

Bases et méthodes du développement



© TIMDAVIDCOLLECTION/stock.adobe.com

Septembre 2024

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

Traduction française : Christine Rabl

Table des matières

1) Le développement de ENP basé sur des preuves.....	3
2) La recherche systématique de la meilleure évidence possible.....	4
2.1 Stratégie et éléments centraux de la recherche bibliographique systématique, internationale	4
2.1.1 La question de recherche.....	5
2.1.2 Dédution de termes de recherche	6
2.1.3 Critères d'inclusion et d'exclusion	7
2.1.4 Définition des bases de données et d'autres sources de recherche	8
2.1.5 Phrases de recherche et options d'influence sur le déroulement de la recherche	9
2.1.6 Abstract screening et procuration des textes intégraux.....	13
2.2 Critères et outils appliqués pour l'évaluation de l'évidence, la sélection d'évidence et l'extraction de données.....	14
2.2.1 Critical appraisal de publications dans le processus de développement de ENP	15
2.2.2 Synthèse d'évidence et évaluation d'éléments/concepts d'actions individuels	27
2.2.3 Tableau d'évidence pour évaluer des études dans le contexte du développement ENP	27
2.3 La suite du processus de développement de ENP	39
Bibliographie	40
Annexe I : Tableau d'évidence rempli à titre d'illustration issu du développement de ENP depuis 2021 ..	46
Annexe II : Protocole de documentation d'une recherche documentaire systématique	47

Bases et méthodes du développement

1) Le développement de ENP basé sur des preuves

ENP veut dire European Nursing care Pathways et représente un langage professionnel de soins standardisé aux structures monohiérarchiques qui classe et met à disposition l'expertise soignante sous forme de lignes directrices de pratique. La figure ci-après permet de comprendre les démarches systématiques du processus de perfectionnement de ENP qui est optimisé sans cesse. En règle générale, une nouvelle version base de données de ENP est mise à disposition chaque année ; les publications en livre sont généralement réalisées à des intervalles de trois ans.

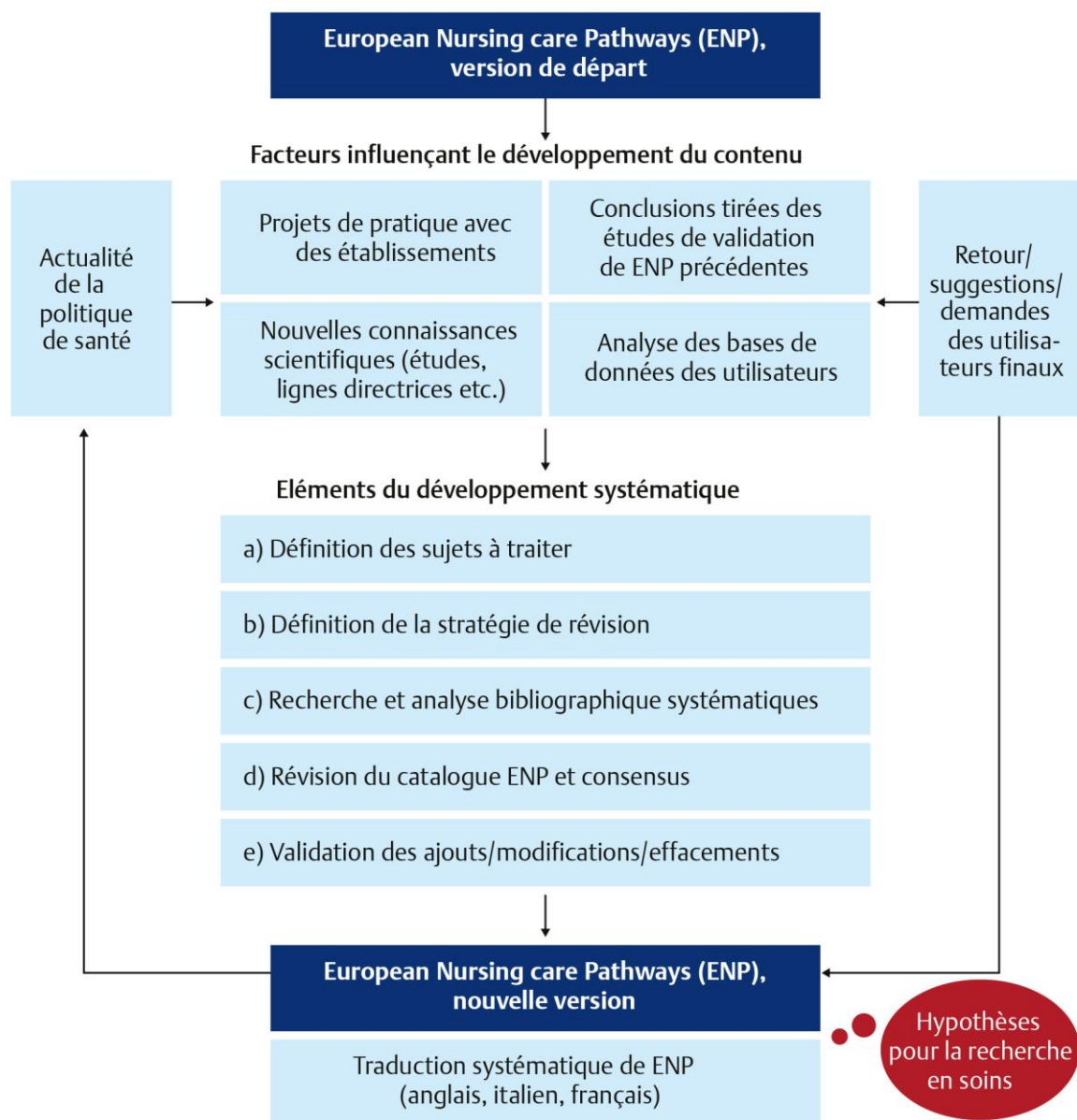


Figure 1 : Représentation schématique du processus du développement systématique de ENP

En fonction des décisions prises en politique sanitaire, des retours utilisateurs et des dernières connaissances des sciences de soins et des disciplines avoisinantes de la santé, chaque année un choix de lignes directrices de pratique ENP est défini qui sera soumis à une vérification systématique et, si besoin, à une révision. La qualité de développement des lignes directrices de pratique basées sur le diagnostic soignant est cruciale pour les utilisateurs de la classification des soins car elle assure une sécurité élevée pour un travail factuel dans le processus de prise de décision. D'autant plus importante est la description la plus transparente possible des méthodes et démarches scientifiques appliquées pour perfectionner les lignes directrices ENP existantes ou pour en développer de nouvelles.

2) La recherche systématique de la meilleure évidence possible

En règle générale, le développement ou le perfectionnement systématique d'une ligne directrice de pratique ENP suit un processus de plusieurs étapes. Les questions suivantes y jouent un rôle essentiel :

- Est-ce que le phénomène observé représente un concept diagnostique soignant ?
- Est-ce que la littérature spécialisée contient des caractéristiques/facteurs de risque/étiologies/ressources basés sur des preuves qui confirment le diagnostic soignant ou qui présentent un lien étiologique/influençant avec le diagnostic soignant ?
- Quels sont les objectifs soignants ciblés ?
- Est-ce que la littérature spécialisée confirme les concepts d'intervention visant à pallier et/ou soulager une problématique soignante d'un diagnostic soignant ?
- Quelle est la dépense en ressources, temporelles et autres, applicable à une nouvelle intervention de soins et quelles mises en correspondance (mapping) avec d'autres systèmes et outils sont requises ?
- ...

Un élément central pour répondre à ces questions est la planification et réalisation d'une recherche et évaluation systématique de la littérature internationale. En fonction de la question ciblée lors du développement et/ou de la mise à jour de la ligne directrice de pratique ENP, une stratégie de recherche est élaborée à l'aide de laquelle les publications pertinentes peuvent être identifiées sans laisser des « taches aveugles ». En même temps, une stratégie de recherche bien pensée permettra de réagir de manière adéquate aux contingences qui peuvent surgir au cours de la recherche bibliographique, par ex. en termes de qualité (trop de résultats de recherche aux contenus divergents) ou de quantité (clairement trop peu de résultats de recherche suite à une stratégie de recherche trop étroite ou bien une abondance de résultats non gérable suite à une stratégie trop générale). Les chapitres suivants détailleront le développement d'une stratégie de recherche, la réalisation elle-même de la recherche bibliographique et notamment l'évaluation individuelle de l'évidence des publications identifiées au sens d'un « critical appraisal » en fonction du type de publication ou de la conception de la recherche.

2.1 Stratégie et éléments centraux de la recherche bibliographique systématique, internationale

Le perfectionnement de ENP respecte scrupuleusement les démarches classiques d'une recherche bibliographique systématique internationale (cf. par ex. Mayer, Raphaelis, & Kobleder, 2021 ; Nordhausen & Hirt, 2020). Les étapes engagées sont les suivantes :

1. Délimitation du sujet
2. Formulation de recherches et hypothèses de recherche le plus précises possibles
3. Dédution de termes de recherche appropriés à base des questions de recherche et variation des termes (synonymes, traduction anglaise etc.)
4. Définition des critères d'inclusion et d'exclusion (par ex. in-/exclusion de conceptions d'étude déterminées, délimitation de la période de publication et des langues de publication etc.)
5. Définition des bases de données et autres sources à utiliser pour la recherche

6. Documentation de la stratégie et du déroulement de la recherche
7. Réalisation de la recherche bibliographique
8. Tri et évaluation des résultats de recherche (abstract screening)
9. Procuration et gestion de la littérature pertinente
10. Lecture critique des textes intégraux (critical appraisal, cf. chapitre 2.2)

Toutes les démarches et décisions sont documentées et elles sont concertées au sein de l'équipe de développement à des moments essentiels du processus de recherche. A cette fin, un protocole a été établi pour la documentation d'une recherche bibliographique systématique (cf. annexe II) qui doit être appliqué pour toute réalisation d'une recherche systématique. En fonction de la quantité et la qualité des résultats de recherche, la stratégie est affinée lors du screening des résumés pour obtenir un résultat le plus complet possible tout en restant maniable. Cela veut dire qu'au cours du processus une recherche bibliographique systématique peut être rejetée de quelques étapes. Lorsqu'il s'agit d'une question mineure, ponctuelle (telle que la vérification d'un possible ajout d'une caractéristique relative à un diagnostic soignant donné), on peut décider de réaliser une recherche bibliographique « mineure ». Dans ce cas-là, le protocole de la recherche bibliographique systématique peut être négligé, les sources trouvées sont évaluées en fonction des critères décrits dans le chapitre 2.2. Toutes les personnes impliquées dans le processus de développement disposent d'une formation pratique soignante ou thérapeutique et d'au moins d'un diplôme d'une filière soignante ou thérapeutique.

2.1.1 La question de recherche

La question de recherche est d'une importance essentielle pour la recherche de l'évidence et pour l'extraction de données. Plus la question est précise, meilleurs seront les résultats obtenus par la recherche. Le point de départ du processus de recherche pour établir ou mettre à jour une ligne directrice de pratique et/ou des éléments déterminés est donc la formulation explicite et précise d'une ou de plusieurs questions. Une recherche bibliographique d'orientation (l'entrée dans un sujet avec une approche ciblée sans être trop systématique) et la spécification de différentes conditions ou facteurs cadres de la thématique présentent des bonnes pratiques pour délimiter un sujet ou préciser une question. Des aspects pertinents peuvent être :

- Le groupe cible ou les personnes d'intérêt
- Le setting
- Le contexte
- L'acte (soignant), le phénomène, la situation, ... d'intérêt
- Les questions essentielles : qui, quand, quoi, comment, pourquoi ?
- La délimitation par rapport à des thématiques avoisinantes ou apparentées

Pour formuler une question, il faut veiller à deux choses fondamentales : d'un côté, il doit s'agir d'une question qui peut être répondue, mais de l'autre, la structure même de la question ne doit pas anticiper le résultat (présupposé). Il faut, en plus, veiller absolument à ce que la question ne contienne pas de présupposés et soit formulée sans parti pris pour ne pas viser (inconsciemment) le présupposé. Ainsi, la question « Lequel des concepts soignants A ou B a un effet plus avantageux sur la qualité de vie des personnes atteintes d'une démence ? » est à privilégier à la question « Est-ce que l'intervention de soins A présente vraiment un effet plus avantageux sur la qualité de vie des personnes atteintes d'une démence que l'intervention de soins B ? ».

Surtout dans le contexte clinique et pour des questions autour des concepts d'intervention de soins, y compris dans ENP, s'impose la méthode PICO, depuis longtemps établie en médecine basée sur les preuves, (Behrens & Langer, 2004 ; Panfil, Ivanović, & Conrad, 2011). L'acronyme PICO désigne les paramètres **P**atient/**P**opulation – **I**ntervention – **C**ontrôle/**C**omparaison – **O**utcome. Selon cette approche, une question de recherche devrait impliquer tous ces paramètres :

- P (patient/population) : description du groupe cible
- I (intervention) : technologie, méthode ou intervention faisant l'objet de l'étude ou de la recherche
- C (contrôle/comparaison) : alternatives auxquelles on peut comparer la technologie, la méthode ou l'intervention
- O (outcome) : la grandeur cible. Qu'est-ce que l'on veut atteindre ou comment peut-on mesurer le résultat ?

Lorsqu'il s'agit d'une recherche portant sur un concept d'intervention de soins pour le développement de ENP, la méthode PICO peut être appliquée telle quelle. Mais il faut toujours prendre en compte le contexte – ce qui illustrera l'exemple suivant :

	Question
Population/bénéficiaire de soins/contexte diagnostique	Est-ce que les patient-e-s avec démence hospitalisé-e-s présentent
Intervention	grâce à l'instauration d'un café nocturne,
Contrôle	par rapport à des cliniques sans une telle offre,
Résultat	des résultats différents lors du Mini Mental Status Test (MMST) ?

Tableau 1 : Exemple 1 pour une question portant sur des interventions de soins, formulée selon la méthode PICO

	Question
Population/bénéficiaire de soins/contexte diagnostique	Est-ce que des bénéficiaires de soins atteint-e-s de fatigue pathologique perçoivent,
Intervention	grâce à des interventions de soins complémentaires d'acupression,
Contrôle	par rapport à l'absence d'autres interventions de soins complémentaires,
Résultat	un changement quant à la fatigue subjectivement ressentie ?

Tableau 2 : Exemple 2 pour une question portant sur des interventions de soins, formulée selon la méthode PICO

Le développement/perfectionnement des lignes directrices de pratique ENP ne concerne pas uniquement les concepts d'interventions de soins, mais l'accent peut également être mis sur autres composants d'un diagnostic soignant – à savoir le diagnostic proprement dit ou les caractéristiques, étiologies et ressources correspondantes. Dans ce cas-là, il est possible d'adapter la méthode PICO de sorte qu'elle soit applicable pour l'élaboration de questions dans ce contexte. L'exemple suivant en illustre l'idée :

	Question
Population/bénéficiaire de soins/contexte diagnostique	Les bénéficiaires des soins atteint-e-s d'une fatigue pathologique,
Caractéristiques et symptômes observables/mesurables	quelles caractéristiques/symptômes mesurables, observables ou rapportés présentent-ils/elles,
Résultat	qui permettent une vérification univoque de la présence du diagnostic soignant « présente une fatigue pathologique » ?

Tableau 3 : Exemple d'une question de recherche ciblant les caractéristiques d'un diagnostic d'après la méthode PICO

2.1.2 Déduction de termes de recherche

Une question formulée avec le plus de clarté et précision possible présente une base idéale pour déduire des termes de recherche ou mots clés à utiliser pour une recherche de base de données. En règle générale, on peut subdiviser une question de recherche en plusieurs aspects ou concepts centraux qui devraient également être saisis en tant que concepts individuels. Voici un exemple :

« Quel est l'effet du **Snoezelen** par rapport à la **validation intégrative** sur l'apparition de **comportements défis** chez des **résident-e-s dément-e-s en institution** ? »

Les concepts et mots-clés centraux qu'on peut identifier à partir de cette question sont les suivants :

- résident-e-s avec démence en institution (population)
- Snoezelen (intervention)
- validation intégrative (intervention de contrôle)
- comportements défis (résultat)

Le plus souvent, la variation des termes de recherche pour chaque concept central de la question de recherche augmente considérablement les chances d'obtenir des résultats pertinents tout en réduisant le risque des « tâches aveugles » dans la réalisation de la recherche documentaire. Pour multiplier les termes de recherche, on peut notamment avoir recours à des synonymes, termes génériques et spécifiques et à des traductions anglaises avec leurs synonymes respectifs. Sans oublier les différentes formes fléchies d'un mot. Tous les éléments/concepts centraux de la question de recherche devraient d'abord être analysés un par un selon cette approche ; les termes de recherche qui en résultent peuvent ensuite être combinés pour former des phrases de recherche pertinentes. L'établissement d'une matrice tabulaire s'est avéré une aide pratique pour identifier et multiplier les termes et mots-clés de recherche. Dans l'exemple suivant qui reprend la question présentée ci-haut, les champs de la matrice ne sont remplis que partiellement à titre illustratif.

	Concept central 1 Résident-e en institution	Concept central 2 Comportement défi	Concept central 3 Démence	Concept central 4 Validation (intégrative)	Concept central 5 Snoezelen
Terme générique	Institution de soins, bénéficiaires de soins	Comportement	Troubles cognitifs	Communication	Stimulation sensorielle, aménagement de l'environnement
Terme spécifique	--	Wandering, agressivité,...	Maladie d'Alzheimer	Acceptation	--
Synonymes	--	Anomalie comportementale	--	--	--
Traduction anglaise	resident, nursing home,...	challenging behavior, defiantness,...	dementia, Alzheimer disease, mental disorder,...	validation,...	Snoezelen, snoezeling

Tableau 4 : Exemple d'une matrice tabulaire pour identifier et multiplier les termes de recherche pertinents

2.1.3 Critères d'inclusion et d'exclusion

Les critères d'inclusion et d'exclusion relatifs aux questions de recherche sont décisifs pour l'extraction de données. L'objectif en est de fixer le cadre formel de la recherche et de mettre l'accent sur ce qui est vraiment pertinent et praticable (filtrage supplémentaire). Parmi les critères possibles : la langue de publication, la période de parution, les types de publication/d'étude pertinents, la focalisation sur un groupe d'âge déterminé, le sexe, le setting, les professions, le choix des bases de données ou autres sources à utiliser. Un autre aspect important du choix, c'est que tout critère d'inclusion ou d'exclusion appliqué doit être justifiable, l'argumentation correspondante est à documenter. Pour l'exemple présenté, les critères d'inclusion/d'exclusion pourraient être les suivants :

Critères d'inclusion	Critères d'exclusion
Groupe d'âge : > 65 ans (Justification : les démences préséniles sont très rares)	Types de publication : exclusion de rapports d'expérience, études de cas particuliers et autres semblables (Justification : peu de résultats quant à l'efficacité à attendre)
Période de parution des publications : à partir de 1990 (Justification : avant cette date, ni la validation ni le Snoezelen n'ont été appliqués à une échelle considérable)	
Langues des publications : allemand et anglais (justification : les traductions professionnelles sont très coûteuses en argent et en temps)	

Tableau 5 : Critères d'inclusion et d'exclusion à titre illustratif

Souvent, la formulation définitive des critères d'inclusion et d'exclusion n'est possible que pendant la réalisation de la recherche : ce n'est pas rare qu'au cours de la recherche surgissent des aspects thématiques ou formels qui soulèvent à nouveau la question d'une prise en compte ou d'une exclusion.

