○ Controlled clinical studies, e.g. intervention studies (cross-over studies), comparative studies	○ Non-analytical studies: case series, case reports ○ Non-experimental descriptive studies ○ Delphi method	○ Cross-sectional studies ○ Prevalence/incidence studies ○ Correlation study
Very low level of trustworthiness	○ Non-experimental descriptive studies ○ Non-analytical studies: case series, case reports ○ Qualitative research work ○ Expert opinion from recognized authorities (reference book level, e.g. recognized nursing textbook literature) ○ Delphi method	○ Case series, case reports ○ Expert opinion from recognized authorities (reference book level, e.g. recognized nursing textbook literature)	○ Case series, case reports ○ Expert opinion from recognized authorities (reference book level, e.g. recognized nursing textbook literature) ○ Delphi method

Table 16: Initial classification of the summarizing evidence level in the context of ENP development based on GRADE

* Interpretative studies: documentation analysis, narrative analysis, conversation analysis, content analysis, hermeneutic methods, phenomenology, grounded theory

Aspects that reduce or increase evidence in the ENP research context

The various factors (adapted and based on GRADE) that can lead to a reduction or increase in the summarized evaluation of the evidence level are presented below. Any reduction or increase in the level of evidence does not reflect rigid categories, but a continuum within each category. If the evidence is only moderate for a particular factor, the decision on whether a study/reference falls above or below the threshold for decreasing or increasing quality (by one or more factors) depends on the rater's judgment. This also involves assessing the significance of the respective factor for the overall result (Schünemann et al., 2013).

Factors that can reduce the level of evidence	
Factor	Consequence
Limitations in study design or implementation (high risk of bias), e.g. • Lack of allocation concealment • Insufficient blinding • Insufficient follow-up • Selective reporting of results • Measuring instrument inappropriate for collecting measured values • Unclear "inclusion and exclusion criteria" of study participants and/or studies • High absenteeism of participants • Distortion due to confounding • Approach not appropriate to the subject and question • Unclear/inaccurate intervention description (Hoffmann et al., 2016) • Differences in the test conditions • Small number of study participants, wide confidence interval (CI)	↓ 1 or 2 steps
Inconsistency of the results	↓ 1 or 2 steps
Indirectness of the evidence	↓ 1 or 2 steps

Inaccuracy e.g. • Missing information on the origin of data, reasons for exclusions, etc.	↓ 1 or 2 steps
Publication bias, e.g. • Selective choice of studies for publication • Selective reporting of results	↓ 1 or 2 steps
Lack of theoretical framework or unclear scientific-theoretical positioning, lack of disclosure of prior assumptions	↓ 1 or 2 steps
Factors that can increase the quality of the evidence	
Factor	Consequence
• Well described and conclusively derived study design and conduct (low risk of bias) • Particularly transparent, plausible and theory-based justification for subject- and question-based decisions in research design • Plausible selection of participants with large samples that can be expected to achieve saturation • Plausible constancy of the test conditions • Transparent reporting of dropouts AND low proportion of these (intention-to-treat analysis) with specification of statistical effect • Method triangulation combined with homogeneous results • Sound theoretical derivation with suitable research design	↑ 1 level
High effectiveness/high effect size, e.g. • Narrow confidence interval for RCTs	↑ 1 or 2 levels
All plausible confounding factors would reduce the proven effect or increase the effect. Therefore, if no effect was observed, this risk can be excluded	↑ 1 level

Table 9: Factors that can reduce or increase the summarized level of evidence of a study/publication (source: own and adapted presentation based on Schünemann et al., 2013)

2.3 The further course of the ENP development process

As part of the ENP development process, the review proposals are generally agreed by the team. At the consensus meetings, all participants are provided with the literature search protocol, the critical appraisal documents, the evidence tables and the original studies. A random sample of 10% of the references is evaluated by a second person. If there is a high level of agreement of 80% or more, the further study assessments are carried out solely by the person primarily responsible for working on the topic previously defined in the team. The literature search and the evaluation of the studies are reflected upon at weekly meetings throughout the entire development process. At the end of the editing process, each amendment proposal developed for implementation in ENP is discussed in a meeting and jointly agreed. If a consensus cannot be reached immediately on an aspect, the literature is consulted again and/or other experts from nursing science and practice with proven expertise are consulted in order to find further evidence.