2.1.4 Définition des bases de données et d'autres sources de recherche

Un aspect central dans l'élaboration d'une stratégie de recherche est la question de savoir où chercher avec les plus grandes chances de succès. Outre les sources classiques telles que les catalogues de bibliothèques ou les bibliographies, une multitude énorme et presque inépuisable d'outils et d'options de recherche est disponible notamment en ce qui concerne la recherche numérique. Dans le contexte du travail scientifique, une recherche (peu structurée) à l'aide d'un moteur de recherche général comme Google a évidemment moins de pertinence qu'une recherche numérique par exemple dans une base de données spécialisée. La figure suivante résume les options et sources essentielles d'une recherche documentaire systématique :

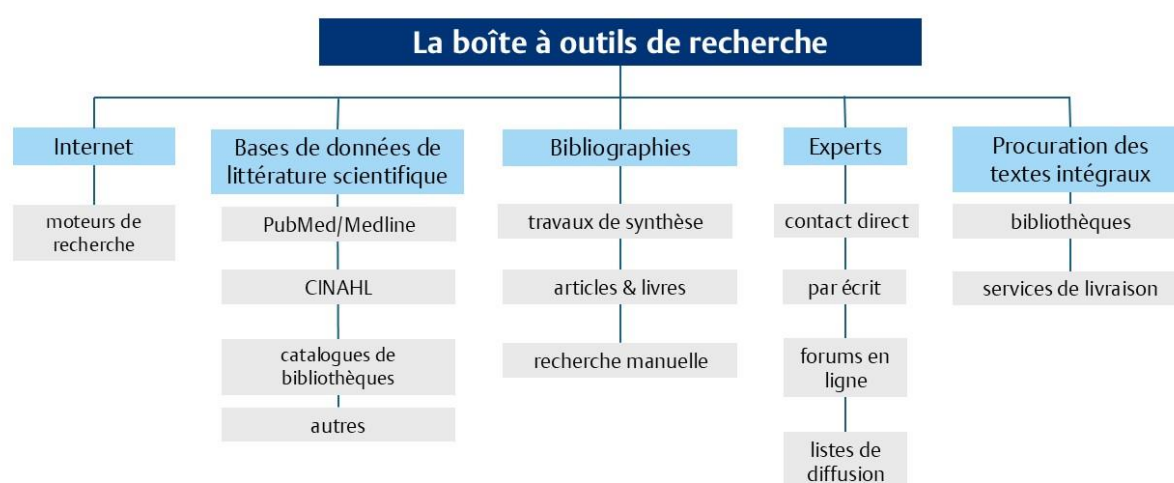


Figure 2 : Sources et options essentielles d'une recherche documentaire systématique (source : RECOM d'après Simon, 2018)

Les sources documentaires les plus pertinentes pour le développement de ENP sont les bases de données spécialisées telles que PubMed/MEDLINE, CINAHL (Cumulative Index to Nursing and Allied Health Literature), Cochrane Library, GeroLit, PsycNet et bien plus encore. Les lignes directrices soignantes et médicales des associations, sociétés et éditeurs spécialisés représentent également un riche fonds de connaissances pour le développement de ENP : souvent, il s'agit de travaux de synthèse systématiques qui sont élaborés et vérifiés avec des procédures sophistiquées et fournissent des recommandations basées

sur les preuves. Elles n'ont pas de caractère obligatoire, mais elles résument l'état des connaissances sur des questions très spécifiques et proposent des approches de solutions basées sur les preuves. Des références pertinentes pour trouver des lignes directrices sont par ex. le groupe de travail des sociétés scientifiques et médicales allemandes (AWMF), la base de données des lignes directrices et des standards du centre allemand pour la qualité des soins (ZQP), le National Institute for Health and Clinical Excellence (NICE), les Lignes directrices sur les pratiques exemplaires de l'Association des Infirmières et Infirmiers autorisés de l'Ontario (RNAO) ou le Scottish Intercollegiate Guideline Network (SIGN).

Si besoin, d'autres approches peuvent compléter la recherche documentaire systématique dans les bases de données spécialisées, les catalogues de lignes directrices, bibliothèques et autres sources structurées. Prenons comme illustration « Google Scholar », un dérivé spécifique du moteur de recherche Google, conçu pour la recherche générale de documents scientifiques. La méthode dite « boule de neige » (« footnote chasing » en anglais) peut également représenter une stratégie de recherche utile. Cela désigne la quête à base des références bibliographiques de publications spécialisées, récemment trouvées ou déjà connues. Cette méthode est particulièrement prometteuse avec des travaux de synthèse pour découvrir d'autres publications d'intérêts ou des termes de recherche complémentaires pour réaliser de nouvelles recherches.

Pour documenter la recherche bibliographique systématique, les points suivants concernant le choix des bases de données spécialisées et autres sources devraient être retenus dans le protocole de recherche :

- Le choix des bases de données numériques avec une brève justification
- Le choix des catalogues de bibliothèques/ressources en accès libre avec une brève justification
- Le choix des catalogues de lignes directrices (par ex. AWMF) avec une brève justification
- Le choix d'autres endroits de recherche (revues spécialisées, littérature grise, moteurs de recherche en ligne, méthode boule de neige etc.) avec une brève justification

2.1.5 Phrases de recherche et options d'influence sur le déroulement de la recherche

Pour assurer une recherche systématique la plus ciblée, maniable et complète possible et sans laisser des « tâches aveugles », mais aussi pour être en mesure de réagir de manière adéquate aux résultats de recherche non prévisibles en termes de qualité et quantité (trop ou trop peu de résultats, la plupart des résultats hors sujet...), il est fortement recommandable de générer différentes phrases de recherche. Outre les phrases de recherche « justes » supposées, il convient de considérer des alternatives pour limiter ou élargir la recherche. Les techniques ou aspects suivants devraient être pris en compte pour élaborer les phrases de recherche :

- **Recherche par champs** (utilisation de la structure des références bibliographiques dans une base de données, par ex. recherche par auteur, par titre, par abstract, ...)
- **Filtrage** surtout avec les critères d'inclusion/d'exclusion (par ex. période de publication, langue, groupe d'âge...)
- Utilisation de la **recherche par mots-clés (ou termes MeSH¹)** qui caractérisent une publication et permettent une détermination rapide des contenus. Les mots-clés ne doivent pas obligatoirement être contenus dans le texte proprement dit de la publication.
- **Troncature et masquage**, c'est-à-dire l'utilisation de jokers pour éviter la saisie de toutes les variantes grammaticales d'un terme de recherche ou d'une combinaison de mots. La troncature (souvent avec le caractère *) cherche les racines de mots avec toute sorte de terminaison ou de suffixe, tandis que le masquage remplace un seul caractère dans un terme de recherche (souvent avec le caractère # ou ?)

¹L'acronyme MeSH (pour « Medical Subject Headings ») décrit un thésaurus utilisé pour indexer des articles et livres en médecine et sciences de la santé. Il est notamment appliqué dans la base de données MEDLINE (PubMed).

- Les **opérateurs booléens** qui servent à relier des termes de recherche de manière ciblée. Les opérateurs booléens étant l'une des fonctions les plus puissantes pour une recherche documentaire systématique, il vaut mieux vérifier deux fois leur application (source d'erreur courante). Les opérateurs les plus essentiels sont :
 - **AND** (rarement ET en français. N'affiche que les résultats qui contiennent tous les termes connectés – « et...et... »).
 - **OR** (rarement OU en français. Affiche tous les résultats qui contiennent le terme A ou le terme B ou le terme C (...) – « ou... ou ... »).
 - **NOT** (rarement SAUF en français. Exclut tous les résultats qui contiennent le(s) terme(s) correspondant(s) – « en aucun cas »).

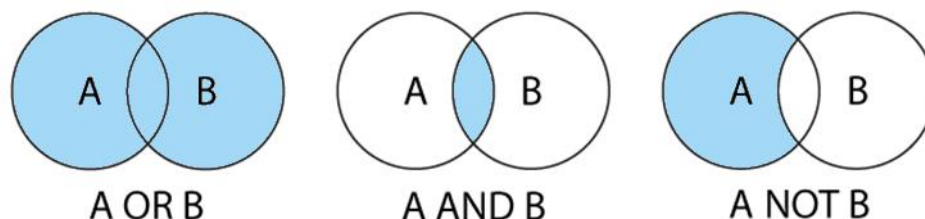


Figure 3 : Les opérateurs booléens et leurs effets sur la recherche documentaire (source : présentation originale)

- La **phrase de recherche** avec laquelle les termes saisis sont recherchés exclusivement dans l'ordre exact indiqué. Cette méthode peut être très utile notamment pour rechercher des concepts avec une combinaison fixe de plusieurs mots (par ex. « aidant familial »). La recherche par phrase est souvent réalisée en mettant les termes recherchés entre guillemets (« ... ») ou en activant une option spécifique dans le masque de recherche de la base de données (par ex. la fonction « ordre exact des mots »).

Il s'avère souvent utile de séparer, dans une première démarche, les différents concepts essentiels d'une phrase de recherche, d'élaborer des phrases de recherche à cette base – et d'assembler ces phrases de recherche finalement pour les dernières opérations de recherche. Pour construire des phrases partielles, mais aussi pour la phrase complète finale, l'utilisation d'une matrice peut faciliter la tâche, offrant une vue d'ensemble des termes et techniques de recherche (recherche par champs, troncature, opérateurs booléens etc.). Sans prétendre être exhaustive, la matrice présentée ci-après illustre la construction de phrases partielles/la phrase complète à partir des aspects discutés jusqu'ici.

Phrase de recherche partielle pour le concept central 1 Résident-e en institution			
Opérateur connecteur (AND, OR)	Terme de recherche (si besoin tronqué, masqué ou sous forme de phrase)	Recherche par champs ? (titre, MeSH,...)	Alternatives ?
	Terme principal resident	si besoin recherche par champs et abstract	
OR	Synonyme 1 « nursing home »	si besoin recherche par champs et abstract	
OR	Synonyme 2 « elder care »	si besoin recherche par champs et abstract	« elderly care », « geriatric care »
OR	Synonyme 3		
Termes à exclure pour ce concept ?			
NOT	Hospital OR « outpatient care » OR « home care »		

Tableau 6 : Construction d'une phrase de recherche partielle pour un concept central de la question de recherche

Phrase de recherche partielle pour le concept central 2 Comportement défi			
Opérateur connecteur (AND, OR)	Terme de recherche (si besoin tronqué, masqué ou sous forme de phrase)	Recherche par champs ? (titre, MeSH,...)	Alternatives ?
	Terme principal « comportement défi »	si besoin recherche par champs et abstract	
OR	Synonyme 1 defiantness	si besoin recherche par champs et abstract	
OR	Synonyme 2 wandering	si besoin recherche par champs et abstract	
OR	Synonyme 3 aggressi*		
Termes à exclure pour ce concept ?			
NOT			

Tableau 7 : Construction d'une phrase de recherche partielle pour un autre concept central de la question de recherche

Une fois les phrases de recherche et possibles alternatives élaborées pour tous les concepts centraux, il s'agit d'assembler toutes les facettes pour obtenir une phrase de recherche complète (et des alternatives correspondantes). A cette fin, on peut se servir souvent de l'historique de la recherche dans les bases de données spécialisées (par ex. en combinant la « recherche concept central 1 » avec l'opérateur « AND » et la « recherche concept central 2 »). Pour assurer une documentation cohérente et pour être en mesure de préciser, au besoin, la recherche finale, toutes les « recherches complètes » y comprises les recherches alternatives (si le résultat obtenu s'avère être trop large, trop étroit ou inadéquat), devraient être réfléchies au préalable et documentées sous forme de tableaux. C'est généralement au moment d'assembler les phrases partielles à la phrase complète que les critères d'inclusion et d'exclusion sont appliqués.

Phrase de recherche complète – variante 1				
	Opérateur connecteur (AND, OR, NOT)		Opérateur connecteur (AND, OR, NOT)	
Phrase de recherche partielle pour concept central 1	AND	Phrase de recherche partielle pour concept central 2	AND	Phrase de recherche partielle pour concept central 3
Termes à exclure pour la phrase de recherche complète ?				
			NOT	
Filtres de recherche (critères d'inclusion/d'exclusion)				
Période de publication	Langue	Conceptions d'étude/de publication	Groupe d'âge	...
Phrase de recherche complète – variante 2				
	Opérateur connecteur (AND, OR, NOT)		Opérateur connecteur (AND, OR, NOT)	
Phrase de recherche partielle pour concept central 1	AND	Phrase de recherche partielle pour concept central 2	AND	Phrase de recherche partielle pour concept central 3
Termes à exclure pour la phrase de recherche complète ?				
			NOT	

Filtres de recherche (critères d'inclusion/d'exclusion)				
Période de publication	Langue	Conceptions d'étude/de publication	Groupe d'âge	...

Tableau 8 : Exemple d'une matrice pour établir des phrases de recherche et assembler les recherches ciblées sur les différents concepts centraux de la question de recherche

En complément à la représentation des phrases de recherche en tableau, il peut être utile de documenter les phrases sous forme textuelle avec une syntaxe spécifique aux bases de données. Tout comme en mathématiques, il faut veiller à une utilisation correcte des parenthèses et respecter les particularités et techniques propres à chaque base de données. Les parenthèses expriment la priorité dans le traitement de la phrase dans la base de données : les éléments à l'intérieur d'une parenthèse sont traités en premier, ensuite les éléments à l'extérieur de la parenthèse. Généralement, il est également possible d'imbriquer des parenthèses. Pour rechercher un terme dans un champ spécifique (par ex. uniquement dans le titre ou dans l'abstract), il faut appliquer différentes techniques en fonction de la base de données utilisée – souvent cela est réalisé en plaçant le champ de recherche souhaité entre crochets []. La composition manuelle d'une phrase de recherche à partir d'une syntaxe requiert une attention accrue et, dans l'idéal, la vérification par une autre personne pour éviter des erreurs susceptibles d'entraîner des conséquences non souhaitées sur les résultats de recherche. Pour illustrer la problématique, voici un bref exemple d'une syntaxe de recherche relative à la question de recherche discutée :

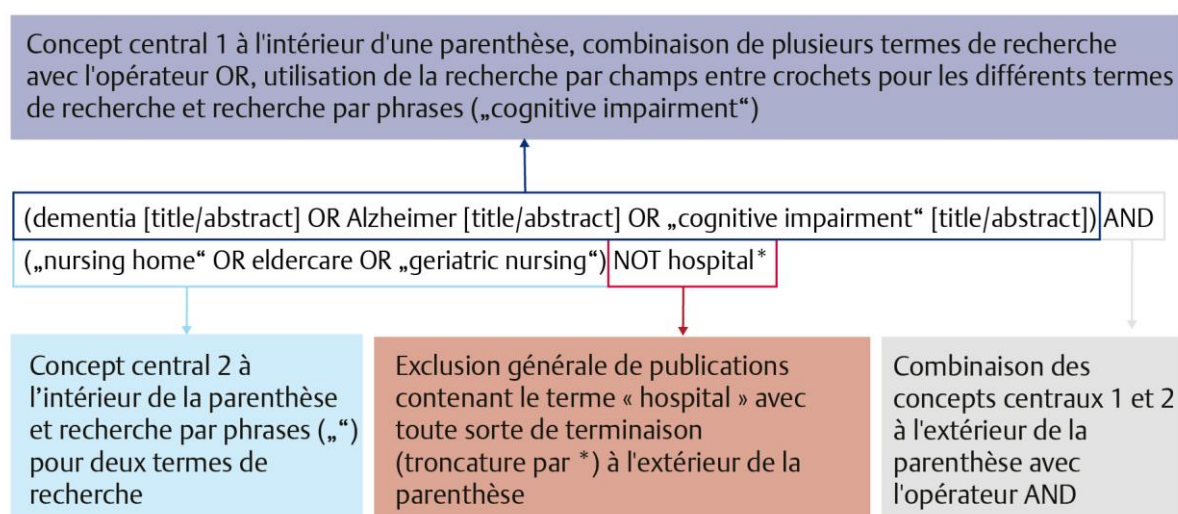


Figure 4 : Composition textuelle d'une phrase de recherche connectant plusieurs concepts centraux à base de la syntaxe spécifique à une base de données

Pour saisir le déroulement de la recherche avec les modifications au niveau de la quantité et qualité des résultats dans différentes bases de données spécialisées (et autres endroits de recherche, si besoin), il est recommandable de documenter l'historique de toutes les opérations de recherche réalisées. Toutes les bases de données et catalogues spécialisés importants offrent la possibilité de consulter et copier (par ex. en capture d'écran) l'historique de la recherche ou de l'enregistrer pour une utilisation ultérieure. En règle générale, cela requiert la création d'un compte d'utilisateur (dans la plupart des cas gratuit) pour la base de données correspondante.

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		

Figure 5 : Exemple de la saisie d'une opération de recherche, réalisée dans la base de données PubMed (<https://pubmed.ncbi.nlm.nih.gov/>)

2.1.6 Abstract screening et procuration des textes intégraux

Un compte d'utilisateur pour les bases de données pertinentes peut être un outil de gestion très utile au moment de trier les résultats obtenus à l'aide des phrases de recherche finales pour décider quelles publications se procurer pour analyser et évaluer ensuite les textes intégraux. Dans la plupart des cas, cette décision est prise à base de ce que l'on appelle « abstract screening ».

Lors de cette démarche, tous les résultats de l'opération de recherche finale sont vérifiés pour déterminer s'ils remplissent ou non les critères d'inclusion en matière de contenu et autres. Pour ce faire, le résumé/synopsis – « abstract » – de la publication qui est normalement consultable directement dans la base de données est étudié et évalué suivant un schéma/questionnaire à définir au préalable. Ensuite, toutes les publications qui remplissent les critères d'inclusion ou pour lesquelles l'abstract seul ne suffit pas pour prendre une décision avec assurance, sont procurées en version intégrale. Tous les résultats qui peuvent être exclus lors du screening avec certitude et justification ne seront plus pris en compte dans la suite. A chaque fois que c'est possible, le screening des abstracts devrait être réalisé par deux personnes de manière indépendante qui comparent ensuite leurs résultats et les discutent au besoin, si des décalages importants sont constatés. Tant les exclusions comme les inclusions des résultats de recherche sont documentées avec une brève justification dans le protocole de recherche.

Outre la description textuelle² du nombre de résultats initiaux trouvés dans différentes bases de données et autres sources, du nombre de doubles et du nombre de résultats exclus par abstract screening ou encore analyse du texte intégral, il est en plus recommandé – comme le préconise par exemple le PRISMA statement (Moher et al., 2011 ; Ziegler & König, 2011) – d'établir un diagramme de flux pour représenter les différentes phases de la recherche documentaire ou les phases de tri des résultats/publications. La figure suivante illustre un tel diagramme à titre d'exemple.

² Par ex. « Parmi le total de <XX> résultats, <YY> publications ont été évaluées comme pertinentes pour l'obtention en texte intégral. Parmi ces <YY> publications pertinentes, <ZZ> ont finalement pu être obtenues en texte intégral. »

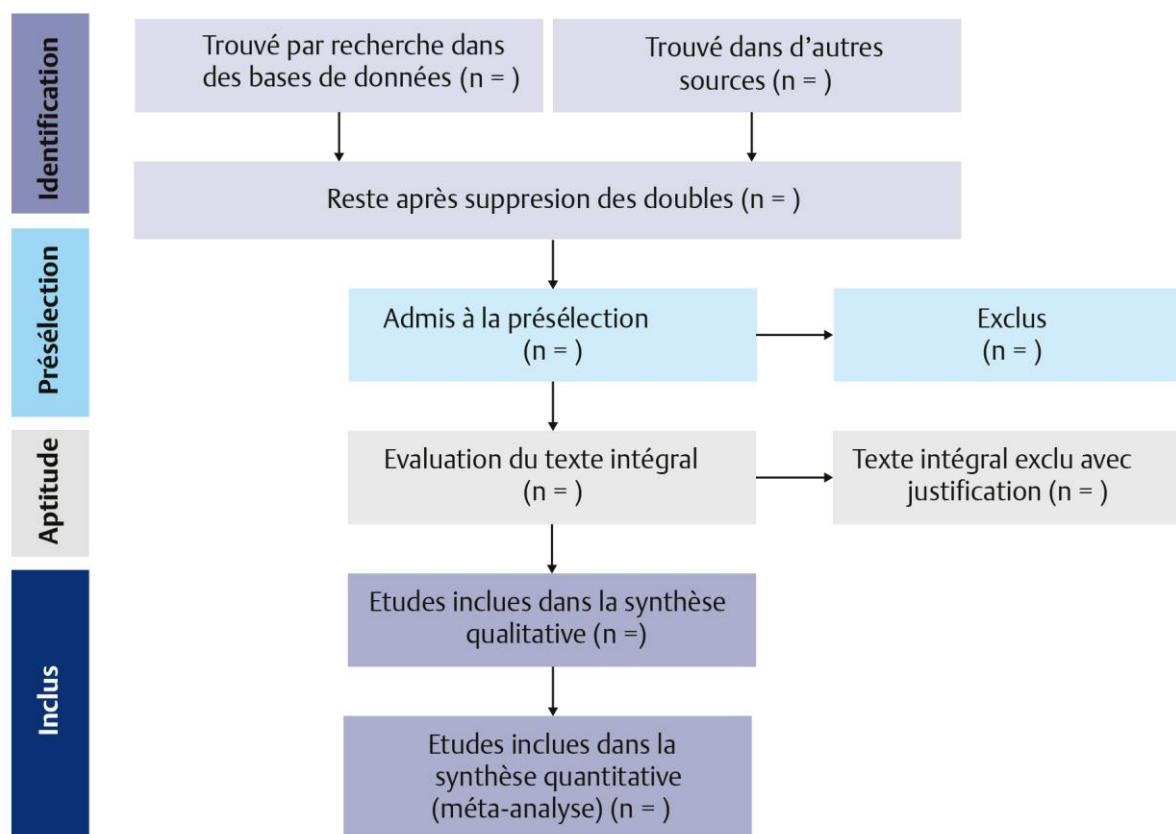


Figure 6 : Diagramme de flux pour illustrer l'exclusion de publications au cours de la recherche (source : RECOM d'après Ziegler & König, 2011, p. e11)

En complément des accès disponibles (en ligne) aux bibliothèques, éditions numériques de revues et différents éditeurs, on peut avoir recours à des services de livraison de documents, comme par ex. Subito³. La gestion des textes intégraux et de l'ensemble des références bibliographiques se réalise avec le programme de gestion bibliographique EndNote.

2.2 Critères et outils appliqués pour l'évaluation de l'évidence, la sélection d'évidence et l'extraction de données

Lors (de la poursuite) du développement des lignes directrices de pratique ENP, on procède souvent par analogie au développement de lignes directrices du groupe de travail des sociétés scientifiques et médicales (AWMF). En règle générale, on opte pour une approche de recherche itérative, hiérarchique pour élaborer ou réviser les lignes directrices de pratique ENP (Schmucker et al., 2017). Cela veut dire qu'on privilégie les évidences déjà « pré-évaluées » suivant la hiérarchie des « 6S » (Dicenso, Bayley, & Haynes, 2009). Cette hiérarchie de l'évidence pré-évaluée prévoit une prise en compte successive de a) lignes directrices de haute qualité, puis b) l'évidence agrégée (travaux de synthèse systématiques) et c) la littérature primaire. Si l'évidence d'un concept d'intervention soignante et/ou un rapport diagnostique n'est pas appuyée par des lignes directrices/travaux de synthèse de haute qualité, on consulte la littérature primaire pertinente et autre littérature spécialisée. En fonction du type de publication ou du design de recherche concerné, différentes méthodes d'évaluation (critical appraisal) sont appliquées pour déterminer la qualité. Le chapitre suivant précisera les détails de ces méthodes.

³ Cf. <https://www.subito-doc.de>

2.2.1 Critical appraisal de publications dans le processus de développement de ENP

Dans le cadre du développement ENP, différents formulaires d'évaluation ont été utilisés de 2004 à 2021 pour la lecture critique (cf. figure 7). En 2021, ces formulaires ont été remplacés par des versions entièrement nouvelles qui déterminent depuis les processus de développement. Le pourquoi de cette nouvelle conception des outils d'évaluation s'explique notamment par l'orientation de plus en plus internationale des nouvelles approches ainsi que par de nombreuses nouvelles connaissances obtenues dans ce contexte. Il existe un large éventail d'outils conçus pour évaluer la qualité d'études ou publications (Browman et al., 2017 ; Cochrane Deutschland et al. (éd.), 2021; Dreier et al., 2010; Ma et al., 2020; Platz 2018 ; Thole et al., 2007). Un rapport de l'année 2010, portant sur un Health Technology Assessment (HTA), a identifié déjà 125 différents outils liés à des études d'intervention, d'observation et de diagnostic ainsi que des revues et méta-analyses et en a constaté des variances importantes en termes de qualité (Dreier et al., 2010). Les auteurs décrivent en plus différentes opérationnalisations de la qualité des études dans leur recherche bibliographique systématique. Ils recommandent pour l'évaluation de différencier la qualité méthodologique d'un travail de recherche du risque d'un effet de distorsion systématique (biais) avant, pendant et après la réalisation et de l'évaluer à part. En plus, une distinction explicite entre la qualité du rapport et la qualité de l'étude est préconisée, un mélange des deux pouvant entraîner de mauvaises interprétations. L'évaluation d'études cible en grande partie la validité interne, c'est-à-dire la validité des énoncés formulés. Chaque type d'étude ou design de recherche présente ses propres risques de distorsion. Ainsi, pour évaluer l'efficacité des mesures d'acupression appliquées en soins pour réduire la nausée, la consistance des structures et des observations entre le groupe d'intervention et le groupe de contrôle est essentielle, tandis que l'évaluation de travaux de synthèse systématiques dépend notamment de la qualité des études individuelles choisies ainsi que des critères d'inclusion ou d'exclusion.

Le type et l'étendu des études utilisées pour le développement ou perfectionnement des lignes directrices de pratique ENP varie en fonction des questions de recherche posées et des études disponibles. S'il s'agit par exemple de trouver des preuves étiologiques quant à un diagnostic soignant, la question formulée est généralement très précise. Les résultats de recherche obtenus par une question précise sont normalement moins nombreux, alors que des questions plutôt abstraites peuvent facilement produire un nombre beaucoup plus important de résultats. Au cours du processus de recherche visant le développement ENP qui est marqué par une approche itérative, une recherche documentaire supplémentaire ciblant par exemple un concept d'action spécifique peut s'imposer. Toute publication qui sert d'appui pour un item ENP est d'abord évaluée en termes de qualité. Ce processus complexe qui est l'évaluation critique de toute publication utilisée est réalisée avec des outils d'évaluation adaptés au type de référence concerné. Différents types d'étude ou designs de recherche se prêtent en fonction de la question soulevée lors du développement ENP. Prenons par exemple les concepts d'intervention de soins : les essais randomisés contrôlés (en anglais : randomized controlled trial – RCT) y sont très pertinents puisqu'ils présentent, avec une réalisation adéquate, le risque de distorsion le plus faible. Dans l'idéal, il existe en plus des travaux de synthèse et/ou des méta-analyses qui utilisent des RCT comme études primaires. Quant à des recherches ciblant plutôt des caractéristiques et/ou étiologies relatives à un diagnostic soignant ou des expériences vécues par les bénéficiaires de soins, on privilégie d'autres types d'études, tels que des designs d'études qualitatifs. Pour la plupart des types d'étude et de publication, un large éventail d'outils est disponible pour la lecture critique. La liste ci-après résume les outils utilisés pour le développement ENP, suivie par des précisions individuelles sur chacune des approches et, le cas échéant, les adaptations spécifiques apportées pour l'usage interne.

- 2.2.1.1 Revues systématiques (RS) & méta-analyses** : évaluation selon AMSTAR 2 (Assessment of Multiple SysTemAtic Reviews)
- 2.2.1.2 Etudes observationnelles (études cas-témoins, de cohorte, transversales)** – Critical Appraisal Tools du JBI (Joanna Briggs Institute)
- 2.2.1.3 Essais randomisés contrôlés (RCT)** – systématique d'évaluation sur le modèle des recommandations Cochrane selon Platz (2021)
- 2.2.1.4 Travaux de recherche qualitatifs** – Critical Appraisal Checklist for Qualitative Research : systématique d'évaluation du JBI pour la recherche qualitative (par ex. Grounded Theory/théorie ancrée, analyse de contenu...)
- 2.2.1.5 Lignes directrices et recommandations pour la pratique clinique (RPC)** – systématique d'évaluation sur le modèle de l'outil AGREE-II
- 2.2.1.6 Travaux de recherche par les méthodes mixtes** – systématique d'évaluation (par extraits) du MMAT (Mixed Methods Appraisal Tool)
- 2.2.1.7 Livres spécialisés/manuels** – réalisation individuelle des critères d'évaluation de la bibliothèque de l'université Cornell à Ithaca, New York
- 2.2.1.8 Littérature grise et sites web** – critères d'évaluation de la AACODS checklist (Authority, Accuracy, Coverage, Objectivity, Date, Significance)

La littérature pertinente pour chaque ligne directrice de pratique ENP est systématiquement évaluée selon les critères de qualité fixés en fonction du design d'étude ou du type de publication. Des check-lists soutiennent l'équipe de développement de ENP dans le processus d'évaluation. La qualité de l'évidence est ensuite indiquée selon un schéma de classification (degré d'évidence, level of evidence). La structure du tableau d'évidence et son utilisation est précisée dans la partie 2.2.3.

Les outils utilisés pour la lecture critique sont présentés de manière approfondie au préalable. La figure suivante montre un outil rempli à titre d'illustration pour donner une idée de l'application concrète des check-lists appliqués pour le développement de ENP.

Critical appraisal d'essais randomisés contrôlés (RCT) selon Platz (2021) :¹

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

Réviseur 1 : Wieteck, Pia

Réviseur 2 : Kraus, Sebastian

Critical appraisal d'essais randomisés contrôlés (RCT) selon Platz (2021)	Oui	Non	Imprécis	Non évaluable
1. Les critères d'inclusion et d'exclusion ont-ils été clairement définis ? (validité externe)	☒	☐	☐	☐
2. Y a-t-il une définition claire et une évaluation adéquate des grandeurs cibles (outcomes) de l'étude (pertinence clinique) ?	☒	☐	☐	☐
3. Est-ce que des effets secondaires et l'acceptabilité des interventions ont été rapportés ?	☒	☐	☐	☐
4. Y a-t-il eu un examen « follow-up » adéquat (saisie des effets à long terme) ?	☐	☐	☐	☒
5. Y a-t-il une définition et une description claires des conditions expérimentales et de contrôle (contraste de l'étude) ?	☒	☐	☐	☐
6. Les participants sont-ils assignés de manière randomisée (biais de sélection) ?	☒	☐	☐	☐
7. L'assignation a-t-elle été maintenue secrète (« allocation concealment », biais de sélection) ?	☒	☐	☐	☐
8. Le groupe expérimental et celui de contrôle ont-ils été comparables au début de l'étude (biais de sélection) ?	☒	☐	☐	☐
9. Les patients et les intervenants ont-ils été mis en aveugle pendant le traitement et les traitements ont-ils été comparables dans les groupes randomisés, en dehors des effets examinés (biais de réalisation) ?	☒	☐	☐	☐
10. Les résultats de l'étude ont-ils été déterminés en aveugle (« biais de détection ») ?	☒	☐	☐	☐
11. La rédaction des comptes rendus n'a-t-elle pas été sélective (« reporting bias ») ?	☒	☐	☐	☐
12. Les données de résultat ont-elles été (presque) complètes (« attrition bias ») ?	☒	☐	☐	☐
13. Est-ce qu'une analyse en intention de traiter a été rapportée („attrition bias“) ?	☒	☐	☐	☐
14. Les résultats appuient-ils suffisamment les conclusions rapportées ?	☒	☐	☐	☐

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

Figure 7 : Check-list pour la lecture critique d'études (ici : essais randomisés contrôlés) rempli à titre d'illustration

2.2.1.1 Revues systématiques (RS) & méta-analyses – évaluation selon AMSTAR 2

En médecine, on prend souvent des revues systématiques comme base pour déterminer l'évidence externe au moment de sélectionner des thérapies ou approches diagnostiques appropriées. Les méthodes et processus utilisés pour les revues systématiques sont établis depuis longtemps. Cela vaut de même pour toute une série d'outils conçus pour évaluer la qualité de RS. Il existe de nombreuses méthodes d'évaluation permettant de classer la qualité de travaux de synthèse systématiques (Zeng et al., 2015). Seul le manuel pour « évaluer le risque de distorsion des travaux de synthèse systématiques » proposé par Schmucker et al. (2017) indique 20 différents outils. L'outil AMSTAR (Assessment of Multiple SysTemAtic Reviews) conçu pour évaluer des travaux de synthèse basés sur des essais randomisés contrôlés (RCT), est classé comme un outil validé dont la fréquence d'utilisation est la plus élevée – même en dehors de la médecine (Lühnen, Berger-Höger, & Richter, 2021 ; Schmucker et al., 2017). En plus, on a élaboré ce que l'on appelle des « reporting statements » pour des revues systématiques, fournissant différents critères pour rapporter ou publier un travail de synthèse systématique. Un exemple fréquemment utilisé est le PRISMA statement : « Preferred Reporting Items for Systematic reviews and Meta-Analyses » (Moher et al., 2011 ; Thangatorai, Lim, & Nalliah, 2018 ; Ziegler & König, 2011).

La qualité de l'AMSTAR, proposé par un groupe de travail canadien, a été confirmée par des études. Quant à une version révisée de l'outil (R-AMSTAR), publiée ultérieurement, les caractéristiques de mesures n'ont pas encore été suffisamment étudiées (Pieper et al., 2015). AMSTAR est également disponible en application en ligne (Shea et al., 2017). Les 11 questions directrices de l'outil initial sont répondues par « oui », « non », « peu clair » ou « non applicable ». Une évaluation globale n'est pas prévue. La version révisée (R-AMSTAR), par contre, attribue des points de score de 1 à 4 par question pour obtenir ainsi une meilleure quantification de la qualité totale d'une publication (Schmucker et al., 2017). Outre la version révisée, il existe une autre version remaniée, appelée AMSTAR 2. Le but de ce remaniement a été l'élargissement de l'application de l'AMSTAR aux travaux de synthèse, prenant en compte tant les études non randomisées que les méta-analyses (Shea et al., 2017). Etant donné que les études sans répartition aléatoire (non randomisées) sont assez répandues en soins et font par conséquent partie des travaux de synthèse correspondants, il paraît utile et efficace d'utiliser AMSTAR 2. A différence de la version initiale, AMSTAR 2 comprend 16 items à répondre, en fonction du critère d'évaluation concerné, par « oui », « partiellement rempli », « non » ou « non applicable ». La check-list précisant l'interprétation est gratuitement accessible en ligne⁴ (Shea et al., 2017). AMSTAR 2 attend toujours sa validation globale. Les examens réalisés jusqu'ici ont vérifié, à l'aide de statistiques kappa, la fiabilité inter-observateur, lui attestant essentiellement une qualité acceptable. L'ergonomie d'AMSTAR 2 a également fait l'objet d'une évaluation, avec le résultat qu'il faut compter une durée de 15 à 32 minutes pour l'application. Cela est justifiable par le fait que 10 questions de l'outil original ont été maintenues et que le développement est basé sur un large consensus d'experts et sur des retours utilisateurs (Shea et al., 2017). L'AMSTAR 2 est disponible en version française (Pallot & Rostagno 2021).

Lors de l'évaluation de la littérature, l'équipe ENP utilise les explications et aides à l'interprétation qui sont disponibles pour tous les 16 critères d'évaluation de l'AMSTAR 2.

L'AMSTAR 2, de même que la version originale, ne prévoit pas l'attribution d'un score pour quantifier et évaluer la qualité globale d'un travail de synthèse ou d'une méta-analyse. Au lieu de cela, on propose une évaluation sommaire à base du jugement de personnes qui appliquent l'AMSTAR 2. Shea et collègues (2017) proposent à cet effet un barème d'évaluation (cf. le tableau ci-après), tout en faisant remarquer que d'autres approches peuvent être utilisées.

⁴ cf. https://amstar.ca/Amstar_Checklist.php (consulté le 02/03/2022)

Evaluation de la confiance générale dans les résultats de l'évaluation d'une RS/méta-analyse	
Elevée	Aucune ou faiblesse non critique : La revue systématique fournit un résumé précis et compréhensif des résultats des études disponibles qui traitent la question concernée.
Modérée	Plus d'une faiblesse non critique : La revue systématique présente plus d'une faiblesse, mais pas de déficience critique. Elle peut fournir un résumé des études disponibles qui ont été prises en compte dans l'analyse.
Faible	Une déficience critique avec ou sans faiblesses non critiques : La revue présente une erreur critique (risque de biais) et pourrait ne pas être en mesure de fournir un résumé précis et compréhensif des études disponibles qui traitent la question concernée.
Critiquement faible	Plus d'une erreur critique avec ou sans faiblesses non critiques : La revue présente plus d'une erreur critique et ne devrait pas être considérée comme fournissant un résumé précis et compréhensif des études disponibles.

Tableau 9 : Evaluation sommaire de la confiance globale dans les résultats d'une revue systématique/méta-analyse (source : présentation et traduction originales selon Shea et al., 2017, p. 6)

Dans le développement ENP, l'évaluation synthétique d'une étude ou publication se réalise avec une méthode d'évaluation faisant référence à GRADE (Grading of Recommendations, Assessment, Development and Evaluation) qui ressemble fortement à ce schéma d'évaluation globale (cf. chapitre 2.2.3).

2.2.1.2 Etudes observationnelles (études cas-témoins, de cohorte, transversales) – Critical Appraisal Tools du JBI (Joanna Briggs Institute)

Les études observationnelles poursuivent des objectifs très différents. Allant de la découverte de nouveaux liens jusqu'à la confirmation d'éléments déjà connus. Les trois conceptions d'études les plus fréquentes, appliquées en fonction des objectifs pertinents (études cas-témoins, de cohorte, transversales) peuvent être évaluées avec le STROBE statement (Strengthening the Reporting of Observational Studies in Epidemiology) très répandu. Pour le développement du STROBE statement, on a pu profiter des expériences faites auparavant avec les outils conçus pour relater des études observationnelles. Notamment le CONSORT statement (Consolidated Standards of Reporting Trials), pensé surtout pour rapporter les RCT (Cuschieri, 2019a ; Moher et al., 2010). Le STROBE statement a été élaboré par un groupe de recherche en tenant compte de preuves empiriques et de réflexions théoriques. L'outil comprend 22 items dont une partie est conçue spécifiquement pour les designs de recherche des études cas-témoins, de cohorte ou des études transversales. Le STROBE statement offre aux auteurs un outil de soutien pour optimiser la rédaction des rapports d'études. On peut, certes, appliquer les critères du STROBE statement également pour interpréter les travaux de recherche et évaluer leur qualité (Cuschieri, 2019b ; Vandembroucke et al., 2014 ; von Elm et al., 2014) – sauf que cela ne correspond pas à leur finalité d'application. Cette utilisation détournée est vivement critiquée (Haile, 2021). Le STROBE statement (en sa version allemande) peut être utilisée à titre d'orientation et d'inspiration ou pour la rédaction de rapports portant sur le travail de développement de ENP.

Pour la lecture critique d'études observationnelles, de nombreux outils spécifiques sont disponibles qui sont souvent adaptés aux particularités du design d'étude pertinent. Dans ce contexte, Dreier et collègues (2010) mentionnent 30 outils conçus pour la lecture critique et le rapport des études observationnelles et ce nombre a encore augmenté depuis. A titre d'illustration, mentionnons les check-lists du **Scottish Intercollegiate Guidelines Network (SIGN)**, l'outil **ROBINS-I (Risk of Bias In Non-randomized Studies of Interventions)**, l'échelle **Newcastle-Ottawa (NOS)** ou l'outil **AXIS (Appraisal tool for Cross-Sectional Studies)** (Cochrane Deutschland et al. (éd.), 2021 ; Ma et al., 2020). L'outil ROBINS-I est utilisé, par exemple, par le groupe de travail des sociétés scientifiques et médicales (AWMF) pour la synthèse des évidences lors du développement de lignes directrices (Cochrane Deutschland et al. (éd.), 2021 ; Ma et al., 2020). Il s'agit d'un outil complexe à plusieurs étapes, pensé pour évaluer le risque de distorsion systématique (risk of bias) des résultats d'études interventionnelles non randomisées. Tout comme nombre d'autres outils, ROBINS-I implique certaines contraintes (Page, McKenzie, & Higgins, 2018). L'outil se limite aux études

quantitatives qui analysent l'utilité ou la nocivité d'une intervention. Cela comprend entre autres les études de cohorte, les études cas-témoins ou les études avant/après.

Après comparaison de différents outils conçus pour évaluer des études observationnelles, on a opté pour utiliser les check-lists proposées par l'institut Joanna Biggs (JBI, situé à Adelaïde, Australie), dans le contexte du développement de ENP (Ma et al., 2020). Méthode reconnue, elle couvre le spectre requis d'études observationnelles (Munn et al., 2020). Les outils du JBI sont largement utilisés et établis en soins, médecine et d'autres professions de la santé (Dietrich et al., 2016 ; Hopper et al., 2019 ; Porritt, Gomersall, & Lockwood, 2014 ; Stratil et al., 2021 ; Williams, Boylan, & Nunan, 2020). En plus, des explication ou aides à l'interprétation des différents critères d'évaluation sont disponibles pour chaque critical appraisal du JBI.