References

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

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: 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. *Evid Based Nurs, 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

Appendix I: Evidence table as part of the ENP development from 2022

(Exemplary with assessment grid of RCT studies)

Evidence table: Name of the study: _____ Name of the reviewer: _____

Ref. nr.	Author, year, level of evidence	Study type, number of studies, number of participants	Search date, searched databases, search algorithm	Population (e.g. age, gender distribution, disease duration, severity)	☐ Intervention and control measures ☐ Nursing diagnosis ☐ Nursing diagnostic phenomena	Outcome measurement Follow-up period (Observation period of a study)	Key findings Effect sizes Key features of the nursing diagnostic concept Causal relationships, risk factors... Risk of bias	Validity assessment Checklist: Click or tap here to enter text. (F x-F y)	Conclusion (based on PICO; results, effect estimate, benefit-harm ratio and acceptability, subgroup analyses; methodological weaknesses / risk of bias, imprecision, heterogeneity; relevance of the results for clinical practice, relevant main points and conclusions)
	Author(s), year GRADE - level of evidence ○ ○ ○ ○	• .. •			• ..	*How and with what results was the outcome measured, e.g. assessment values or measured values, etc.?* •..	*Central results with details of the effect sizes* •..	F1: yes F2: no F3: no F4: yes F5: no F6: no F7: no F8: no F9: yes F10: no F11: yes F12: yes F13: yes F14: yes	*Conclusions from the main results and the evidence assessment of the study - recommendation for the inclusion of items in ENP* • ..

Questions for RCTs according Platz 2021		Yes	No	Unclear	Not applicable
1.	Were the inclusion and exclusion criteria clearly defined? (external validity)	☐	☐	☐	☐
2.	Is there a clear definition and adequate assessment of the outcomes of the study (clinical relevance)?	☐	☐	☐	☐
3.	Were side effects and acceptability of the interventions reported?	☐	☐	☐	☐
4.	Did an adequate follow-up study take place (recording of long-term effects)?	☐	☐	☐	☐
5.	Is a clear definition and description of the experimental and control conditions given (study contrast)?	☐	☐	☐	☐
6.	Are the participants randomly assigned (selection bias)?	☐	☐	☐	☐
7.	Was the allocation kept under wraps (“allocation concealment”, selection bias)?	☐	☐	☐	☐
8.	Were the experimental and control groups comparable at the beginning of the study (selection bias)?	☐	☐	☐	☐
9.	Were the patients and staff blinded during the treatment, and were the treatments in the randomized groups comparable (implementation bias) outside of the effects investigated?	☐	☐	☐	☐

10. Were the study results collected in a blinded manner (“detection bias”)?	☐	☐	☐	☐
11. Was the reporting not selective (“reporting bias”)?	☐	☐	☐	☐
12. Were the results data (almost) complete (“attrition bias”)?	☐	☐	☐	☐
13. Was an “intention-to-treat” analysis reported?	☐	☐	☐	☐
14. Do the results sufficiently support the reported conclusions?	☐	☐	☐	☐

Appendix II: Former Critical Appraisal Forms (2004-2021)

Evaluation for RCT

RECOM®
Pflege Competence Centrum

(Source reference, title of the study)

Edited by: Date:

Evaluation form for studies **No.2**

Study type	**Randomized controlled trials (RCTs)** (Description: experimental study in which participants are randomly assigned to an intervention group or a control group and are prospectively observed, double-blind study: the investigators do not know who belongs to which group; "gold standard")
Hierarchy level	Ib
Labeling	A: good research-based evidence

Checklist	**RCT** (based on Lancaster et al.1997 in Schreier/Bartholomeyczik 2004)
General questions (Hamer, Collinson 2000 in Schreier/Bartholomeyczik 2004)	**Answers**
1. are the results of the study valid?	Yes☐ No☐ Criticism:
2. is the study directed at a clearly focused problem, a clearly formulated question?	Yes☐ No☐ Criticism:
3. is the study directed/focused on the conditions: a) how the population was selected b) how the intervention (investigation/experiment) was carried out? c) how the results are considered/evaluated?	a) Yes☐ No☐ Criticism: b) Yes☐ No☐ Criticism: c) Yes☐ No☐ Criticism:
4. was the allocation of patients to treatment randomized (random)?	Yes☐ No☐ Criticism:
5. were all patients included in the study correctly informed about their participation in the study?	Yes☐ No☐ Criticism:
6. have ethical criteria been examined and taken into account?	Yes☐ No☐ Criticism:
7. was the follow-up examination complete?	Yes☐ No☐ Criticism:
Yes☐ No☐ Criticism:	Yes☐ No☐ Deviations in:
Detailed questions	**Answers**
9. were patients, assistants and examination staff unaware (blind) of the treatment method	Yes☐ No☐ Criticism:

(blind study)?	
10. was the study group homogeneous (similar) at baseline in terms of factors that could affect the results (e.g. age, gender and social class)?	Yes☐ No☐ Criticism:
11. apart from the experimental interventions, were the groups treated equally?	Yes☐ No☐ Criticism:
Evaluation of results	
12. how great was the treatment effect?	
13. what results could be measured?	
14 How precise was the assessment of the treatment effect?	
15 How wide were the confidence intervals?	
16. what was the failure rate, how was the failure data analyzed?	

Abstract:

Key findings of the study:

Relevance for the ENP treatment pathway:

Evaluation for qualitative studies

(Source reference, title of the study)

Edited by: Date:

Evaluation form for studies No.6

Study type	**Qualitative study** (aim is to discover unknown relationships and social phenomena in their natural environment; method triangulation is useful)
Hierarchy level	III
Labeling	B: fairly good research-based evidence

Checklist	**Qualitative study** (according to Schreier/Bartholomeyczik 2004)
General questions (according to Hamer, Collinson 2000 in Schreier/Bartholomeyczik 2004, Prakke, Wurtser 1999 in Schreier/Bartholomeyczik 2004 and according to Oxmann et al. 1994, Brown 1999, Sackett et al 2000 in Behrens/Langer 2004)	**Answers**
1. was the underlying qualitative research method clearly aligned with the objectives of the study?	Yes☐ No☐ Criticism: Target: Method:
2. was the research question clearly formulated?	Yes☐ No☐ Criticism:
3. was a literature search carried out?	Yes☐ No☐ Criticism:
4. were the criteria for population selection and the type of sample clearly described?	Yes☐ No☐ Criticism:
5. what method was used to select the study participants?	
6. is the sample selection justified?	Yes☐ No☐ Criticism:
7. are characteristic features of the sample such as gender, cultural origin, social class, etc. described?	Yes☐ No☐ Criticism:
8. have ethical criteria been examined and taken into account?	Yes☐ No☐ Criticism:
9. was the sample appropriate?	Yes☐ No☐ Criticism:
10. was the method of data collection adequately explained? (described in detail)	Yes☐ No☐ Criticism:
11. was the approach to the type of data collection guided by rules?	Yes☐ No☐ Criticism:
12. how was the collected data subsequently analyzed and interpreted?	
13. did the data collection reach saturation?	Yes☐ No☐ Criticism:

(Signs of validity, assessment by number, duration of interviews/observations, time span, method triangulation in data collection and analysis) → Quality criterion	
14. have quality criteria been documented?	Yes☐ No☐ Criticism:
15. are the results presented in a detailed and comprehensible manner?	Yes☐ No☐ Criticism:
16. do the results accurately reflect the phenomena described?	Yes☐ No☐ Criticism:
17. were the results confirmed? (participants' conclusions, matching procedures, evidence of data saturation)	Yes☐ No☐ Criticism:
18. do the results help to better understand the people studied in their environment?	Yes☐ No☐ Criticism:
19 Where and how can the new knowledge be applied in practice?	