Parmi les études observationnelles, on distingue généralement les études analytiques et les études descriptives. Les études descriptives analysent les données saisies sans chercher à trouver des liens causaux. Les études transversales (ou : cross-sectionnelles) et les études longitudinales comptent parmi les études observationnelles descriptives. Les études analytiques, par contre, se consacrent à chercher de principes cause-effet, autres liens ou des facteurs de risque spécifiques (exposition). Deux exemples d'approches analytiques sont les études cas-témoins (études de contrôle de cas/case-control studies) ou les études de cohorte (cohort studies).

Les **études cas-témoins** peuvent avoir une conception rétrospective ou prospective. Les personnes présentant une maladie donnée ou un autre phénomène d'intérêt soignant sont comparées à des personnes non atteintes de cette maladie/ce phénomène. L'équipe ENP utilise, pour l'évaluation critique des études cas-témoins, la « Critical Appraisal Checklist for Case Control Studies » du JBI (Moola et al., 2020).

Dans une **étude de cohorte**, les participants sont sélectionnés selon des caractéristiques définies, puis cet échantillon sera observé sur une période prolongée. On peut saisir par ex. les facteurs de risque des sujets (par ex. obésité, diabète) et l'apparition ou non d'événements déterminés (par ex. escarre), par rapport à d'autres sujets non exposés. Pour les études de ce type, le JBI offre la « Checklist for Cohort Studies » (Moola et al., 2020) comme critical appraisal.

Les **études transversales** saisissent et décrivent des données ou paramètres obtenus une fois, à un moment défini, auprès d'un échantillon, et présentent donc une vue instantanée. L'outil d'évaluation du JBI permettant l'évaluation des études de cette conception est la « Checklist for Analytical Cross Sectional Studies » (Moola et al., 2020).

Pour les publications basées sur un **design de recherche quasi-expérimental** (à savoir toutes les études expérimentales qui travaillent avec des groupes non aléatoires et donc sans randomisation), on applique également la check-list du JBI prévue à cet effet, la « Checklist for Quasi-Experimental Studies ».

L'ensemble des outils de lecture critique du JBI avec aides à l'interprétation est accessible via une présence sur internet prévue à cet effet : <https://jbi.global/critical-appraisal-tools> (consulté le 24/05/2024).

2.2.1.3 Essais randomisés contrôlés (RCT) – systématique d'évaluation sur le modèle des recommandations Cochrane selon Platz (2021)

Pour les essais randomisés contrôlés (RCT de l'anglais *randomized controlled trial*) aussi, de nombreux outils de rapport et de lecture critique sont disponibles (Zeng et al., 2015). Un outil assez répandu pour les rapports de RCT est le CONSORT statement (Consolidated Standards of Reporting Trials) (Cuschieri, 2019a ; Moher et al., 2010). Cet outil qui comprend 25 items a été développé à base du SORT Statement (Standards of Reporting Trials) dont les faiblesses ont été maintes fois critiquées. Une rédaction solide des résultats des RCT est essentielle – vu que ce design de recherche, souvent considéré comme l'étalon-or en médecine et parfois en soins, n'est pas dénué de faiblesses méthodologiques qui risquent de distordre les résultats. Le CONSORT Statement peut donc aider dans le développement ENP à identifier les

rédactions fiables. En complément, il est préconisé d'étudier les explications et interprétations des questions/items du CONSORT Statement (cf. Knippschild et al., 2015 ; Moher et al., 2010).

Outre les listes de vérification pour les rédactions, il existe un grand nombre d'outils facilitant la lecture critique des essais randomisés contrôlés. Ainsi, le manuel Cochrane pour évaluer le risque de distorsion systématique (risk of bias) des études randomisées est largement utilisé pour l'élaboration de lignes directrices. Une deuxième version révisée de l'année 2021 est disponible (Cochrane Deutschland et al. (éd.), 2021 ; Schmucker et al., 2016). Généralement, l'outil peut être appliqué tant pour les essais randomisés individuels avec des groupes parallèles que pour les essais randomisés en cluster avec des groupes parallèles ou les essais croisés randomisés individuels (Cochrane Deutschland et al. (éd.), 2021). Partant des recommandations de Cochrane concernant le risque de distorsion, Platz (2021) a établi un critical appraisal pour les essais randomisés contrôlés. Epruvé par le groupe de travail des sociétés scientifiques et médicales (AWMF) lors du développement des lignes directrices, cet outil dont les critères d'évaluation portent l'empreinte des recommandations Cochrane est utilisé pour l'évaluation de RCT dans le développement de ENP.

Platz propose d'évaluer, suite à l'évaluation des 14 questions, la validité globale de l'étude : très bonne validité (++), bonne validité (+), validité insuffisante (-), faible validité (--) (Platz, 2021). Dans le développement de ENP, l'évaluation globale d'une publication utilisée est réalisée de façon similaire pour tous les designs de recherche et types de publication – mais sur la base du système GRADE (cf. chapitre 2.2.3).

2.2.1.4 Travaux de recherche qualitatifs – systématique d'évaluation sur le modèle de la JBI Critical Appraisal Checklist for Qualitative Research

Dans la recherche en soins basée sur des preuves, les designs de recherche qualitatifs s'imposent de plus en plus. Cela s'explique par le besoin de décrire et comprendre la complexité des situations, expériences, perspectives et cadres de vie des bénéficiaires de soins, de leurs familles et des professionnels dans le contexte de différents phénomènes liés à la santé et la maladie. La recherche qualitative peut donner un aperçu exploratoire guidé par les questions du « comment » ou « pourquoi » (Fringer & Schrems, 2018 ; Williams, Boylan, & Nunan, 2019). En fonction de la question posée et de l'objectif poursuivi, la recherche qualitative dispose d'un large éventail de designs de recherche : phénoménologique, ethnographique ou la méthode de la théorie ancrée, pour en nommer que quelques-uns (Behrens & Langer, 2016). Pour le développement de ENP, il est indispensable de vérifier et d'évaluer les études qualitatives quant à leur qualité et leur aptitude dans le contexte de différentes lignes directrices de pratique. Quand il s'agit de juger la crédibilité d'une étude et de ses résultats, le respect des critères essentiels des designs de recherche qualitatifs est indispensable (Moorley & Cathala, 2019 ; Noble & Smith, 2015 ; Williams et al., 2020). Face à une multitude d'outils ou check-lists disponibles, il n'existe pas de consensus international sur l'outil le plus apte à évaluer les études qualitatives (Hannes, 2011 ; Munthe-Kaas et al., 2019). Un outil établi au niveau international est la Critical Appraisal Checklist for Qualitative Research, proposée par des collaborateurs de l'institut Joanna Briggs et approuvée par le comité scientifique de l'institut après procédure de peer review (Lockwood, Munn, & Porritt, 2015 ; Williams et al., 2020). La checklist sert à évaluer l'ensemble des études qualitatives, sans prendre en compte d'éventuelles particularités de différentes approches qualitatives. L'outil de critical appraisal avec aides à l'interprétation est accessible via une présence du JBI sur internet prévue à cet effet : <https://jbi.global/critical-appraisal-tools> (consulté le 13/07/2022).

Le besoin de développer des outils, spécifiques aux méthodes et scientifiquement examinés, pour évaluer des études qualitatives, se situe dans le contexte de la recherche sanitaire basée sur des preuves (Williams et al., 2020). Jusqu'ici, l'évaluation critique des études qualitatives dans le développement de ENP n'est pas spécifique à la méthode de l'étude qualitative en question, mais les progrès qui se font dans ce domaine

sont continuellement observés pour être en mesure d'apporter, le cas échéant, des ajouts aux outils d'évaluation.

2.2.1.5 Lignes directrices et recommandations pour la pratique clinique (RPC) – systématique d'évaluation sur le modèle de l'outil AGREE II

La qualité méthodique du processus d'élaboration des lignes directrices manque souvent de transparence, et l'interprétation des énoncés et recommandations y contenues n'est pas toujours claire. Il est donc peu surprenant qu'un éventail d'outils a été conçu pour l'évaluation des lignes directrices et des RPC, tant pour la lecture critique que pour la rédaction des résultats. Un exemple pour un outil de rapport de lignes directrices est la checklist RIGHT (Reporting Items for practice Guidelines in HealThcare), conçue par une équipe internationale multidisciplinaire avec des décideurs politiques dans le cadre d'une démarche méthodique à quatre étapes pour améliorer la qualité de publication des lignes directrices (Chen et al., 2017).

Un exemple appliqué dans l'évaluation critique de lignes directrices est l'outil AGREE (Appraisal of Guidelines for Research and Evaluation), établi au niveau international et validé (Burgers et al., 2004 ; Chiappini et al., 2017; MacDermid et al., 2005 ; Werner et al., 2016). En 2009, une version révisée (AGREE II) de l'outil a été publiée (Browman et al., 2017). Cette mise à jour se distingue entre autres par une échelle de notation à 7 degrés, allant de « fortement en désaccord » (1) à « fortement en accord » (7), qui remplace les réponses oui/non de la première proposition. En plus, une base de calcul a été développée qui permet de comparer les lignes directrices évaluées en termes de qualité. Toutes les modifications réalisées par rapport à la première version de la grille AGREE, des aides à l'interprétation et des explications sur les 6 domaines et 23 items de la version AGREE II se trouvent dans le manuel d'utilisation (Browman et al., 2017). Pour la sphère germanophone, il existe outre la traduction de la grille originale (The AGREE Collaboration (Ed.), 2001) une version adaptée et complétée de l'outil AGREE, appelée DELBI (Deutsches Instrument zur methodischen Leitlinien-Bewertung – outil allemand pour l'évaluation méthodique de lignes directrices) (Kopp et al., 2008 ; Thole et al., 2007). Dès sa première publication en 2005, l'outil DELBI comprend un et plus tard (en 2008) deux domaines supplémentaires ; et à la différence des réponses oui/non de la première évaluation AGREE, il propose une évaluation des items à l'aide d'une échelle de 1 (n'est pas vrai du tout) à 4 (est entièrement vrai). Les deux domaines ajoutés focalisent, d'un côté, l'applicabilité de la ligne directrice à évaluer dans le système de santé allemand (domaine 7), et de l'autre l'exactitude méthodique du développement de lignes directrices lors de l'utilisation des lignes directrices existantes (domaine 8) ; les éléments des autres domaines correspondent essentiellement aux items AGREE. Au développement du DELBI ont contribué les expériences faites par le groupe de travail des sociétés scientifiques et médicales (AWMF), par le Centre médical pour la qualité en médecine (ÄZQ) et leurs partenaires, en collaboration avec les développeurs du AGREE et le réseau international pour le développement de lignes directrices (Kopp et al., 2008).

En plus, la version plus récente, AGREE II, est disponible en traduction allemande qui correspond en termes de contenu et d'utilisation à l'original anglophone (The AGREE Next Steps Consortium (Ed.), 2013). Il semble qu'une mise à jour du DELBI a été mis sur les rails : DELBI 2.0 qui est basé sur une variante du AGREE II adaptée à la sphère germanophone. Apparemment, DELBI 2.0 n'a pas réussi à sortir de l'état de version de consultation (Dippmann et al., 2013) et n'a jusqu'ici pas été publiée officiellement⁵. La quête d'une justification ou d'une explication de la situation est restée infructueuse. Cela est d'autant plus regrettable et étonnant que, selon la version la plus récente (mars 2022), la validité du premier DELBI expire au milieu de l'année 2015.

« Le DELBI (outil allemand pour l'évaluation méthodique de lignes directrices) a été publié le 7 juin 2005 (version 2005/2006), complété le 22 septembre 2008 (version 2005/2006 + domaine 8) et reste valable jusqu'à sa

⁵ Même la présence sur internet officielle du DELBI n'affiche uniquement la version de 2008 complétée (avec 8 domaines) de 2008, aucune trace du DELBI 2.0 n'est trouvable (cf. <https://www.leitlinien.de/hintergrund/leitliniengrundlagen#delbi> – dernier accès le 08/03/2022).

prochaine révision ou au plus tard jusqu'au milieu de l'année 2013. Un processus de mise à jour globale a été initié en conséquence. Après vérification par les éditeurs, la validité a été prolongée pour le moment jusqu'au milieu de l'année 2015 » (Kopp et al., 2008, p. 4).

Face à la situation exposée, notamment l'état incertain de la version actuelle de l'outil DELBI, l'équipe ENP utilise pour la lecture critique de lignes directrices et recommandations pour la pratique clinique une version légèrement adaptée de la version allemande de la grille AGREE II.

Pour AGREE II, tout comme pour nombre d'autres outils d'évaluation, il y a des aides à l'interprétation et des explications relatives aux 23 items ; la description de l'application en fournit des détails (The AGREE Next Steps Consortium (Ed.), 2013).

Selon le manuel, tous les items du AGREE II sont évalués sur une échelle de Likert à 7 niveaux (1 : fortement en désaccord – 7 : fortement en accord). La procédure d'évaluation prévoit la détermination d'un score pour chacun des six domaines. Ces scores des domaines sont indépendants et ne devraient pas être regroupés en un score global de qualité pour la ligne directrice évaluée. Pour ce faire, les scores accordés à chaque élément d'un domaine sont additionnés, puis multipliés par le nombre d'évaluateurs pour traduire le résultat en pourcentage du score maximal pour ce domaine (multiplié, là aussi, par le nombre d'évaluateurs). Cette façon de procéder veut permettre la comparaison de lignes directrices et faciliter ainsi la prise de décision quant à la recommandation d'une ligne directrice. Le consortium AGREE renonce consciemment à indiquer une valeur « cut off » pour distinguer les « bons » ou « mauvais » scores des domaines ; les évaluateurs doivent prendre cette décision plutôt dans le contexte de la ligne directrice concernée (The AGREE Next Steps Consortium (Ed.), 2013). L'évaluation globale de la ligne directrice se réalise à la suite des évaluations de 6 domaines/23 éléments avec deux autres critères de conclusion. Il s'agit, d'un côté, de juger la qualité globale de la ligne directrice, toujours sur une échelle à 7 niveaux (1 : de très mauvaise qualité – 7 : d'excellente qualité) et de l'autre, de donner une recommandation quant à l'utilisation de la ligne directrice (oui – oui, avec certaines modifications – non).

Généralement, ce schéma d'évaluation doit rester possible dans le développement de ENP ; c'est pourquoi le score individuel de chaque élément devrait toujours être noté. Dans le développement de ENP, l'évaluation synthétique d'une publication est réalisée avec une méthode d'après GRADE (cf. chapitre 2.2.3) et lors de l'évaluation d'une publication par plusieurs évaluateurs, la fiabilité inter-évaluateurs est calculée sous forme du pourcentage de conformité ; pour cette raison, on n'applique pas les deux critères d'évaluation globale de la grille AGREE II. Pour l'usage interne de la grille AGREE II, cherchant à déterminer la fiabilité inter-évaluateurs, on a regroupé les 7 niveaux de réponse proposés pour évaluer chaque élément, additionnellement en trois catégories : 1–3 (plutôt désaccord), 4 (neutre) et 5–7 (plutôt accord).

2.2.1.6 Travaux de recherche par les méthodes mixtes – systématique d'évaluation (par extraits) du MMAT (Mixed Methods Appraisal Tool)

Le terme « méthodes mixtes » désigne généralement la combinaison de méthodes de recherche quantitatives et qualitatives dans le même design d'étude. En réalité, la recherche méthodes mixtes va beaucoup plus loin que d'autres approches alliant les deux conceptions de recherche, comme par ex. la triangulation ou la combinaison de méthodes (Fringer, 2018 ; Kelle, 2019). Une caractéristique essentielle des méthodes mixtes est l'intégration et la synthèse systématique d'éléments qualitatifs et quantitatifs – tant pour la saisie que pour l'analyse des données et pour le rapport des résultats.

A la différence des disciplines comme la recherche sociale ou éducative où la combinaison de méthodes quantitatives et qualitatives dans une même et seule étude est de pratique courante depuis longtemps, les sciences sanitaires et particulièrement la science soignante commencent tout juste à se pencher vers les méthodes mixtes (Kelle, 2019 ; Knappertsbusch, Langfeldt, & Kelle, 2021 ; Niederberger & Peter, 2018). Et pourtant, notamment en matière de (recherche en) soins – agissant dans ce système complexe qui est la prise en charge sanitaire, interdisciplinaire et transsectoriel – cette approche paraît particulièrement

importante puisque la prise en charge de personnes dépendantes applique un large éventail de perspectives diagnostiques et de concepts d'interventions. Les contextes de prise en charge soignante se caractérisent souvent par une complexité accrue, avec plusieurs phénomènes soignants qui peuvent s'impacter mutuellement. Les approches mono-méthodiques qui essaient de saisir une telle thématique avec des méthodes purement quantitatives ou qualitatives ne sont pas toujours à l'échelle de la multitude d'aspects à considérer (Quasdorf & Holle, 2018). La polémique qui règne entre les adhérents des deux traditions, souvent marquée par une dépréciation réciproque de la validité et la scientificité des méthodes, complique inutilement la saisie de la prise en charge soignante, déjà assez compliquée en soi (Kelle, 2019 ; Quasdorf & Holle, 2018). Selon Niederberger & Peter (2018), les études de méthodes mixtes représentent une approche prometteuse : leur potentiel de connaissances, c'est tant la saisie des causalités que la compréhension intersubjective de nombreux (nouveaux) phénomènes et elles permettent en plus une vision de la pratique (soignante) orientée vers la personne et basée sur des preuves.

Cette augmentation de l'importance et de l'utilisation des méthodes mixtes dans la recherche en santé et en soins entraîne une montée du besoin en outils aptes à évaluer les travaux de recherche correspondants. La quête d'outils appropriés pour la lecture critique des études de méthodes mixtes a fait émerger une série de différentes approches. A titre d'exemple, citons l'instrument « Quality Assessment with Diverse Studies » (QuADS) (Harrison et al., 2021) au champ d'application assez vaste en vue des articles à évaluer, avec des critères pensés pour des travaux tant qualitatifs que quantitatifs. Pour évaluer des études qui se servent d'une combinaison de méthodes ou de l'approche des méthodes mixtes, il est recommandé d'appliquer tous les critères proposés par l'instrument. Faute de critères d'évaluation explicite visant la caractéristique centrale de la recherche des méthodes mixtes – à savoir l'intégration et la synthèse systématiques d'éléments quantitatifs et qualitatifs dans plusieurs ou toutes les phases essentielles du processus de recherche – on n'a pas retenu l'instrument QuADS pour l'utilisation dans le développement de ENP.

Très focalisés sur la recherche des méthodes mixtes sont par contre les outils « Mixed Methods Impact Evaluation Tool » (MMIE) et « Mixed Methods Systematic Reviews Appraisal Tool » (MMSR) (Jimenez et al., 2018a, 2018b). Ils ciblent spécifiquement les travaux de méthodes mixtes visant à évaluer l'effet des interventions et des revues systématiques qui se servent de méthodes mixtes pour obtenir une synthèse des évidences quant à une question déterminée. Les deux outils prennent en compte le facteur intégratif particulier des aspects quantitatifs et qualitatifs de la recherche des méthodes mixtes avec des critères d'évaluation spécifiques ; mais jusqu'ici, la qualité du MMIE et du MMSR n'est pas encore suffisamment examinée. Dû à ce fait et à la focalisation trop étroite sur certains designs de recherche des méthodes mixtes, on a décidé de ne pas appliquer, en ce moment, ces deux outils d'évaluation.

Finalement, le choix pour la lecture critique des études de méthodes mixtes dans le développement de ENP a été porté sur l'outil « Mixed Methods Appraisal Tool » (MMAT) (Hong et al., 2018; Pluye et al., 2009; Taylor & Hignett, 2014). Développé initialement en 2006, la révision actuelle du MMAT a été publiée en 2018. L'outil juge les aspects quantitatifs et qualitatifs d'une étude de méthodes mixtes dans 5 catégories à 5 critères chacune. Les critères visent la qualité des designs de recherche suivants : recherche qualitative, RCT, études non randomisées à orientation quantitative, études descriptives quantitatives, études de méthodes mixtes proprement dites. Ces cinq catégories sont précédées de deux critères d'évaluation généraux qui s'appliquent à tous les designs de recherche. Tous les critères sont répondus par « oui/non/ne sait pas » avec la possibilité de laisser un commentaire pour chaque critère d'évaluation. Il est fortement déconseillé de calculer un score global à base de l'évaluation des différents critères. Au lieu de cela, il est conseillé de préciser les évaluations attribuées aux différents critères. Le MMAT vise, en premier lieu, les études empiriques et ne pas la recherche théorique ou les revues. L'outil devrait être appliqué par au moins deux personnes indépendantes.