Six general quality criteria of qualitative research according to Mayring 1999:	**Assessment**:
1. procedural documentation: Exact documentation down to the last detail so that the research process is comprehensible for everyone a) Explication of the preliminary understanding b) Compilation of the analytical instruments c) Implementation and evaluation of the data collection	completed☐ has not been completed ☐ Criticism: completed☐ has not been completed ☐ Criticism: clean☐ unclean☐ Criticism
2. argumentative interpretation validation: Argumentatively justified interpretation a) adequate prior understanding of the respective interpretation has been established (interpretation is meaningful and theory-based) b) interpretation is coherent, breaks are explained c) alternative interpretations have been sought and checked in order to refute negative interpretations (validity justification)	completed☐ has not been completed ☐ Criticism: conclusive☐ not conclusive☐ Criticism: completed☐ has not been completed☐ Criticism:
3. rule-based: Systematic approach through step-by-step, sequential procedure/ procedural regulation a) Analysis steps defined in advance b) Material divided into meaningful units c) Analysis carried out from one step to the next	completed☐ has not been completed☐ Criticism: conclusive☐ not conclusive☐ Criticism: completed☐ has not been completed☐ Criticism:
4. proximity to the object: Appropriateness to the object as a guiding principle a) linked as closely as possible to the everyday world of the researched subjects b)striving for a match of interests with the	completed☐ has not been completed☐ Criticism: conclusive☐ not conclusive☐ Criticism:

researched subjects c) constantly checking whether proximity to interests is achieved in the research process	completed☐ has not been completed☐ Criticism: conclusive☐ not conclusive☐ Criticism:
5. communicative validation: Validation of the reconstruction of subjective meaning a) Validity of the results/interpretations checked together with the interviewees b) Important arguments on the relevance of the results obtained from the dialog with the interviewees	completed☐ has not been completed☐ Criticism: conclusive☐ not conclusive☐ Criticism:
6. triangulation: Increasing the quality of qualitative research by combining several analytical processes a) qualitative and quantitative methods are used to process the data material	completed☐ has not been completed☐ Criticism:

Abstract:

Key findings of the study:

Relevance for the ENP treatment pathway:

Rating for survey

RECOM Pflege Competence Centrum

Edited by: Date:

Evaluation form for studies **No. 5**

Study type	**Survey** (description: cross-sectional collection of non-experimental data, e.g. through in-depth interviews or observations, to determine the prevalence, distribution and interrelationships of variables)
Hierarchy level	III
Labeling	B: fairly good research-based evidence

Checklist	**Survey** (according to Schreier/Bartholomeyczik 2004)
General questions (according to Hamer, Collinson 2000 in Schreier/Bartholomeyczik 2004)	**Answers**
1. how was the sample determined and selected?	Sample:
2. how were the questions constructed? (open, closed...)	
3. was an already known survey instrument used?	Instrument:
4. what was the response rate?	% absolute
5. was denier bias discussed?	Yes☐ No☐ Criticism:
6. were all results given?	Yes☐ No☐ Criticism:
7. how was the survey conducted/the questionnaire distributed?	
8. were reminders to return the questionnaire included in the data collection process?	Yes☐ No☐ Criticism:
9. have ethical criteria been examined and taken into account?	Yes☐ No☐ Criticism:

Abstract:

Key findings of the study:

Relevance for the ENP treatment pathway:

Evaluation for meta-analyses

RECOM®
Pflege Competence Centrum

(citation, title of the study)

Edited by: Date:

Evaluation form for studies RECOM **No. 1**

Study type	**Meta-analyses** (systematic review) and **Systematic Overview**of randomized controlled trials (RCTs) (Description: Meta-analysis is a method of statistical integration of quantitative research results, results of several studies on the same topic are combined and treated as a closed data complex; one unit of analysis then corresponds to one study)
Hierarchy level	Ia
Labeling	A: good research-based evidence

Checklist (based on existing assessment tools: Oxmann et al. 1994, Brown 1999, Sackett et al. 2000 in Behrens/Langer 2004)	**Systematic Review/Overview** (based on Lancaster et al.1997 in Schreier/Bartholomeyczik 2004)
Questions	**Answers**
1. was the question of the systematic overview/review clearly formulated?	Yes ☐ No ☐ Criticism:
2. has the search for relevant studies been carried out thoroughly?	Yes ☐ No ☐ Criticism: Search locations:
3. were the studies used similar (homogeneity of the patient group studied)	Yes ☐ No ☐ Criticism:
4. are the included criteria suitable? Inclusion criteria: Patient group, intervention, setting, outcome measure, methodological criteria (follow-up)	Yes ☐ No ☐ Criticism:
5. has the validity of the included studies been adequately checked/assessed?	Yes ☐ No ☐ Criticism:
6. has missing information been collected by the investigator?	Yes ☐ No ☐ What is missing:
7. how sensitive/changeable are the results with regard to the way in which they were collected?	Valid ☐ Reliable ☐ Objective ☐
8. how precise are the results (odds ratio)?	
9. is there an overview table with the results of the studies used? (Assessment comprehensible?)	Yes ☐ No ☐ What is missing:
10. is the conclusion of the summary consistent with the work reviewed?	Yes ☐ No ☐ Deviations in:
11. are the recommendations related to the assessment level of the knowledge base?	Yes ☐ No ☐ What is missing:
12. are the assessments of advantages/benefits clear?	Yes ☐ No ☐ Criticism:

13. were analyses of subdivisions of the studies interpreted carefully enough?	Yes ☐ No ☐ Criticism:
14. are the results transferable to our patient group?	Yes ☐ No ☐ Criticism:
15 Is the benefit worth the potential risks and costs (Number Needed-To-Treat?	Yes ☐ No ☐ Criticism:

Abstract:

Key findings of the study:

Relevance for the ENP treatment pathway:

Evaluation for cohort study

(citation, title of the study)

Edited by: Date:

Evaluation form for studies No.4

Study type	(A) **Prospective** cohort **study** ☐ (B) **Individual cohort study** ☐ (Description: Used to determine the prognosis or incidence of a disease, group of people is observed over a period of time)
Hierarchy level	IIa II☐ ☐
Labeling	A: good research-based evidence B: fairly good research-based evidence

Checklist	**Cohort study** (according to Schreier/Bartholomeyczik 2004)
General questions (according to Hamer, Collinson 2000 in Schreier/Bartholomeyczik 2004)	**Answers**
1. was a clear hypothesis put forward?	Yes ☐ No☐ Criticism:
2. were the criteria for participation or non-participation in the cohort clearly formulated?	Yes ☐ No☐ Criticism:
3. how were the cohorts selected/ determined?	Factor:
4. have ethical criteria been examined and taken into account?	Yes ☐ No☐ Criticism:
5. have the diagnostic criteria been clearly presented?	Yes ☐ No☐ Criticism:
6. which measurement methods were used?	Methods:
7. is the validity of the measurement instruments known?	Yes ☐ No☐ Criticism:
8. are the characteristics described in detail?	Yes ☐ No☐ Criticism:

Abstract:

Key findings of the study:

Relevance for the ENP treatment pathway:

Evaluation for case-control studies

(citation, title of the study)

Edited by: Date:

Evaluation form for studies **No. 3**

Study type	(A) **Prospective** case-control **study** ☐ (B) **Retrospective case-control study** ☐ (Description: the characteristic to be investigated already exists in the intervention group, the control group is formed from individuals who are as similar as possible)
Hierarchy level	IIa ☐ IIb ☐
Labeling	A: good research-based evidence B: fairly good research-based evidence