L'utilisation du MMAT est un processus à deux étapes. Avant de procéder à l'évaluation des critères, on identifie d'abord à l'aide d'un diagramme de flux la catégorie d'étude adéquate à utiliser avec le travail de recherche donné.

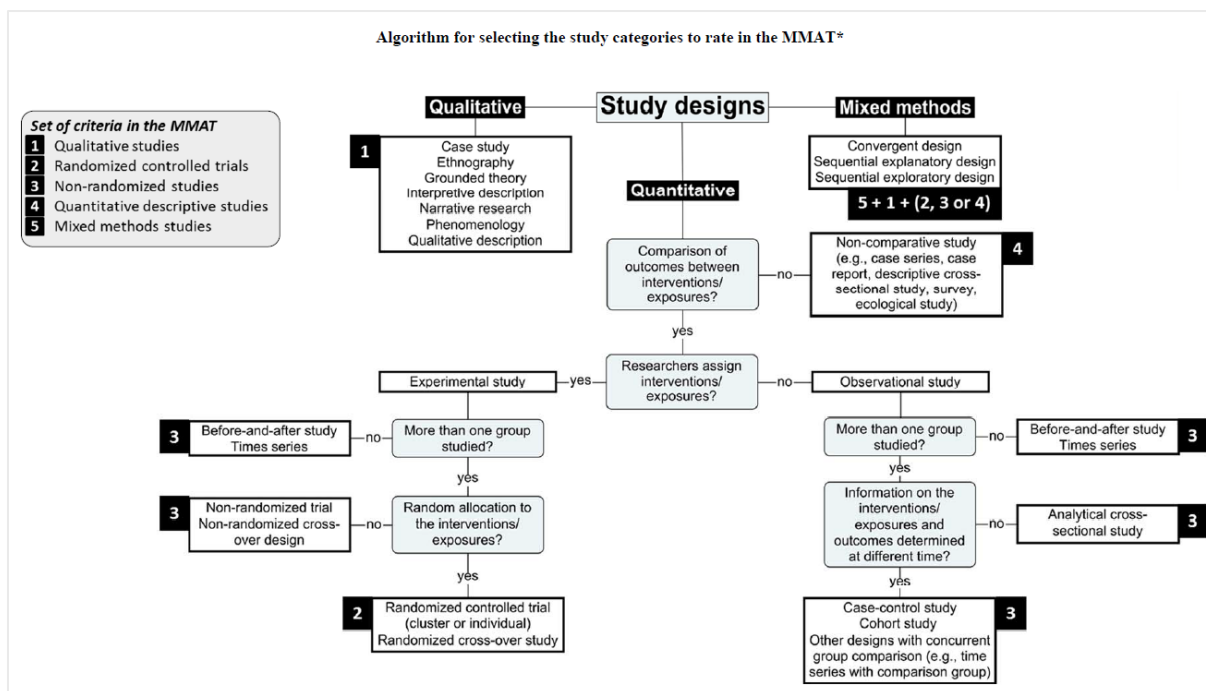


Figure 8 : Diagramme de flux pour sélectionner la catégorie d'étude et utiliser adéquatement le MMAT (source : Hong et al., 2018, p. 8)

Si le travail concerné implique par exemple, outre une approche qualitative, une partie quantitative sous forme d'une étude (partielle) randomisée contrôlée, l'évaluation se réaliserait à base des catégories 1, 2 et 5 du MMAT afin de juger, d'un côté, les éléments quantitatifs et qualitatifs tout en réfléchissant de manière critique les particularités de l'approche de méthodes mixtes. Pour toutes les catégories et critères d'évaluation, des explications et aides à l'interprétation sont mises à disposition qui précisent l'importance et l'intention de chaque élément. Outre l'original anglophone, le MMAT est également disponible en version française, brésilienne, turque et chinoise – une version officielle allemande n'existe pas encore. Quant à la qualité du MMAT, jusqu'ici on en a vérifié l'utilité, la fiabilité, l'efficacité, la validité de contenu ainsi que la fiabilité inter-juges dans des études respectives⁶.

Pour la lecture critique des travaux de méthodes mixtes pertinents pour le développement de ENP, le MMAT n'est appliqué que par extraits ; la version originale de l'outil intégral peut être consultée dans la publication correspondante (Hong et al., 2018). C'est, plus précisément, la cinquième catégorie de l'outil qui s'applique car elle cible explicitement les particularités de ce design de recherche intégratif. Pour l'évaluation des autres catégories du MMAT (recherche qualitative, RCT, études non randomisées à orientation quantitative, études descriptives quantitatives) il existe d'autres outils, en partie plus différenciés et mieux vérifiés, dont les descriptions se trouvent dans les chapitres précédents. La combinaison de la cinquième catégorie du MMAT avec les barèmes d'évaluation spécifiques pour les recherches quantitatives et qualitatives laisse espérer une évaluation plus expressive encore des travaux de méthodes mixtes.

2.2.1.7 Livres spécialisés/manuels – réalisation individuelle des critères d'évaluation de la bibliothèque de l'université Cornell à Ithaca, New York

En cas de besoin, le développement de ENP a recours à des livres spécialisés ou des manuels en soins ou autres professions avoisinantes, surtout si la recherche sur une thématique précise s'avère être encore largement inexploitée. Si une publication spécialisée en (chapitres d'un) livre représente une forme de

⁶Une vue d'ensemble des publications correspondantes est disponible en ligne sous <http://mixedmethodsappraisaltoolpublic.pbworks.com/w/page/127425458/Publications%20on%20the%20MMAT> (consulté le 23/02/2022).

publier des études ou des travaux de recherche, ou si une synthèse basée sur la littérature a été établie sur une thématique précise, les critères d'évaluation décrites ci-dessus s'utilisent en fonction du design de recherche. Tous les autres cas requièrent une évaluation spécifique du type de publication « livre spécialisé/manuel » (ouvrage intégral ou chapitres individuels). La quête d'outils d'évaluation appropriés dans le contexte de la prise en charge soignante ou sanitaire pour cette fin spécifique n'a mis au jour que peu de résultats qui suivent deux approches différentes. Il s'agit, d'une part, de la « Checklist for Text and Opinion » de la série des lectures critiques de l'institut Joanna Briggs (Joanna Briggs Institute (Ed.), 2017). La check-list comprend un total de six critères à répondre par « oui », « non » ou « imprécis ». Finalement, une évaluation globale est réalisée sur l'inclusion ou l'exclusion de la publication ou le besoin de se procurer de plus amples informations. Même si la check-list, soumise à un processus de développement et évaluée à plusieurs reprises, est selon sa définition applicable à l'évidence basée sur des textes et opinions, issue d'opinions d'experts, consensus, commentaires, hypothèses et affirmations et publiés dans des revues spécialisées, magazines, monographies et rapports (Aromataris & Munn, 2020 ; McArthur et al., 2015), on n'a pas retenu cet outil pour l'évaluation de livres spécialisés et manuels. Cette décision s'explique par le caractère particulier des livres spécialisés et manuels : généralement, ils n'essaient pas de faire passer des opinions, consensus ou commentaires mais ils transmettent un corpus de connaissances (non exclusivement scientifiques) préparées pour la pratique et la formation sur un ou plusieurs sujets. Par conséquent, les éléments de la « Checklist for Text and Opinion » ne sont pas entièrement appropriés pour évaluer des livres spécialisés et manuels classiques⁷.

L'outil qu'on a finalement retenu est construit à base d'un catalogue de critères généraux pour évaluer des sources d'information de la bibliothèque de l'université Cornell à Ithaca, New York⁸, et pensé spécifiquement pour l'évaluation de (chapitres de) livres (Datta, Funnell, & Ramuscak, 2012). Selon nos recherches, cet outil ne présente malheureusement ni historique de développement systématique ni contrôles de qualité. De toute façon, les (chapitres de) livres auxquels ne s'applique aucun des autres outils spécifiques présentés ci-dessus correspondent à un niveau d'évidence très faible – indépendamment du résultat d'évaluation obtenu ; pour cette raison, l'adéquation des critères d'évaluation pour des livres spécialisés/manuels prévaut dans ce cas sur le niveau de développement scientifique. C'est aussi la justification pourquoi la réalisation des critères d'évaluation présentée par Datta et collègues (2012) s'est vue adaptée aux besoins du développement de ENP et réduite en ampleur. Pour l'usage interne, on a traduit les éléments de l'original anglophone vers l'allemand. Le résultat en sont dix critères à évaluer par « oui » « partiellement » ou « non » pour obtenir une évaluation synthétique (fort, moyen, faible).

2.2.1.8 Littérature grise et sites web – critères d'évaluation de la checklist AACODS (Authority, Accuracy, Coverage, Objectivity, Date, Significance)

Pour le développement ENP, les informations issues de sites web et de la littérature grise qui ne peuvent pas être évaluées avec les outils présentés jusqu'ici ne jouent qu'un rôle subordonné ou subsidiaire. La « littérature grise » désigne dans ce contexte toutes les publications imprimées ou électroniques produites en dehors de l'édition commerciale. Les publications qui entrent typiquement dans ce concept sont les mémoires (de fin) d'études, actes de congrès, écrits d'entreprise etc., mais aussi des présences/sites web génériques (Paez, 2017; Woods, Phillips, & Dudash, 2020). Si aucun autre outil plus spécifique n'est disponible pour l'évaluation de l'information concernée, la « AACODS checklist », conçue pour évaluer la littérature grise, sera utilisée pour toutes les publications auxquelles aucun autre outil ne s'applique. La checklist AACODS a été publiée en 2010 à l'université Flinders (Adelaïde, Australie) (Tyndall, 2008, 2010). Elle est composée de six catégories larges dont les abréviations donnent l'acronyme qui désigne l'outil :

⁷Un exemple pour illustrer le manque d'adéquation est l'élément 3 qui demande si les intérêts de la population concernée sont situés au centre de l'opinion.

⁸ Cf. https://guides.library.cornell.edu/critically_analyzing/home (consulté le 23/02/2022)

- Authority (Compétence)
- Accuracy (Exactitude)
- Coverage (Etendue)
- Objectivity (Objectivité)
- Date (Date)
- Significance (Portée)

Aujourd'hui, la grille AACODS s'utilise couramment pour évaluer la littérature grise, aussi pour établir des synthèses de connaissances dans le domaine de la santé (Vaska et al., 2019).

2.2.2 Synthèse d'évidence et évaluation d'éléments/concepts d'actions individuels

La synthèse d'évidence par rapport aux différents concepts d'actions d'un diagnostic soignant ENP est déterminée lors de l'évaluation de la littérature, mais elle ne fait actuellement pas partie des publications. Les concepts d'actions d'une ligne directrice de pratique représentent les options d'action soignantes préconisées pour un diagnostic soignant ; en ce moment, la priorisation des concepts d'action et/ou l'évidence relative à chaque concept d'action n'est pas précisée dans les publications. La documentation interne du développement ENP par contre prend bien compte du niveau d'évidence synthétisé des concepts. Mais dans la plupart des cas, l'état de la recherche soignante est toujours trop limité pour préciser l'évidence (externe) des concepts d'action liés aux diagnostics soignants. De nombreux concepts soignants se trouvent encore à un niveau de « bonne pratique ». D'un autre côté, l'indication du niveau d'évidence des concepts d'action risquerait d'exercer une influence unilatérale sur le processus de prise de décision des personnes soignantes, faisant la part entre l'évidence interne et externe. Cela est problématique dans la mesure où le choix d'interventions de soins, adéquat dans la situation individuelle d'une personne dépendante, peut diverger pour maintes raisons du choix le plus « évident ». ENP fait intentionnellement omission de concepts non confirmés dans des études ou publications spécialisées – même si les interventions correspondantes s'appliquent (toujours) dans la pratique soignante ou si des demandes d'intégration sont présentées à l'équipe ENP par des praticiens. Le but de la synthèse d'évidence et de l'évaluation d'éléments/concepts d'action individuels est d'assurer qu'il s'agit de « meilleures pratiques » ou de concepts/éléments basés sur des preuves dont la validité (au moins une évidence faible) a été confirmée. Quant à la synthèse d'évidence des caractéristiques et étiologies pour décrire un énoncé diagnostique, le degré d'évidence de chaque élément n'est pas non plus indiqué. Le développement de ENP se focalise sur une représentation la plus complète possible des liens et/ou d'apparitions confirmés par la littérature. Cela sert à atteindre « l'intégralité » requise pour représenter un lien soignant et pour profiter ultérieurement de ces données comme base pour la recherche soignante. La détermination du degré adéquat de granularité (degré de spécification/abstraction) est un défi majeur dans ce contexte – entre autres pour éviter les chevauchements de contenu.

En résumé : Les interventions de soins et/ou caractéristiques/facteurs de risque/étiologies pour lesquelles une taille d'effet négative et/ou une importance négative est prouvée dans le contexte de la ligne directrice de pratique ENP ne sont systématiquement pas admises. Les éléments qui n'ont pas pu être étayés par des références fiables sont également exclus.

2.2.3 Tableau d'évidence pour évaluer des études dans le contexte du développement ENP

Pour documenter les résultats de la lecture critique (critical appraisal) pendant le développement de ENP, on utilise un tableau dit d'évidence. Ce tableau assure la documentation de la qualité des publications utilisées pour le développement de ENP. Tant pour l'équipe ENP que pour des tiers intéressés il est important de pouvoir consulter des résumés des évidences – et cela contribue en plus à augmenter la

transparence du processus de développement de ENP. L'équipe ENP a conçu le nouveau tableau d'évidence par analogie approximative au développement de lignes directrices réalisé par la Société allemande de neurologie (DGN) et la Société allemande de neuroréhabilitation (DGNR) (Platz, 2021). Comme expliqué précédemment, un critical appraisal d'une publication donnée est réalisé en fonction du design de l'étude ; l'évaluation et les résultats obtenus sont ensuite documentés sous forme condensée dans le tableau d'évidence. L'appréciation globale des aspects évalués y est également documentée, de même que les résultats essentiels et des renseignements liés à l'étude. Par la suite la structure de base du tableau sera présentée.

2.2.3.1 Structure du tableau d'évidence

Le tableau d'évidence conçu pour le perfectionnement de ENP repose sur un document Microsoft Excel aux structures et fonctions bien définies. Le tableau retient les informations et résultats essentiels concernant les études et publications consultées, l'évaluation de leur qualité (évaluation de l'évidence) ainsi que les conclusions par rapport à ENP. Une fonction de filtrage spécifique adaptée aux informations et données saisies permet d'obtenir un aperçu rapide quant à des aspects spécifiques – même si le tableau contient des dizaines de références. La figure ci-après présente les éléments essentiels du tableau d'évidence qui sont ensuite expliqués brièvement.

Figure 9 : Les contenus du tableau d'évidence conçu pour l'évaluation et l'analyse de publications

Les champs dans la partie supérieure du tableau représentent les fonctions de filtrage qui permettent d'afficher ou de masquer de manière ciblée certains contenus afin de regrouper rapidement des informations ou données en fonction de la question posée.

Les colonnes suivantes contiennent les contenus promement dit :

- **N° de réf. :** Numéro de travail de la publication, commençant par 1, ; ce numéro courant constitue une référence facilement maniable dans le cadre du travail de développement de ENP ou dans autres documents.
- **Auteur(s) :** Les noms des auteurs-trices ou éditeurs-trices des publications. S'il y a plus de trois auteurs-trices, seul le-la premier-ère auteur-trice est mentionné, les autres personnes contributantes sont indiquées par « et al. » (selon la 7e édition du manuel de publication de l'Association américaine de psychologie APA).
- **Année :** L'année de parution de la publication/étude.
- **Type d'étude :** Informations concises sur le design de recherche de la publication ou étude.
- La colonne **population, nombre de participants/d`études :** informations sur les données clés de l'étude ou de la publication en fonction du type d'étude. Pour les revues, il s'agit par exemple du nombre d'études individuelles incluses ; dans le cas de traveaux empiriques, on y trouve les nombres de participants et/ou les tailles des groupes.
- Les colonnes **Evaluation de l'évidence (évaluateur 1 et évaluateur 2)** saisissent l'évaluation globale du niveau de preuve de la publication/étude en tenant compte de tous les critères d'évaluation (et, le cas échéant, d'augmentation ou de dégradation) décrits dans les chapitres

suivants 2.2.3.2 et 2.2.3.3. Dans la mesure du possible, une deuxième personne assure une évaluation indépendante de l'évidence pour au moins 10 % de toutes les publications. Si la concordance entre les résultats d'évaluation obtenus par les deux personnes est de 80 % ou plus, les évaluations ultérieures seront réalisées uniquement par la personne responsable principale de la gestion du sujet, concerté auparavant en équipe. Si le taux de concordance est inférieur, l'évaluation sera poursuivie par deux personnes.

- **Problématique, méthode, mesure de l'outcome et résultats principaux :** Cette colonne est prévue pour de brèves informations essentielles portant sur la question clé de recherche de la publication et sur l'accent thématique. Il convient en plus de préciser de quelle façon et à quel moment (par ex. follow-up) les résultats du travail ont été saisis ou mesurés (méthode). Les résultats principaux de la publication ou de l'étude sont également synthétisés. En fonction du type d'étude, de brèves informations sont fournies sur les tailles d'effet (par ex. valeur p) ou d'autres résultats statistiques.
- **Conclusion et pertinence ENP :** Cette colonne sert à formuler les conclusions essentielles des résultats quant à l'intégration/modification dans ENP, visant le transfert des connaissances obtenues vers des items ENP adéquats.
- **Type d'item ENP :** Indication du type d'item ENP à ajouter, modifier ou mettre en sommeil par suite de la publication (caractéristiques, étiologies, ressources, objectifs, interventions de soins etc.) Cette information permet le filtrage du tableau d'évidence pour afficher les publications en fonction du type d'item en question.
- **Source (formatée à partir de EndNote) :** Cette colonne est réservée à la citation complète de la publication ou étude dans le style de citation correct issu du programme de gestion bibliographique EndNote.

Un tableau d'évidence en cours d'élaboration et d'évaluation, et donc pas encore entièrement rempli, est présenté à titre d'exemple dans l'annexe I.

Pour la transformation des résultats individuels obtenus par les systèmes d'évaluation (cf. chapitre 2.2.1) en une évaluation globale du degré d'évidence et de la fiabilité d'une publication, il est primordial de considérer l'importance et l'impact de questions/critères particuliers sur la fiabilité des propos soignants. Si un aspect critique qui a très probablement un impact sur la validité des énoncés est qualifié de « non accompli », cela est autrement plus grave qu'un critère « non accompli » dont l'impact sur la validité globale des résultats de l'étude est limité.

2.2.3.2 Réflexions préliminaires sur les méthodes de classement dans l'évaluation de l'évidence

Au cours des dernières décennies, une série de systématiques de classement d'évidence s'est formée qui s'est vue exposée, à plusieurs reprises, à différents types de critique. Ainsi, il est connu dans la recherche de soins que l'évidence des essais randomisés contrôlés – considérés en médecine souvent l'étalon-or en matière d'évidence – n'est pas forcément supérieure au degré d'évidence des études observationnelles (Behrens & Langer, 2004 ; Evans, 2003 ; Grypdonck, 2004 ; Lavin et al., 2002). Il existe des questions soignantes qui ne se prêtent pas aux preuves d'évidence par RCT ou bien les RCT ne sont pas justifiables pour des raisons éthiques (Roberts & DiCenso, 1999). On reconnaît de plus en plus que les études observationnelles et les séries de cas peuvent contribuer parfaitement à obtenir des évidences au niveau des interventions et que les travaux qualitatifs ou les travaux de méthodes mixtes peuvent, eux aussi, fournir des preuves précieuses pour le développement de diagnostics soignants. Le design d'étude qui est le mieux à même de répondre à une question précise du développement de ENP dépend toujours du contexte donné (Petticrew & Roberts, 2003). Quant aux phénomènes soignants tels que la souffrance liée au deuil ou à la solitude, les études qualitatives de la phénoménologie ou les études observationnelles se prêtent particulièrement à la déduction de caractéristiques du monde vécu par les personnes concernées pour décrire un concept soignant (Panke-Kochinke, 2012 ; Ploeg, 1999). En résumé, pour élucider des

phénomènes diagnostiques, les études observationnelles, études qualitatives, analyses de concepts etc. s'imposent tandis que les essais randomisés contrôlés (RCT) sont à privilégier pour la recherche de concepts thérapeutiques utiles. Lorsqu'il s'agit de liens de causalité ou d'effets par ex. de facteurs environnementaux sur notre santé, les études de cohorte ou des études cas-témoins peuvent s'avérer appropriées. Selon le sujet traité, les analyses synoptiques systématiques ou les méta-analyses évaluent sommairement plusieurs travaux de recherche pour répondre à une question de recherche. Si le choix et la qualité des études primaires incluses sont bons, les résultats des méta-analyses et des travaux synoptiques présentent, en règle générale, le niveau d'évidence le plus élevé. Le tableau suivant résume les designs d'études essentiels avec leurs caractéristiques (sans prétendre être exhaustif).