Checklist	**Case-control study** (based on Lancaster et al.1997 in Schreier/Bartholomeyczik 2004)
General questions (from Hamer, Collinson 2000 in Schreier/Bartholomeyczik 2004)	**Answers**
1 Is the aim of the study clearly formulated?	Yes ☐ No ☐ Criticism:
2. were the criteria for participation or non-participation in the population groups of this study clearly formulated?	Yes ☐ No ☐ Criticism:
3. how were the participants in the "cases" selected/determined?	Same factor:
4. have ethical criteria been examined and taken into account?	Yes ☐ No ☐ Criticism:
5. what were the "matching" criteria (criteria for control group formation)? Matching = distribution of known independent variables between the two study groups (gender, age), which can correlate with the dependent variables (Diekmann 1999)	Criteria:
6. how was the control group selected and what criteria were used?	Criteria:
7. how was the measurement of the independent variable expressed?	
8. which analysis was assured?	

Abstract:

Key findings of the study:

Relevance for the ENP treatment pathway:

Appendix III: Protocol for the documentation of a systematic literature search

Protocol for the documentation of a systematic literature search

Occasion and context of the systematic literature review:

Performing person(s):

Search period:

1. The question(s) and their specification

Topic:

Method chosen for narrowing down (e.g. orienting research, application of the PICO scheme, etc.) and key findings from this:

Specific question(s), observe PIKE:

2. Derivation of search terms

The multiplication of search terms increases the prospect of relevant hits! Relevant parameters: synonyms, generic and subordinate terms, English translations and their synonyms, attention to different inflectional forms.

According to this procedure, all core elements/core concepts of the research question(s) should first be considered separately in order to be able to combine the resulting search terms into search phrases in a meaningful way.

The following matrix may be suitable as an auxiliary construct.

	Core concept 1	Core concept 2	Core concept 3	Core concept 4
Generic term				
Subordinate term				
Synonyms				
English translations				

3. Defining inclusion and exclusion criteria

The aim is to clarify the formal framework of the research and focus on what is really of interest and practicable (additional filtering).

Possible criteria here are, for example:

Language of publications

Publication period

Publication / study types

Age groups, gender, etc.

Databases /search locations

Settings, professions

...

Important: All inclusion and exclusion criteria must be justifiable!

Inclusion criteria	Exclusion criteria

4. Definition of specialist databases and other points of contact for research

Depending on the topic to be dealt with, the question of “where” to search for literature must also be answered.