Designs d'étude	
Perspective/caractéristiques	Description, particularités
Perspective temporelle	Rétrospective : par ex. analyse ultérieure de données issues de dossiers patients, systèmes d'information hospitaliers, documentations du processus de soins etc. Prospective : saisie de données permanente à partir du début de l'étude pendant une certaine période, méthode définie, ensuite analyse
Durée d'observation	Etude transversale : « vue instantanée », par ex. mesure ou interrogation unique Etude longitudinale : observation au fil du temps ou plusieurs mesures à des écarts temporels
Groupe de contrôle	Sans groupe de contrôle (par ex. séries de cas, observations d'application, interrogations, analyses de documents) Avec groupe de contrôle (par ex. RCT)
Triangulation	L'utilisation de plusieurs méthodes dans un travail de recherche pour combiner différentes perspectives. Il peut s'agir d'études qualitatives ou quantitatives.
Les études analytiques cherchent à identifier des principes causes-effets par ex. le rapport entre des facteurs de risque déterminés et un phénomène de soins.	Etude de cohorte (cohort study) : suivie des participants en vue de critères déterminés dans un échantillon sélectionné pendant une période prolongée. Etude cas-témoin (case control) : comparaison de bénéficiaires de soins avec ou sans un phénomène de soins spécifique en vue de par ex. grandeurs d'influence, facteurs de risque, caractéristiques etc.
Les études descriptives sont des études purement descriptives : les données sont présentées et analysées sans chercher d'expliquer des causes ou liens étiologiques.	Etude observationnelle : saisie d'un(e) événement/résultat/apparition pertinent(e) en soins à un moment donné ; sans preuve de causalité. Une étude observationnelle est un type d'étude épidémiologique et peut prendre la forme par ex. d'une étude de cohorte, cas-témoin ou d'une étude transversale (Cuschieri, 2019b). Etude épidémiologique : analyse par ex. de facteurs de risque et leur répartition dans la population, le plus souvent des études observationnelles Etudes de cas particuliers ; études de cas (case reports) ; séries de cas (case series) Etudes de corrélation : appropriées pour former des hypothèses. On vérifie à base de données s'il existe par ex. un lien entre des facteurs observés déterminés et un diagnostic soignant. Etude transversale, par ex. étude de prévalence/d'incidence : par ex. portant sur l'apparition d'événements indésirables dans un établissement ou l'apparition de phénomènes de soins dans une population
Les études expérimentales englobent une série de designs de recherche pour lesquels les scientifiques manipulent une ou plusieurs variables pour contrôler et mesurer toute altération	Essai d'intervention randomisé, contrôlé (RCT) : en plus, répartition aléatoire des participants dans les groupes ; en (double) aveugle randomisé contrôlé : en plus mise en aveugle des participants (et des intervenants) Etude quasi-expérimentale : La conception est expérimentale, mais il manque des caractéristiques essentielles telles que la répartition aléatoire des participants ou le groupe de contrôle. Il n'y a par ex. pas de sélection aléatoire des

provoquée au niveau d'autres variables. Cela permet d'analyser dans quelle mesure un facteur déterminé (variable) agit sur une situation/un comportement ou un état.	participants, au moins une variable indépendante est analysée activement par la conception expérimentale du design d'étude. Essai d'intervention croisé/essai croisé : les groupes d'intervention et de contrôle sont changés au milieu de l'essai Etude d'intervention non randomisée contrôlée : les participants sont assignés à différentes interventions, puis contrôlés : l'étude détermine l'intervention, il existe un groupe de contrôle Etude avant-après : Les participants d'une étude sont examinés par ex. avant et après une intervention. Etude multicentrique : Il s'agit d'études cliniques qui sont réalisées en même temps dans plusieurs établissements avec un design d'étude multicentrique et plusieurs équipes de recherche.
Designs d'étude qualitatifs	Analyse qualitative du contenu : Repose sur l'analyse de données textuelles, par ex. entretiens, documentations, matériel vidéo avec du langage parlé etc. Etudes interprétatives, par ex. analyses de communication, analyses narratives, analyses documentaires Théorie ancrée (grounded theory) : méthode de recherche qualitative avec la particularité d'alterner constamment entre la collecte de l'analyse des données Les scientifiques identifient, affinent et intègrent des catégories des données obtenues afin d'en déduire et/ou confirmer une théorie. Méthode Delphi : une démarche pour recueillir des prévisions à base d'opinions d'experts et/ou pour exprimer l'accord ou désaccord qui existe entre des opinions d'experts. Un procédé de communication de groupe très structuré permet d'obtenir une solution/un accord portant sur une problématique/question complexe à partir des contributions individuelles. Design d'étude phénoménologique : Il s'agit d'études qui décrivent des phénomènes soignants spécifiques tels que le deuil, la solitude, le désespoir (Mehrholtz, 2010). Les études phénoménologiques sont particulièrement importantes dans le contexte du diagnostic soignant. Ce type d'étude compte parmi les approches descriptives. Analyse de concepts (concept analysis) : très utile pour saisir et décrire avec précision des phénomènes soignants observés dans la pratique des soins
Résumé d'études individuelles	Revue/travaux de synthèse : Analyse de la littérature pertinente actuelle avec pour but d'évaluer de manière synthétique des études similaires portant sur la même problématique. Une revue « systématique » suivant un standard fixe compare, dans l'idéal, des études ayant le même point final et un niveau élevé, par ex. des RCT, pour arriver à un constat d'évidence. Méta-analyse : Revue systématique qui résume les résultats des études individuelles avec des méthodes statistiques. Les études primaires devraient être le plus similaires possibles quant à la population étudiée et toutes les études devraient avoir, dans l'idéal, le même point final primaire. Etudes multicentriques : Etudes cliniques qui sont réalisées parallèlement dans plusieurs établissements et dont l'expressivité scientifique augmente avec la participation de différent-e-s intervenant-e-s.

Tableau 10 : Résumé des designs d'étude les plus importants avec leurs caractéristiques centrales (source : présentation originale, complétée d'après Rojahn, 2016, p. 17). Publications utilisées pour compléter le tableau : Bartholomeyczik et al., 2008 ; Behrens & Langer, 2004 ; Moorley & Cathala, 2019 ; Morse et al., 1996 ; Niederberger & Drejack, 2020 ; Panke-Kochinke, 2012 ; Roberts & DiCenso, 1999 ; Rojahn, 2016.

Différents designs d'étude entraînent différents risques d'erreur systématique ou d'effet de distorsion systématique (risk of bias). Etant donné qu'il existe de multiples sources d'erreurs dans la génération de connaissances, il n'est pas possible de définir une hiérarchie unidimensionnelle de types d'étude quant au niveau d'évidence⁹. Un risque de distorsion systématique réside par ex. dans la falsification des résultats

⁹ Level of Evidence (degré/niveau d'évidence) : preuve scientifique qui indique le degré de généralisabilité ou de validité de résultats de recherche selon un barème de critères (Bartholomeyczik et al., 2008).

d'études provoquée par une méthode inappropriée lors de la sélection des participants (biais de sélection). Les RCT essaient d'éviter cet effet de distorsion potentiel avec un choix aléatoire des sujets à étudier et avec la mise en aveugle tant des personnes participantes que des personnes intervenantes. S'il n'y avait que cette source d'erreurs, une détermination unidimensionnelle de la hiérarchie des évidences serait possible. Parmi les nombreuses possibilités d'erreurs systématiques qui risquent de corrompre les résultats d'une étude comptent par ex. le choix inapproprié des outils de mesure ou le manque de cadre théorique du travail de recherche. Il est donc peu surprenant que la « hiérarchie des évidences » classique soit controversée même en médecine (Petticrew & Roberts, 2003). Malgré cela, la hiérarchie des évidences est un outil fréquemment utilisé dans la prise de décision clinique pour déduire et évaluer la qualité et la fiabilité sur la base des designs d'étude. Le tableau ci-après présente la hiérarchie des évidences répandue qui se trouve appliquée, avec quelques adaptations, dans de nombreuses publications – surtout médicales.

Hiérarchie des évidences classique en médecine	
1	Revue systématique et méta-analyse
2	Essais randomisés contrôlés avec des résultats définitifs
3	Essais randomisés contrôlés avec des résultats imprécis
4	Etudes de cohorte
5	Etudes cas-témoins
6	Etudes transversales
7	Etudes de cas particuliers

Tableau 11 : Hiérarchie des évidences classique en médecine (source : présentation originale selon Petticrew & Roberts, 2003)

La comparaison de différentes hiérarchies d'évidence laisse apparaître des similitudes et manifeste en plus que, dans la plupart des cas, les designs de recherche qualitatifs ne sont pas représentés. La hiérarchie des évidences du Scottish Intercollegiate Guidelines Network (SIGN), par contre, implique déjà un degré de réalisation qualitative de l'étude.

La hiérarchie des évidences selon le système de gradation SIGN	
1++	Méta-analyses de qualité élevée, revues systématiques d'essais contrôlés randomisés ou essais contrôlés randomisés avec un risque de biais très faible
1+	Méta-analyses bien menées, revues systématiques ou essais contrôlés randomisés avec un risque de biais faible
1-	Méta-analyses, revues systématiques ou essais contrôlés randomisés avec un risque de biais élevé
2++	Revue systématique de qualité élevée d'études cas-témoins ou de cohorte Etudes cas-témoins ou de cohorte avec un très faible risque d'effet de facteurs de confusion ou de biais et une probabilité élevée que la relation est causale
2+	Etudes cas-témoins ou études de cohorte bien menées avec un faible risque d'effet de facteurs de confusion ou de biais et une probabilité modérée que la relation est causale
2-	Etudes cas-témoins ou de cohorte avec un risque élevé d'effet de facteurs de confusion ou de biais et un risque significatif que la relation ne soit pas causale
3	Etudes non analytiques, par ex. rapports de cas, séries de cas
4	Opinion d'experts

Tableau 12 : Hiérarchie des évidences selon le système de gradation SIGN (source : présentation originale d'après Harbour & Miller, 2001)

L'Oxford Center for Evidence-based Medicine propose d'orienter la hiérarchisation du niveau de preuve non sur le seul design d'étude mais sur la question de recherche sous-jacente (Durieux, Vandenput, & Pasleau, 2013 ; Oxford Centre for Evidence-Based Medicine, 2009). Aujourd'hui, la conviction se répand de plus en plus que les constats d'évidence en recherche interventionnelle ne doivent pas forcément s'appuyer sur des essais contrôlés randomisés pour atteindre un niveau de preuve supérieur (Behrens & Langer, 2004 ; Gastinger et al., 2002 ; Petticrew & Roberts, 2003). Il y a même des indices scientifiques selon lesquels les RCT n'occupent pas toujours le niveau supérieur de l'échelle des évidences et les études observationnelles

non plus les échelons inférieurs (Petticrew & Roberts, 2003). En plus, la conviction gagne du terrain que les analyses quantitatives ne peuvent apporter un bénéfice réel pour le groupe cible concerné que lorsqu'elles sont intégrées dans la recherche et les connaissances qualitatives (Grypdonck, 2004). Il y avait déjà plusieurs tentatives de créer une hiérarchie des niveaux de preuve spécifique à la recherche des soins. Ainsi, Lavin et al. (2002) ont établi une classification soignante avec la nécessité de prendre en compte les questions de recherche sous-jacentes, basée sur les différences qui existent entre Evidence-Based Nursing et Evidence-Based Medicine. Une autre approche pour définir une hiérarchie des niveaux de preuve en recherche de soins est proposée par Evans (2003). Il focalise trois dimensions : l'efficacité, l'adéquation et la faisabilité des interventions dans le domaine de la santé. Les études observationnelles, de même que les essais contrôlés randomisés, sont situées au niveau « bon ».

Niveau d'évidence	Efficacité	Adéquation	Faisabilité
Excellent	Travaux de synthèse systématiques Etudes multicentriques	Travaux de synthèse systématiques Etudes multicentriques	Travaux de synthèse systématiques Etudes multicentriques
Bon	RCT Etudes observationnelles	RCT Etudes observationnelles Etudes interprétatives*	RCT Etudes observationnelles Etudes interprétatives*
Modéré	Designs d'études non contrôlés avec résultats clairs Etudes avant-après Essais non randomisés contrôlés	Etudes descriptives Focus groups	Etudes descriptives Etudes interventionnelles Etudes avant-après Focus groups
Faible/mauvais	Etudes descriptives Séries de cas Opinion d'experts Etudes de mauvaise qualité méthodologique	Opinion d'experts Séries de cas Etudes de mauvaise qualité méthodologique	Opinion d'experts Séries de cas Etudes de mauvaise qualité méthodologique

Tableau 13 : Hiérarchie des évidences selon Evans pour évaluer des interventions dans le domaine de la santé (source : présentation originale d'après Evans, 2003)

* **Etudes interprétatives** : cela englobe, entre autres, les analyses documentaires, analyses narratives, analyses conversationnelles, analyses de contenu, méthodes herméneutiques, phénoménologie, théorie ancrée

Outre les réflexions jusqu'ici exposées, la qualité de réalisation d'une étude publiée joue également un rôle important dans l'évaluation de la qualité (Behrens & Langer, 2016). Ainsi, on ne peut pas attribuer automatiquement une qualité élevée à un RCT ou une revue de synthèse systématique. Le tableau ci-après donne une idée grossière des sources d'erreurs systématiques les plus fréquentes observées dans des études qui sont susceptibles d'entraver l'évaluation de la qualité de preuve.

Sources d'erreurs systématiques fréquentes dans la réalisation d'études	
Source d'erreur	Explication
Nombre de cas trop faible	Avec un échantillon trop petit, les différences qui existent ne sont possiblement pas détectées.
Question imprécise ou absence de théorie déduite	Les questions posées ne sont pas claires ou il manque de contextualisation théorique. Remède : cadre théorique, évtl. schéma PICO
Manque de valeurs mesurées	Problème rencontré par ex. dans des méta-analyses qui recourent à des valeurs estimées pour faciliter des calculs ultérieurs, les mesures n'étant pas possibles à cause de la situation inhomogène des études.

Choix inapproprié d'outils de mesure	Surtout en cas d'études interventionnelles, il s'agit de déterminer le résultat (outcome)/la différence. Si l'on utilise des outils imprécis ou inappropriés pour déterminer les effets, on risque de mesurer autre chose que prévue. Remède : évtl. recherche qualitative pour développer des outils de mesure appropriés.
Facteur de confusion, variable confondante (confounder)	Paramètre ou facteur de risque qui est associé à la fois à l'exposition/intervention et à la grandeur cible (par ex. maladie). Ex. : âge, sexe, consommation de nicotine, maladies additionnelles Remède : dans un design d'étude expérimental, par ex. randomisation ; dans une analyse de données par ex. stratification
Variations dans l'évolution de la maladie	Amélioration ou aggravation au cours de la maladie, indépendamment de l'intervention, provoquée par des variables/influences non saisies, par ex. amélioration spontanée des rhumes, progression d'un diabète, poussées de la sclérose en plaques ou du rhumatisme Remède : groupe de contrôle
Biais de sélection (selection bias)	Différences systématiques dans la sélection des participant-e-s ou dans la composition des groupes, par ex. hommes/femmes, patient-e-s présentant des symptômes légers/graves Remède : randomisation, assignation en aveugle
Biais de performance (performance bias)	Différences systématiques entre les groupes par ex. en termes de soins, attention etc. Remède : mise en aveugle des participant-e-s et intervenant-e-s
Biais de mémorisation (recall bias)	Distorsion du fait de la mémorisation, par ex. dans des études rétrospectives : les participants qui ont un souvenir imprécis d'une possible exposition mentionnent celle-ci plus souvent quand ils sont malades que lorsqu'ils ne le sont pas. D'autres exemples peuvent être des données défavorables, lacunaires en cas d'analyses documentaires.
Biais de l'observateur (de constatation) (observer/ascertainment bias)	Différences systématiques dans l'évaluation parce que l'assignation aux groupes n'est pas (le suffisamment) mise en aveugle.
Biais d'information (detection bias)	Différences systématiques entre les groupes dans la détermination du résultat, par ex. : le traitement hormonal chez les femmes entraîne des visites médicales plus fréquentes (par ex. à cause des saignements) ce qui augmente la détection de carcinomes utérins – même si les hormones de favorisent pas les carcinomes.
Déviation du plan d'étude (attrition bias, drop out, loss to follow up)	Différences systématiques entre les groupes en cas de déviations du protocole d'étude, par ex. si dans le groupe d'intervention plus de participant-e-s arrêtent l'étude à cause des effets secondaires que dans le groupe de placebo. Remède : analyse en intention de traiter (ITT)
Biais de publication (publication/reporting bias)	Différences systématiques entre les résultats publiés et non publiés, par ex. parce que les résultats significatifs, souhaités et spectaculaires sont publiés plus souvent, plus vite et dans des journaux plus prestigieux. Remède : application des lignes directrices correspondantes pour la publication et transparence du processus de recherche

Tableau 14 : Sources fréquentes d'erreurs systématiques dans la réalisation d'études (source : présentation originale d'après Grypdonck, 2004 ; Rojahn, 2016)

Les observations présentées jusqu'ici ne laissent pas de doute sur la complexité du classement d'une étude/sa publication dans une hiérarchie des évidences et montrent clairement qu'il ne faut pas prendre le design d'étude pour unique critère. A part cela, le design d'étude comme critère de fiabilité du niveau de preuve mérite un regard critique – notamment vu la multitude de possibles sources d'erreurs. Dans le contexte de la recherche en soins, il faut toujours poser la question méthodologique de savoir si le design de recherche est apte à répondre à la question de recherche (Mayring, 2017 ; Petticrew & Roberts, 2003). Plusieurs auteurs préconisent par conséquent d'évaluer le niveau de preuve d'un travail de recherche toujours à l'aulne de la question de recherche posée (Baker et al., 2010 ; Bartholomeyczik et al., 2008 ; Petticrew & Roberts, 2003). Au lieu d'appliquer des niveaux d'évidence rigides, ils privilégient l'utilisation d'une matrice des évidences sensible aux questions de recherche et au pluralisme méthodologique (Petticrew & Roberts, 2003). Un exemple correspondant se trouve dans la discussion à ce sujet qui a lieu en soins psychiatriques : Une quantité considérable des connaissances et de l'évidence y est déduite

d'études de cas particuliers (Stickley & Phillips, 2008). Dans le champ de tension entre l'individualité et la généralisabilité, la transposabilité des hiérarchies d'évidence rigides suscite des réflexions critiques : elle risque de privilégier les concepts d'action qui sont basés sur les niveaux d'évidence théoriquement les plus élevés quitte à négliger d'autres concepts qui sont parfois aussi ou même plus efficaces dans la situation individuelle de prise en charge. Les arguments exposés sont encore un motif qui explique pourquoi, dans le développement de ENP, les niveaux d'évidence ne sont pas indiqués explicitement pour chaque concept d'intervention soignante.

2.2.3.3 Evaluation de l'évidence dans le développement de ENP d'après le schéma GRADE

Après la lecture critique à base des outils d'évaluation correspondants, un niveau d'évidence est attribué à la publication ou le travail de recherche dans sa totalité. Le but en est d'obtenir, lors (de la poursuite) du développement d'une ligne directrice de pratique ENP, une impression globale de la qualité des références utilisées pour estimer leur validité. Le niveau d'évidence d'une référence reflète le degré de confiance dans l'exactitude et la fiabilité des propos avancés sur des causalités diagnostiques ou des estimations de l'effet d'un concept d'intervention. Le résultat cumulé d'une évaluation globale est saisi dans le tableau d'évidence ENP. Le développement ENP prétend recourir dans la mesure du possible à des références d'un niveau d'évidence élevé pour assurer une haute fiabilité des diagnostics soignants avec leurs éléments associés (caractéristiques, étiologies et ressources) et concepts d'action reliés. L'évaluation globale d'études ou publications en termes de qualité est de pratique courante dans le développement de lignes directrices (Platz, 2021). Beaucoup d'outils de lecture critique proposent aussi des éléments ou méthodes visant l'évaluation globale d'une publication. Quelques-unes de ces approches se trouvent parmi les outils d'évaluation présentés (cf. chapitre 2.2.1). En effet, les différences dans l'opérationnalisation des critères ou questions d'évaluation globale rendent difficile une vue d'ensemble de la qualité des publications consultées.