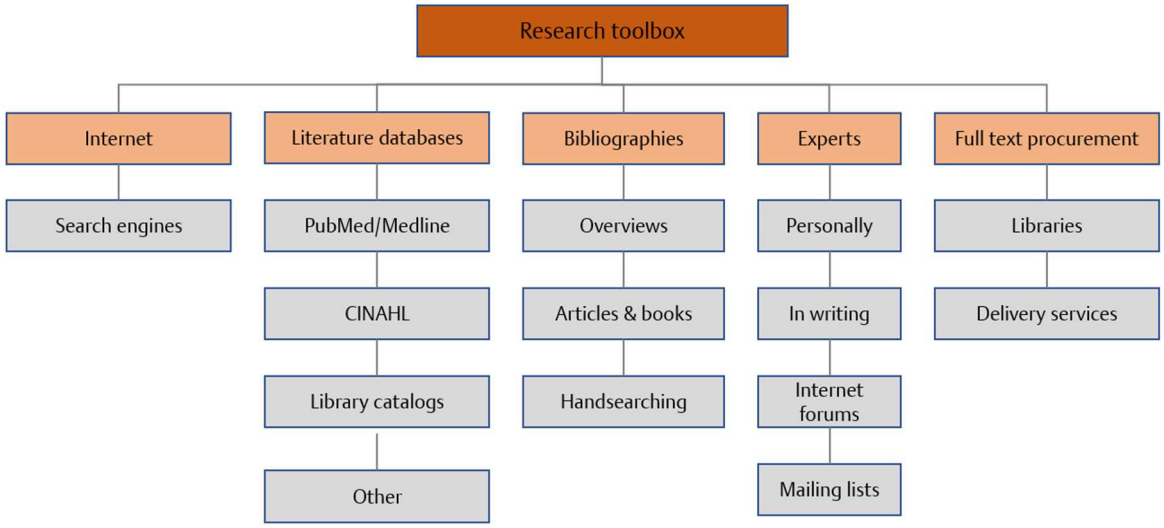


Figure 13: Orientation aid for determining search locations and specialist databases (source: Brandenburg et al. 2013: p.45)

Selection of electronic specialist databases with a brief explanation:

Selection of library catalogs/open access holdings with short reasoning.

If applicable: the selection of guideline catalogs (e.g., AWMF, etc.) with short reasoning:

Selection of further research locations (e.g. journals, gray literature, search engines on the Internet, snowball system, etc.) with a brief explanation:

5. Development of search phrases and alternatives

See 2.1.5 for instructions.

Search phrases to be used and possible alternatives:

Partial search phrase for core concept 1			
Linking operator (AND, OR)	Search term (if necessary truncated, masked or as a phrase)	Field search? (Title, MeSH, ...)	Alternatives?
	Main term		
OR	*Synonym 1*		
OR	*Synonym 2*		
OR	*Synonym 3*		
Terms to be excluded in this core concept?			
NOT			

Partial search phrase for core concept 2			
Linking operator (AND, OR)	Search term (if necessary truncated, masked or as a phrase)	Field search? (Title, MeSH, ...)	Alternatives?
	Main term		
OR	*Synonym 1*		
OR	*Synonym 2*		
OR	*Synonym 3*		
Terms to be excluded in this core concept?			
NOT			
Partial search phrase for core concept 3			
Linking operator (AND, OR)	Search term (if necessary truncated, masked or as a phrase)	Field search? (Title, MeSH, ...)	Alternatives?
	Main term		
OR	*Synonym 1*		
OR	*Synonym 2*		
OR	*Synonym 3*		
Terms to be excluded in this core concept?			
NOT			

Linking the partial search phrases to the individual core concepts of the research question(s) to form an overall search phrase and developing possible alternatives if the search result is too narrow, too broad or thematically inappropriate:

Overall search phrase - variant 1				
	Linking operator (AND, OR, NOT)		Linking operator (AND, OR, NOT)	
Partial search phrase for core concept 1	*AND*	*Partial search phrase for core concept 2*	*AND*	*Partial search phrase for core concept 3*
				Terms to be excluded from the overall search phrase?
			NOT	

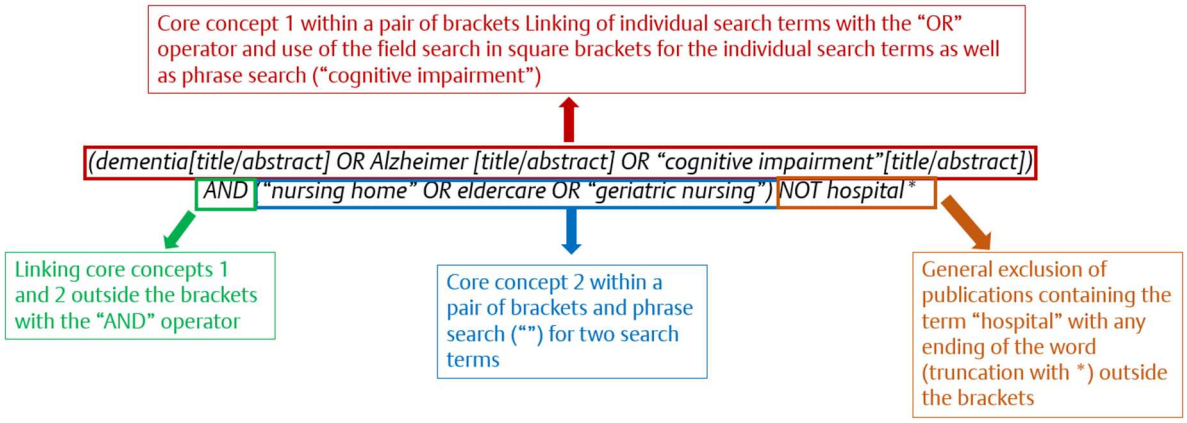
Search filters to be used (inclusion/exclusion criteria)				
Publication period	Language	Study/publication designs	Age group	…

Overall search phrase - variant 2				
	Linking operator (AND, OR, NOT)		Linking operator (AND, OR, NOT)	
Partial search phrase for core concept 1	*AND*	*Partial search phrase for core concept 2*	*AND*	*Partial search phrase for core concept 3*
Terms to be excluded from the overall search phrase?				
			NOT	
Search filters to be used (inclusion/exclusion criteria)				
Publication period	Language	Study/publication designs	Age group	…

A useful optional alternative to the tabular presentation of the search phrase(s) may be the text-only implementation or the text-only draft of the same.

Here is a brief example to illustrate this:

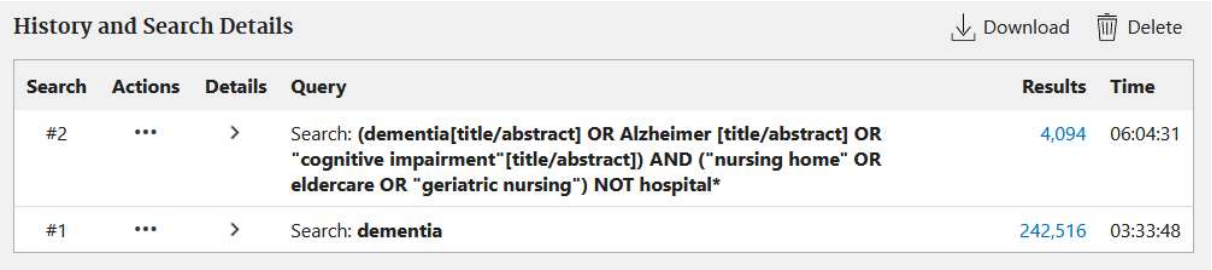
*(dementia[title/abstract] OR Alzheimer [title/abstract] OR "cognitive impairment" [title/abstract]) AND ("nursing home" OR eldercare OR "geriatric nursing") NOT hospital**



Search phrase(s) in text form:

6. Documentation of the results of individual search runs and the search phrase(s) finally used

It is advisable to document the history of all searches carried out in order to make the search process traceable, including the respective changes to the quantity and quality of the search results in various specialist databases (and possibly using other search locations). All major specialist databases and catalogs offer the option of viewing and copying the search history (see screenshot for an example).



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

Search history at <date> in the database <name>:

Search history at <date> in the database <name>:

Search history at <date> in the database <name>:

Final search phrase(s) from <date>, which was used to start the abstract screening:

7. Documentation of the results of abstract screening and full text retrieval:

Abstract Screening of the final search run in the database <name> at <date>:

From a total of <XX> search hits, a total of <YY> publications were rated as relevant for acquisition in full text.
Of these <YY> relevant publications, <ZZ> could actually be obtained as full text during the course.

Abstract Screening of the final search run in the database <name> at <date>:

From a total of <XX> search hits, a total of <YY> publications were rated as relevant for acquisition in full text.
Of these <YY> relevant publications, <ZZ> could actually be obtained as full text during the course.

Abstract Screening of the final search run in the database <name> at <date>:

From a total of <XX> search hits, a total of <YY> publications were rated as relevant for procurement in full text.
Of these <YY> relevant publications, <ZZ> could actually be obtained as full text during the course.