C'est pourquoi l'équipe ENP opte pour une systématique homogène pour évaluer l'évidence globale d'une publication. Pour illustrer l'utilité, ci-après un exemple tiré de la révision d'une ligne directrice de pratique ENP liée à la fatigue avec un volume de 50 études pertinentes qui ont été analysées et évaluées. L'évaluation synthétique affichée dans le tableau permet de saisir rapidement la qualité de la recherche à ce sujet.

Type d'étude	Evaluation de l'évidence d'après GRADE	Source
RS	● ● ● ●	Patterson, E., Wan, Y. W. T., & Sidani, S. (2013). Nonpharmacological nursing interventions for the management of patient fatigue: a literature review. <i>Journal of Clinical Nursing</i> (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. <i>JAMA Oncol</i> , 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. <i>Evid Based Complement Alternat Med</i> , 2011. doi: 10.1155/2011/142913

Tableau 15 : Extrait illustratif de la vue d'ensemble des publications utilisées dans le cadre d'un travail de révision de ENP

Quant à la systématique et la méthode de l'évaluation synthétique de l'évidence d'une publication, le schéma GRADE (Grading of Recommendations, Assessment, Development and Evaluation) est appliqué

(Kunz, Burnand, & Schünemann, 2008; Schünemann et al., 2013). Le constat d'évaluation est déterminé à base d'un schéma à quatre points.

Orientation sur GRADE

L'évaluation qualitative GRADE considère les RCT comme plus crédibles et robustes face au risque de biais que d'autres méthodes de recherche ; par conséquent, ce design est situé d'emblée sur un niveau d'évidence supérieur que par ex. les études observationnelles. Le classement initial d'un design d'étude déterminé peut certes être revu à la hausse ou à la baisse, dépendant de différents facteurs (cf. à titre d'illustration les résultats des deux RCT présentés dans le tableau 15 ci-dessus). L'évaluation qualitative GRADE est initialement conçue pour viser un point final dans le développement de lignes directrices. Pour cette raison, l'application de GRADE dans le contexte du développement de ENP a requis certains ajustements. Ici, l'évaluation porte sur la qualité d'une publication donnée et ne pas sur un point final. Une approche similaire a déjà été utilisée avec succès par d'autres experts (Panfil et al., 2011). La grille GRADE classe la qualité d'évidence en quatre niveaux ou degrés. Le tableau suivant présente d'abord les niveaux d'évidence GRADE, tirés du manuel officiel de l'instrument, traduits vers le français.

Qualité de l'évidence	
Niveau	Définition
Elevé	Nous avons une confiance élevée dans l'estimation de l'effet : celle-ci doit être très proche du véritable effet.
Modéré	Nous avons une confiance modérée dans l'estimation de l'effet : celle-ci est probablement proche du véritable effet, mais il est possible qu'elle soit nettement différente.
Faible	Nous avons une confiance limitée dans l'estimation de l'effet : celle-ci peut être nettement différente du véritable effet.
Très faible	Nous avons très peu de confiance dans l'estimation de l'effet : il est probable que celle-ci soit nettement différente du véritable effet.

Tableau 16 : Niveaux d'évidence selon GRADE (source : représentation et traduction de RECOM selon Schünemann et al., 2013)

Pour ajuster GRADE au développement de diagnostics soignants ENP, le wording des quatre niveaux d'évidence a été adapté, comme présenté dans le tableau ci-après. Les niveaux d'évidence sont représentés par une symbolique à quatre niveaux.

Niveau de preuve	Symbolique	Définition
Qualité élevée	++++ ou ● ● ● ●	Il est très improbable que des études futures changent la confiance que nous avons dans le lien diagnostique et/ou l'effet thérapeutique observés. Il est très probable que le vrai effet soit proche de l'estimation de l'effet.
Qualité modérée	+++ ou ● ● ● ○	Il est probable que de études futures aient un impact important sur la confiance que nous avons dans le lien diagnostique et/ou l'effet thérapeutique observés. Il existe une confiance modérée que le vrai effet soit proche de l'estimation de l'effet. Mais il existe la possibilité qu'il en diffère.
Qualité faible	++ ou ● ● ○ ○	Il est très probable que des études futures changent l'estimation de l'effet diagnostique et/ou l'effet thérapeutique observés. La confiance dans l'estimation de l'effet est limitée et le vrai effet peut différer considérablement.
Qualité très faible	+ ou ● ○ ○ ○	Le lien diagnostique et/ou l'effet thérapeutique observés sont soumis à une grande incertitude. La confiance dans l'estimation de l'effet est très limitée.

Figure 10 : Niveaux de preuve traduits et adaptés pour l'utilisation dans le développement ENP d'après GRADE (sources : Kunz et al., 2008 ; Platz, 2021 ; Schünemann et al., 2013)

Tout comme d'autres systématiques, GRADE définit une hiérarchie d'évidence par rapport au design d'étude qui sert de point de départ pour les décisions de classement qui suivent dans le processus d'évaluation. Le groupe de développement de GRADE a défini des critères pouvant entraîner un surclassement ou un déclassement du niveau de preuve. Là aussi, des ajustements ponctuels ont été

réalisés pour l'application en recherche de soins. Concrètement, les designs de recherche qualitatifs ont été ajoutés à l'évaluation initiale du niveau de preuve et par conséquent, des particularités liées aux études qualitatives ont été ajoutées aux critères de dé/surclassement.

Classement initial des études en fonction du design d'étude et de la question de recherche

Le tableau suivant recueille les classements initiaux pour l'évaluation d'évidence globale dans le cadre du développement ENP. Cette grille complète le schéma GRADE original en plusieurs points. Au niveau des interventions de soins, on a gardé la hiérarchie d'évidence ainsi que les critères de sur/déclassement. Quant au développement de diagnostics soignants, par contre, des ajouts ont été apportés, car dans le cas de ENP il ne s'agit pas seulement d'étudier les relations causales d'interventions de soins mais également de développer des lignes directrices de pratique. Les ajouts s'orientent sur la discussion menée autour du classement des preuves dans le contexte spécifique des soins (Behrens & Langer, 2004 ; Evans, 2003) et sur le diagnostic soignant et les chemins thérapeutiques (Grieser, Hegedüs, & Kunz, 2011). Le classement initial d'une étude est différent en fonction du design de recherche appliqué ou de la vérification de l'effet d'interventions de soins, la confirmation d'un diagnostic soignant avec caractéristiques/symptômes ou encore de la question portant sur des facteurs d'influence ou des liens de causalité. Comme exposé précédemment, il est judicieux pour plusieurs raisons de définir les hiérarchies d'évidence dans le contexte de la question de recherche correspondante. En conséquence, le classement initial des niveaux de preuve de GRADE s'est vu adapté au contexte de recherche de ENP.

Catégorie d'évidence	Fiabilité du design d'étude quant aux preuves d'effet des interventions de soins	Fiabilité du design d'étude quant au phénomène diagnostique et aux caractéristiques	Fiabilité du design d'étude quant aux facteurs d'influence/étiologies d'un diagnostic soignant
Fiabilité élevée	- Travaux de synthèse systématiques ou méta-analyses avec RCT - Etudes multicentriques - Lignes directrices	- Travaux de synthèse systématiques, méta-analyses avec travaux de recherche qualitatifs et/ou quantitatifs, non randomisés - Lignes directrices - Etudes multicentriques	- Travaux de synthèse systématiques, méta-analyses avec travaux de recherche qualitatifs et/ou quantitatifs, non randomisés - Etudes multicentriques
Fiabilité modérée	- RCT - Etudes de cohorte	- Triangulation méthodologique - Etudes observationnelles - Travaux de recherche qualitatifs - Etudes interprétatives*	- RCT - Etudes de cohorte - Etudes cas-témoins
Fiabilité faible	- Etudes observationnelles - Etudes quasi-expérimentales sans randomisation (non-randomized controlled trial) - Travaux de synthèse/méta-analyses d'études cas-témoins ou de cohorte - Etudes avant-après - Etudes cliniques contrôlées, par ex. études interventionnelles (essais)	- Etudes non-analytiques : séries de cas, études de cas - Etudes descriptives, non expérimentales - Méthode Delphi	- Etudes transversales - Etudes de prévalence/d'incidence - Etude de corrélation

	croisés), études comparatives		
Fiabilité très faible	• Etudes descriptives, non expérimentales • Etudes non-analytiques : séries de cas, études de cas • Travaux de recherche qualitatifs • Opinion d'expert-e-s donnée par des autorités reconnues (au niveau de livres spécialisés par ex. manuels en soins reconnus) • Méthode Delphi	• Séries de cas, études de cas • Opinion d'expert-e-s donnée par des autorités reconnues (au niveau de livres spécialisés par ex. manuels en soins reconnus)	• Séries de cas, études de cas • Opinion d'expert-e-s donnée par des autorités reconnues (au niveau de livres spécialisés par ex. manuels en soins reconnus) • Méthode Delphi

Tableau 17 : classement initial du niveau de preuve global dans le cadre du développement de ENP, d'après GRADE

* **Etudes interprétatives** : cela englobe, entre autres, les analyses documentaires, analyses narratives, analyses conversationnelles, analyses de contenu, méthodes herméneutiques, la phénoménologie ou la théorie ancrée.

Facteurs réduisant ou augmentant l'évidence dans la recherche de ENP

Par la suite, nous présentons les différents facteurs (adaptés d'après GRADE) pouvant entraîner une baisse ou une augmentation de l'évaluation globale du niveau de preuve. Une baisse ou augmentation du niveau d'évidence ne reflète jamais une catégorie rigide mais un continuum au sein d'une catégorie. Si les preuves relatives à un facteur déterminé ne sont que modérées, la décision de savoir si une étude/référence dépasse ou non le seuil défini pour la baisse ou l'augmentation de la qualité (d'un ou de plusieurs facteurs) dépend du jugement de la personne évaluatrice. Il s'agit entre autres d'estimer l'impact exercé par le facteur en question sur le résultat global (Schünemann et al., 2013).

Facteurs pouvant <u>réduire</u> le niveau de preuve	
Facteur	Conséquence
Insuffisances de la conception ou réalisation de l'étude (risque élevé de biais), par ex. : Absence de dissimulation de l'assignation Mise en aveugle insuffisante Follow-up insuffisant Rapport sélectif des résultats Outil de mesure inadéquat pour saisir les valeurs de mesure Critères imprécis d'in/exclusion de participant-e-s et/ou d'études Perte élevée de participant-e-s Biais de confusion Démarche non appropriée au sujet et à la question de recherche Description imprécise de l'intervention (Hoffmann et al., 2016) Différences au niveau des circonstances d'étude Nombre réduit de participant-e-s, large intervalle de confiance (IC)	↓ 1 ou 2 niveaux
Inconsistance des résultats	↓ 1 ou 2 niveaux
Preuves indirectes	↓ 1 ou 2 niveaux
Imprécision, par ex. absence d'indications sur l'origine de données, les motifs d'exclusion etc.	↓ 1 ou 2 niveaux
Biais de publication, par ex. : Choix sélectif d'études lors de la publication Rapport sélectif des résultats	↓ 1 ou 2 niveaux
Absence de cadre théorique ou localisation scientifique imprécise, absence de présentation des présupposés	↓ 1 ou 2 niveaux

Facteurs pouvant <u>augmenter</u> le niveau de preuve	
Facteur	Conséquence
Bonne description et déduction cohérente du design et de la réalisation de l'étude (faible risque de biais) Justification particulièrement transparente, cohérente et étayée par la théorie pour les décisions liées au sujet et à la question de recherche dans le design d'étude Sélection plausible des participant·e·s avec de grands échantillons qui laisse espérer une saturation Maintien constant plausible des conditions expérimentales Rapport transparent des participant·e·s qui abandonnent ET faible taux de ceux-ci (analyse en intention de traiter) avec indication d'effet statistique Triangulation de méthodes associée à des résultats homogènes Déduction théorique fondée avec le design de recherche adéquat	↑ 1 niveau
Grande puissance/taille d'effet par ex. RCT avec intervalle de confiance étroit	↑ 1 ou 2 niveaux
Tout facteur perturbateur plausible est susceptible de réduire ou d'augmenter l'effet prouvé. Par conséquent, ce risque peut être exclu si aucun effet n'a été observé.	↑ 1 niveau

Tableau 98 : Facteurs susceptibles de réduire ou d'augmenter le niveau de preuve global d'une étude/publication (source : présentation originale d'après Schünemann et al., 2013)

2.3 La suite du processus de développement de ENP

Dans le développement de ENP, toute proposition d'édition – élaborée au préalable par une personne principalement responsable – fait l'objet d'un consensus final au sein de l'équipe. Lors des séances de consensus, chaque membre de la réunion dispose du protocole de recherche documentaire, des documents de lecture critique ainsi que le tableau d'évidence et des publications originales. 10 % des références sont évaluées par une deuxième personne au hasard, par échantillonnage. Si la concordance déterminée est de 80 % ou plus, les évaluations d'études ultérieures seront réalisées uniquement par la personne responsable principale de la gestion du sujet concerté auparavant en équipe. Tout au long du processus de développement, la recherche documentaire et l'évaluation des études sont analysées ponctuellement lors des réunions hebdomadaires. A la fin du processus, toute proposition de modification de ENP qui en résulte est discutée et accordée en équipe lors d'une séance commune. Si l'on n'arrive pas à un accord pour un aspect déterminé, la littérature spécialisée est consultée de nouveau et/ou on demande l'avis d'autres expert·e·s reconnus en science et pratique soignante afin de trouver d'autres preuves.

Plus d'informations sur le processus systématique de développement et de validation de ENP sont disponibles dans les « Informations scientifiques », consultables en ligne sous <https://www.recom.eu/info-scientifique>.

Bibliographie

- 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 Guidelines 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://cqrmq.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/attach/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 Inception 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.000000000000060
- Mehrholz, J.** (2010). Wissenschaft erklärt: Qualitative Studiendesigns – Weiche Daten erfassen. *Ergopraxis*, 8(02), pp. 13-13.
- Moher, D. et al.** (2010). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Bmj*, 340, pp. c869. doi: 10.1136/bmj.c869
- Moher, D. et al.** (2011). Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. *DMW – Deutsche Medizinische Wochenschrift*, 136(8), pp. 9-15. doi: 10.1055/s-0031-1272978
- Moola, S. et al.** (2020). Chapter 7: Systematic reviews of etiology and risk. In E. Aromataris & Z. Munn (Eds.), *JBI Manual for Evidence Synthesis*: JBI. Retrieved from <https://synthesismanual.jbi.global> (Accessed: 18.07.2022).
- Moorley, C. & Cathala, X.** (2019). How to appraise qualitative research. *Evidence Based Nursing*, 22(1), pp. 10-13. doi: 10.1136/ebnurs-2018-103044
- Morse, J. M. et al.** (1996). Concept analysis in nursing research: a critical appraisal. *Sch Inq Nurs Pract*, 10(3), pp. 253-277.
- Munn, Z. et al.** (2020). Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. *JBI Evidence Synthesis*, 18(10), pp. 2127-2133. doi: <https://doi.org/10.11124/jbisrir-d-19-00099>
- Munthe-Kaas, H. M. et al.** (2019). Systematic mapping of existing tools to appraise methodological strengths and limitations of qualitative research: first stage in the development of the CAMELOT tool. *BMC Medical Research Methodology*, 19(1), pp. 113. doi: 10.1186/s12874-019-0728-6
- Niederberger, M. & Drejack, S.** (2020). Die qualitative Inhaltsanalyse in den Gesundheitswissenschaften. Ergebnisse eines systematischen Reviews einschlägiger Fachzeitschriften. *Forum Qualitative Sozialforschung / Forum: Qualitative Social, Research and Practice in Alzheimer's disease*, 21(1).

- Niederberger, M. & Peter, L.** (2018). Mixed-Methods-Studien in den Gesundheitswissenschaften. Ein kritischer Überblick. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 133, pp. 9-23. doi: <https://doi.org/10.1016/j.zefq.2018.02.008>
- Noble, H. & Smith, J.** (2015). Issues of validity and reliability in qualitative research. *Evidence Based Nursing*, 18(2), pp. 34. doi: 10.1136/eb-2015-102054
- Nordhausen, T. & Hirt, J.** (2020). RefHunter. Manual zur Literaturrecherche in Fachdatenbanken. Version 5.0 vom 13. Oktober 2020. Retrieved from <https://refhunter.eu/manual/> (Accessed: 25.02.2022).
- Oxford Centre for Evidence-Based Medicine.** (2009). Oxford Centre for Evidence-Based Medicine: Levels of Evidence. Retrieved from <https://www.cebm.ox.ac.uk/resources/levels-of-evidence/oxford-centre-for-evidence-based-medicine-levels-of-evidence-march-2009> (Accessed: 31.01.2022).
- Paez, A.** (2017). Gray literature: An important resource in systematic reviews. *J Evid Based Med*, 10(3), pp. 233-240. doi: <https://doi.org/10.1111/jebm.12266>
- Page, M. J., McKenzie, J. E., & Higgins, J. P. T.** (2018). Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open*, 8(3), pp. e019703. doi: 10.1136/bmjopen-2017-019703
- Panfil, E.-M., Ivanović, N., & Conrad, C.** (2011). Methodenpapier. FIT-Nursing Care. Forschungs- und IT-gestützte Expertennetzwerk-Plattform zur Unterstützung von Evidence-based Nursing. Version 1.0 Stand Juni 2011., 1-66. Retrieved from https://www.fit-care.ch/documents/8165518/0/110615_Methodenpapier_FIT%2BNC_Juni.pdf/a6d68f1c-55c9-4d09-982f-b40a374c0b2c (Accessed: 27.01.2022).
- Panke-Kochinke, B.** (2012). Augenscheinlich fehlgeleitet. Evidenz und Empirie. Methodische Postulate für eine qualitative Versorgungsforschung. *Pflege & Gesellschaft*, 17(1), pp. 5-21.
- Petticrew, M. & Roberts, H.** (2003). Evidence, hierarchies, and typologies: horses for courses. *J Epidemiol Community Health*, 57(7), pp. 527-529. doi: 10.1136/jech.57.7.527
- Pieper, D. et al.** (2015). Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. *J Clin Epidemiol*, 68(5), pp. 574-583. doi: 10.1016/j.jclinepi.2014.12.009
- Platz, T.** (2018). Leitlinien für Diagnostik und Therapie in der Neurologie. Therapie des spastischen Syndroms. Entwicklungsstufe S2k. Retrieved from https://www.dgn.org/images/red_leitlinien/LL_2019/PDFs_Download/030078_LL_Spastisches_Syndrom_2019_aktualisiert.pdf (Accessed: 11.05.2020).
- Platz, T.** (2021). Evidenzbasierte Leitlinienentwicklung der Deutschen Gesellschaft für Neurologie (DGN) und der Deutschen Gesellschaft für Neurorehabilitation (DGNR) – Methodik für die systematische Evidenzbasierung. *Fortschr Neurol Psychiatr*, 89(09), pp. 415-423.
- Ploeg, J.** (1999). Identifying the best research design to fit the question. Part 2: qualitative designs. *Evidence Based Nursing*, 2(2), pp. 36-37. doi: 10.1136/ebn.2.2.36
- Pluye, P. et al.** (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. *International Journal of Nursing Studies*, 46(4), pp. 529-546. doi: <https://doi.org/10.1016/j.ijnurstu.2009.01.009>
- Porritt, K., Gomersall, J., & Lockwood, C.** (2014). JBI's Systematic Reviews: Study selection and critical appraisal. *Am J Nurs*, 114(6), pp. 47-52. doi: 10.1097/01.naj.0000450430.97383.64
- Quasdorf, T. & Holle, D.** (2018). Mixed-Methods-Forschungsdesigns. In H. Brandenburg et al. (Eds.), *Pflegewissenschaft 2: Lehr- und Arbeitsbuch zur Einführung in die Methoden der Pflegeforschung*. Bern: Hogrefe, pp. 116-123.
- Roberts, J. & DiCenso, A.** (1999). Identifying the best research design to fit the question. Part 1: quantitative designs. *Evidence-Based Nursing*, 2(1), pp. 4-6.
- Rojahn, J.** (2016). Studien richtig lesen – Teil 1: Studiendesign und Fehlerquellen. *Lege artis – Das Magazin zur ärztlichen Weiterbildung*, 6(01), pp. 16-23.
- Schmucker, C. et al.** (2017). Bewertung des Verzerrungsrisikos von systematischen Übersichtsarbeiten: ein Manual für die Leitlinienerstellung. 1.0. Retrieved from https://www.cochrane.de/sites/cochrane.de/files/uploads/manual_bewertung_sr_fuer_ll.pdf (Accessed: 05.01.2022).

- Schmucker, C. et al.** (2016). Bewertung des Biasrisikos (Risiko systematischer Fehler) in klinischen Studien: ein Manual für die Leitlinienerstellung. Version 1.0 vom Mai 2016. Retrieved from <http://www.cochrane.de/de/news/manual-bewertung-des-biasrisikos-risiko-systematischer-fehler-klinischen-studien-ein-manual-f%C3%BCr> (Accessed: 07.03.2022).
- Schünemann, H. et al.** (2013). GRADE Handbook – Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Retrieved from <https://gdt.gradepro.org/app/handbook/handbook.html> (Accessed: 12.01.2022).
- Shea, B. J. et al.** (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, 358(j4008).
- Simon, M.** (2018). Literaturrecherche. In H. Brandenburg, E.-M. Panfil & H. Mayer (Eds.), *Pflegewissenschaft 2. Lehr- und Arbeitsbuch zur Einführung in die Methoden der Pflegeforschung* (3. ed.). Bern: Hogrefe, pp. 47-71.
- Stickley, T. & Phillips, C.** (2008). Einzelfallstudie und evidenzbasierte Praxis im Vergleich. *PPH*, 14(01), pp. 20-23.
- Stratil, J. M. et al.** (2021). Non-pharmacological measures implemented in the setting of long-term care facilities to prevent SARS-CoV-2 infections and their consequences: a rapid review. *Cochrane Database Syst Rev*, 9(9), pp. Cd015085. doi: 10.1002/14651858.CD015085.pub2
- Taylor, E. & Hignett, S.** (2014). Evaluating Evidence: Defining Levels and Quality Using Critical Appraisal Mixed Methods Tools. *HERD: Health Environments Research & Design Journal*, 7(3), pp. 144-151. doi: 10.1177/193758671400700310
- Thangatorai, R., Lim, F. C., & Nalliah, S.** (2018). Cervical pessary in the prevention of preterm births in multiple pregnancies with a short cervix: PRISMA compliant systematic review and meta-analysis. *J Matern Fetal Neonatal Med*, 31(12), pp. 1638-1645. doi: 10.1080/14767058.2017.1319930
- The AGREE Collaboration (Ed.).** (2001). Appraisal of Guidelines for Research & Evaluation (AGREE) Instrument – Deutschsprachige Version. Retrieved from https://www.agreetrust.org/wp-content/uploads/2013/06/AGREE_Instrument_German.pdf (Accessed: 08.03.2022).
- The AGREE Next Steps Consortium (Ed.).** (2013). Appraisal of Guidelines for Research & Evaluation II. AGREE II Instrument. Deutsche Version 2014. Retrieved from https://www.agreetrust.org/wp-content/uploads/2014/03/AGREE_II_German-Version.pdf (Accessed: 08.03.2022).
- Thole, H. et al.** (2007). Kritische Bewertung von Leitlinien. In R. Kunz et al. (Eds.), *Lehrbuch Evidenz-basierte Medizin in Klinik und Praxis* (2. ed.). Köln: Deutscher Ärzte-Verlag, pp. 177-187.
- Tyndall, J.** (2008). How long can you go? Toward a hierarchy of grey literature. Retrieved from <https://dspace.flinders.edu.au/xmlui/bitstream/handle/2328/3326/Tyndall.pdf?sequence=1&isAllowed=y> (Accessed: 24.02.2022).
- Tyndall, J.** (2010). AAOCDS Checklist. Retrieved from https://dspace.flinders.edu.au/xmlui/bitstream/handle/2328/3326/AACODS_Checklist.pdf?sequence=4&isAllowed=y (Accessed: 15.02.2022).
- Vandenbroucke, J. P. et al.** (2014). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *Int J Surg*, 12(12), pp. 1500-1524. doi: 10.1016/j.ijsu.2014.07.014
- Vaska, M. et al.** (2019). Exploring all that is Grey in the Health Sciences: What is Grey Literature and How to use it for Comprehensive Knowledge Synthesis. *Journal of National Heart Foundation of Bangladesh*, 8(1), pp. 14-19.
- von Elm, E. et al.** (2014). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *Int J Surg*, 12(12), pp. 1495-1499. doi: 10.1016/j.ijsu.2014.07.013
- Werner, R. N. et al.** (2016). The quality of European dermatological guidelines: critical appraisal of the quality of EDF guidelines using the AGREE II instrument. *J Eur Acad Dermatol Venereol*, 30(3), pp. 395-403. doi: 10.1111/jdv.13358
- Williams, V., Boylan, A. M., & Nunan, D.** (2019). Qualitative research as evidence: expanding the paradigm for evidence-based healthcare. *BMJ Evid Based Med*, 24(5), pp. 168-169. doi: 10.1136/bmjebm-2018-111131

- Williams, V., Boylan, A. M., & Nunan, D.** (2020). Critical appraisal of qualitative research: necessity, partialities and the issue of bias. *BMJ Evid Based Med*, 25(1), pp. 9-11. doi: 10.1136/bmjebm-2018-111132
- Woods, S., Phillips, K., & Dudash, A.** (2020). Grey literature citations in top nursing journals: a bibliometric study. *Journal of the Medical Library Association*, 108(2), pp. 262-269. doi: 10.5195/jmla.2020.760
- Zeng, X. et al.** (2015). The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med*, 8(1), pp. 2-10. doi: 10.1111/jebm.12141
- Ziegler, A. & König, I. R.** (2011). Leitlinien für Forschungsberichte: Deutschsprachige Übersetzungen von CONSORT 2010, PRISMA und STARD. *DMW – Deutsche Medizinische Wochenschrift*, 136(08), pp. 357-358. doi: 10.1055/s-0031-1272535

Exemple d'un tableau d'évidence, partiellement rempli, issu du développement de ENP :

46  Thieme RECOM

Annexe II : Protocole de documentation d'une recherche documentaire systématique

Motif et contexte de la recherche documentaire systématique :

-

Personne(s) réalisant la recherche :

-

Période de recherche :

-

1) Question(s) de recherche et précision

Sujet :

-

Méthode de délimitation choisie (par ex. recherche d'orientation, application de la méthode PICO etc.) et conclusions centrales en tirées :

-

Question(s) concrète(s) :

-

2) Dédution de termes de recherche

La multiplication des termes de recherche augmente les chances d'obtenir des résultats pertinents ! On peut notamment recourir à des synonymes, termes génériques et spécifiques ou à des traductions anglaises avec leurs synonymes respectifs.

Tous les éléments/concepts centraux de la question de recherche devraient d'abord être analysés un par un selon cette approche ; les termes de recherche qui en résultent peuvent ensuite être combinés pour former des phrases de recherche pertinentes.

La grille suivante peut servir d'aide pratique.

	Concept central 1 Résident·e en institution	Concept central 2 Comportement défi	Concept central 3 Démence	Concept central 4 Validation (intégrative)	Concept central 5 Snoezelen
Terme générique					
Terme spécifique					
Synonymes					
Traduction anglaise					

3) Définition des critères d'inclusion et d'exclusion

L'objectif : fixer le cadre formel de la recherche et mettre l'accent sur ce qui est vraiment pertinent et praticable (filtrage supplémentaire).

Voici quelques critères possibles :

- Langue des publications
- Période de parution
- Types de publication/d'étude
- Groupes d'âge, sexe etc.
- Bases de données/endroits de recherche
- Settings, professions
- ...

Important : Tout critère d'inclusion ou d'exclusion doit être justifiable !

Critères d'inclusion	Critères d'exclusion

4) Définition des bases de données et d'autres sources de recherche

En fonction du sujet à traiter, il faut également définir l'endroit où chercher la littérature spécialisée.

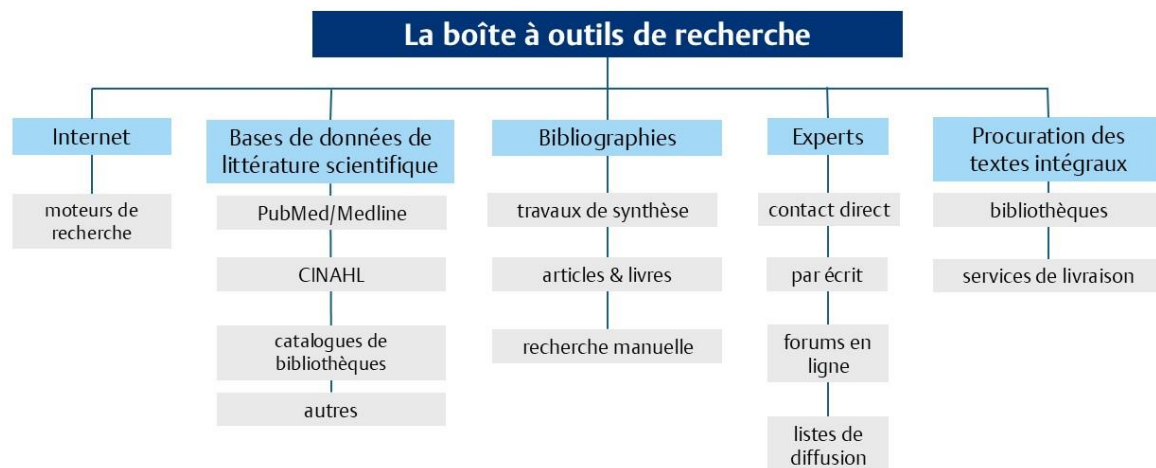


Figure 11 : Sources et options essentielles d'une recherche documentaire systématique (source : RECOM d'après Simon, 2018)

Choix des bases de données numériques avec une brève justification :

-

Choix des catalogues de bibliothèques/ressources en accès libre avec une brève justification :

-

Le cas échéant : choix des catalogues de lignes directrices (par ex. AWMF etc.) avec une brève justification :

-

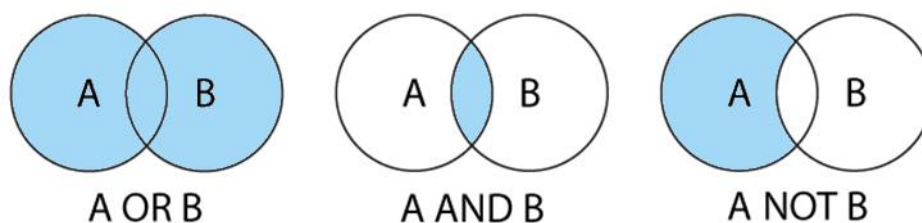
Choix d'autres endroits de recherche (revues spécialisées, littérature grise, moteurs de recherche en ligne, méthode boule de neige etc.) avec une brève justification :

-

5) Développement de phrases de recherche et alternatives

Pour assurer une recherche systématique la plus ciblée, maniable et complète possible et sans laisser des « tâches aveugles », mais aussi pour être en mesure de réagir de manière adéquate aux résultats de recherche non prévisibles en termes de qualité et quantité (trop ou trop peu de résultats, la plupart des résultats hors sujet...), il est fortement recommandable de générer différentes phrases de recherche. Outre les phrases de recherche « justes » supposées, il convient de considérer des alternatives pour limiter ou élargir la recherche. Les techniques ou aspects suivants devraient être pris en compte pour élaborer les phrases de recherche :

- **Recherche par champs** (utilisation de la structure des références bibliographiques dans une base de données, par ex. recherche par auteur, par titre, par abstract, ...)
- **Filtrage** surtout avec les critères d'inclusion/d'exclusion (par ex. période de publication, langue, groupe d'âge...)
- Utilisation de la **recherche par mots-clés (ou termes MeSH)** qui caractérisent une publication et permettent une détermination rapide des contenus. Les mots-clés ne doivent pas obligatoirement être contenus dans le texte proprement dit de la publication.
- **Troncature et masquage**, c'est-à-dire l'utilisation de jokers pour éviter la saisie de toutes les variantes grammaticales d'un terme de recherche ou d'une combinaison de mots. La troncature (souvent avec le caractère *) cherche les racines de mots avec toute sorte de terminaison ou de suffixe, tandis que le masquage remplace un seul caractère dans un terme de recherche (souvent avec le caractère # ou ?)
- Les **opérateurs booléens** qui servent à relier des termes de recherche de manière ciblée. Les opérateurs booléens étant l'une des fonctions les plus puissantes pour une recherche documentaire systématique, il vaut mieux vérifier deux fois leur application (source d'erreur courante). Les opérateurs les plus essentiels sont :
 - ✓ **AND** (rarement ET en français. N'affiche que les résultats qui contiennent tous les termes connectés – « et...et... »).
 - ✓ **OR** (rarement OU en français. Affiche tous les résultats qui contiennent le terme A ou le terme B ou le terme C (...) – « ou... ou ... »).
 - ✓ **NOT** (rarement SAUF en français. Exclut tous les résultats qui contiennent le(s) terme(s) correspondant(s) – « en aucun cas »).



- La **phrase de recherche** avec laquelle les termes saisis sont recherchés exclusivement dans l'ordre exact indiqué. Cette méthode peut être très utile notamment pour rechercher des concepts avec une combinaison fixe de plusieurs mots (par ex. « aidant familial »). La recherche par phrase est souvent réalisée en mettant les termes recherchés entre guillemets (« ... ») ou en activant une option spécifique dans le masque de recherche de la base de données (par ex. la fonction « ordre exact des mots »).

Phrases de recherche à utiliser et possibles alternatives :

Phrase de recherche partielle pour concept central 1			
Opérateur connecteur (AND, OR)	Terme de recherche (si besoin tronqué, masqué ou sous forme de phrase)	Recherche par champs ? (titre, MeSH,...)	Alternatives ?
	Terme principal		
OR	Synonyme 1		
OR	Synonyme 2		
OR	Synonyme 3		
Termes à exclure pour ce concept ?			
NOT			

Phrase de recherche partielle pour concept central 2			
Opérateur connecteur (AND, OR)	Terme de recherche (si besoin tronqué, masqué ou sous forme de phrase)	Recherche par champs ? (titre, MeSH,...)	Alternatives ?
	Terme principal		
OR	Synonyme 1		
OR	Synonyme 2		
OR	Synonyme 3		
Termes à exclure pour ce concept ?			
NOT			
Phrase de recherche partielle pour concept central 3			
Opérateur connecteur (AND, OR)	Terme de recherche (si besoin tronqué, masqué ou sous forme de phrase)	Recherche par champs ? (titre, MeSH,...)	Alternatives ?
	Terme principal		
OR	Synonyme 1		
OR	Synonyme 2		
OR	Synonyme 3		
Termes à exclure pour ce concept ?			
NOT			

Assemblage des phrases de recherche partielles relatives aux concepts centraux de la question de recherche pour former une phrase de recherche globale ; élaboration de possibles alternatives si les résultats obtenus sont trop étroits, larges ou thématiquement inappropriés :

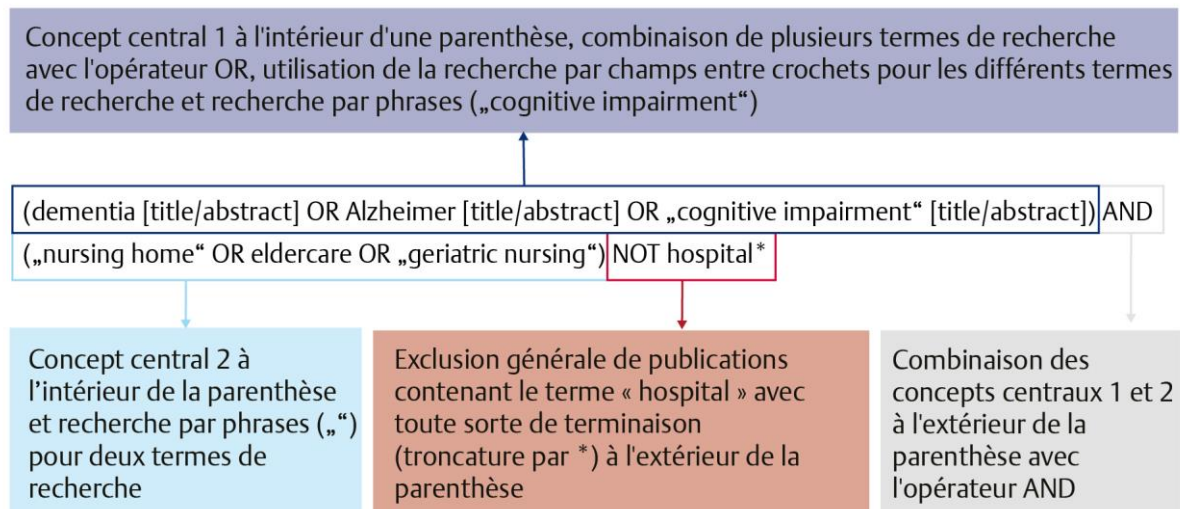
Phrase de recherche complète – variante 1				
	Opérateur connecteur (AND, OR, NOT)		Opérateur connecteur (AND, OR, NOT)	
Phrase de recherche partielle pour concept central 1	AND	Phrase de recherche partielle pour concept central 2	AND	Phrase de recherche partielle pour concept central 3
Termes à exclure pour la phrase de recherche complète ?				
			NOT	
Filtres de recherche (critères d'inclusion/d'exclusion)				
Période de publication	Langue	Conceptions d'étude/de publication	Groupe d'âge	...

Phrase de recherche complète – variante 2				
	Opérateur connecteur (AND, OR, NOT)		Opérateur connecteur (AND, OR, NOT)	
Phrase de recherche partielle pour concept central 1	AND	Phrase de recherche partielle pour concept central 2	AND	Phrase de recherche partielle pour concept central 3
Termes à exclure pour la phrase de recherche complète ?				
			NOT	
Filtres de recherche (critères d'inclusion/d'exclusion)				
Période de publication	Langue	Conceptions d'étude/de publication	Groupe d'âge	...

En complément à la représentation des phrases de recherche en tableau, il peut être utile de documenter les phrases sous forme textuelle. Tout comme en mathématiques, il faut veiller à une utilisation correcte des parenthèses. Les parenthèses expriment la priorité dans le traitement de la phrase dans la base de données : les éléments à l'intérieur d'une parenthèse sont traités en premier, ensuite les éléments à l'extérieur de la parenthèse. Généralement, il est également possible d'imbriquer des parenthèses. Pour rechercher un terme dans un champ spécifique (par ex. uniquement dans le titre ou dans l'abstract), il faut appliquer différentes techniques en fonction de la base de données utilisée – souvent cela est réalisé en plaçant le champ de recherche souhaité entre crochets [].

Voici un bref exemple pour illustration :

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



Phrase(s) de recherche sous forme textuelle :

-

6) Documentation des résultats de chaque opération de recherche et des phrases de recherche définitives

Pour saisir le déroulement de la recherche avec les modifications au niveau de la quantité et qualité des résultats dans différentes bases de données spécialisées (et autres endroits de recherche, si besoin), il est recommandable de documenter l'historique de toutes les opérations de recherche réalisées. Pour ce faire, toutes les bases de données et catalogues spécialisés importants offrent la possibilité de consulter et copier (par ex. en capture d'écran) l'historique de la recherche.

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		

Opération de recherche du <date> dans la base de données <nom> :

-

Opération de recherche du <date> dans la base de données <nom> :

-

Opération de recherche du <date> dans la base de données <nom> :

-

Phrase(s) de recherche finale(s) du <date> ayant déclenché l'abstract screening :

-

7) Documentation des résultats de l'abstract screening et de la procuration des textes intégraux :

Abstract screening de l'opération de recherche finale dans la base de données <nom> du <date> :

- Parmi le total de <XX> résultats, <YY> publications ont été évaluées comme pertinentes pour l'obtention en texte intégral.
- Parmi ces <YY> publications pertinentes, <ZZ> ont finalement pu être obtenues en texte intégral.

Abstract screening de l'opération de recherche finale dans la base de données <nom> du <date> :

- Parmi le total de <XX> résultats, <YY> publications ont été évaluées comme pertinentes pour l'obtention en texte intégral.
- Parmi ces <YY> publications pertinentes, <ZZ> ont finalement pu être obtenues en texte intégral.

Abstract screening de l'opération de recherche finale dans la base de données <nom> du <date> :

- Parmi le total de <XX> résultats, <YY> publications ont été évaluées comme pertinentes pour l'obtention en texte intégral.
- Parmi ces <YY> publications pertinentes, <ZZ> ont finalement pu être obtenues en texte intégral.