



4.16 Annexe à la section 4.5 – Exemples d’outils d’évaluation de la qualité



Formes de données probantes	Exemples d’outils d’évaluation de la qualité
Types de données probantes pour lesquelles des outils d’évaluation de la qualité existent	
Analyse des données	ROBINS-I (riskofbias.info) pour les études d’observation, telles que celles qui examinent les associations entre certains facteurs (y compris les interventions) et certains résultats, lorsqu’il existe un risque de biais provenant de : <ul style="list-style-type: none"> • facteur de confusion (lorsque la relation observée entre un facteur et un résultat diffère de la vraie relation en raison d’un ou plusieurs facteurs supplémentaires qui ne sont pas pris en compte) • sélection des participants à l’étude • classification de(s) intervention(s) • écarts par rapport à l’intervention prévue • données manquantes • mesure des résultats • sélection des résultats qui sont rapportés
Évaluation	Risque de biais (RoB) 2 (riskofbias.info) pour les essais contrôlés randomisés, où le risque de confusion est moindre, mais où il existe un risque de biais provenant de certaines (quoique moins nombreuses) des mêmes sources que ci-dessus : <ul style="list-style-type: none"> • processus de randomisation • écarts par rapport aux interventions prévues • données manquantes (résultats) • mesure des résultats • sélection des résultats qui sont rapportés
Recherche comportementale/ mise en œuvre	Voir les autres lignes pour les types d’études ou de synthèses pertinents
Recherche qualitative	Liste de contrôle d’évaluation critique JBI pour la recherche qualitative (bit.ly/31Lsib1), où des considérations très différentes entrent en jeu, telles que : <ul style="list-style-type: none"> • congruence entre la méthodologie de recherche et la question de recherche, les méthodes de collecte de données, la représentation et l’analyse des données et l’interprétation des résultats, ainsi qu’entre la perspective philosophique déclarée et la méthodologie • réflexivité de la part du chercheur, telles que des déclarations situant le chercheur culturellement et théoriquement, et abordant l’influence du chercheur sur la recherche et vice versa • représentation des participants à l’étude et de leurs voix • flux de conclusions de l’analyse et de l’interprétation des données



Synthèse de données probantes

Voir ci-dessus pour les types d'études pertinents pris en compte dans la synthèse des données probantes

A MeaSurement Tool to Assess systematic Reviews (AMSTAR; amstar.ca) pour la qualité des synthèses de données probantes, où le risque de biais peut découler de :

- identification de toutes les études potentiellement pertinentes grâce à une recherche complète de la littérature publiée et grise, et sans restrictions de langue
- sélection de toutes les études abordant la question de recherche en utilisant des critères explicites sur les conceptions des études et sur les participants, les interventions/facteurs, les comparaisons et les résultats, et avec au moins deux examinateurs appliquant les critères
- évaluation de la qualité et extraction des données de toutes les études incluses
- synthèse des résultats de toutes les études incluses

Notez qu'il existe deux versions d'AMSTAR : 1) la version originale qui peut être appliquée à tous les types de synthèses, bien que certains critères soient supprimés à la fois du numérateur et du dénominateur ; 2) une deuxième version d'AMSTAR plus particulièrement pertinente pour les synthèses d'essais randomisés contrôlés

Grading of Recommendations, Assessment, Development and Evaluations (GRADE ; bit.ly/3C9pMrx) pour la certitude des données probantes des résultats d'une intervention, avec :

- certitude réduite en raison du risque de biais (avec des données probantes issues d'essais contrôlés randomisés commençant avec une certitude élevée et des données probantes provenant d'études observationnelles commençant à une faible qualité, puis ajustées en fonction de RoB2 ou ROBINS-I), imprécision (par exemple, un ou deux petites études), incohérence (par exemple, deux études montrant des résultats très différents), caractère indirect (par exemple, mesures de substitution utilisées ou contextes atypiques étudiés) et biais de publication (par exemple, plus fréquent avec les études d'observation en raison du manque de registres d'études ou avec l'industrie -études financées en raison de l'incitation commerciale à publier des études positives)
- certitude évaluée pour une grande ampleur de l'effet, un gradient dose-réponse, et quand tous les facteurs de confusion résiduels diminueraient l'ampleur de l'effet

GRADE CERQual (cerqual.org) pour la certitude des données probantes de la représentation qualitative d'un phénomène d'intérêt, avec :

- la certitude est évaluée à la baisse en raison de préoccupations concernant les limites méthodologiques (parce que des problèmes dans la façon dont les études ont été conçues ou rapportées ont été identifiés à l'aide d'un outil d'évaluation critique comme celui du JBI ci-dessus), la pertinence (parce que le contexte dans lequel les études primaires ont été menées est substantiellement différent du contexte de la question de synthèse), la cohérence (parce que certaines des données contredisent les résultats ou sont ambiguës) et l'adéquation (parce que les données ne sont pas suffisamment riches ou ne proviennent que d'un petit nombre d'études ou de participants).



Évaluation des technologies / analyse coût-efficacité

Liste de contrôle du International Network of Agencies for Health Technology Assessment (INAHTA) (bit.ly/2YJVMVK) pour la qualité des évaluations de technologies, avec deux des 14 questions portant sur l'approche de la synthèse des données probantes (avec des questions similaires à AMSTAR) et une autre question portant sur si l'évaluation a été contextualisée par une analyse coût-efficacité (avec des données de coûts locales, c'est-à-dire nationales ou infranationales) et la prise en compte des implications juridiques, éthiques et sociales locales

Liste de contrôle Drummond des analyses coût-efficacité (bit.ly/3FbnB8R), et pour les évaluations économiques plus généralement, avec des questions sur la conception de l'étude, la collecte de données, et l'analyse et l'interprétation des résultats

Liste de contrôle Phillips pour les analyses coût-efficacité qui incluent un composant des modèles d'analyse décisionnelle (bit.ly/3FcWBGc) avec des questions sur la structure du modèle (par exemple, justification explicite, hypothèses justifiées et horizon temporel approprié), les données utilisées (par exemple, probabilités de base issues d'études observationnelles, effets du traitement issus d'essais contrôlés randomisés et évaluations de quatre types d'incertitude, à savoir la structure du modèle, les étapes méthodologiques suivies, l'hétérogénéité de la population étudiée et les paramètres utilisés) et la cohérence (interne et externe) - il existe également l'outil complémentaire TRUST pour évaluer les incertitudes dans les modèles d'analyse décisionnelle (bit.ly/3quFSKp)



Lignes directrices

Outil AGREE II (bit.ly/30qyFAB) pour évaluer le développement, le rapport et l'évaluation (ou l'appréciation de la qualité) des lignes directrices, qui utilise 23 éléments regroupés en six domaines, dont chacun est noté indépendamment :

- portée et objectif décrits
- implication des parties prenantes (citoyen/patient et professionnel)
- rigueur du développement (avec des synthèses de données probantes utilisées, un processus de développement de recommandations robuste et des recommandations liées aux données probantes à l'appui)
- clarté de la présentation
- applicabilité
- indépendance éditoriale (par rapport aux conflits d'intérêts des bailleurs de fonds et des membres du panel)

GRADE (bit.ly/3C9pMrx) pour évaluer la force des recommandations, qui utilise quatre considérations clés :

- équilibre entre les résultats souhaitables et indésirables (compromis), en tenant compte des meilleures estimations de l'ampleur des effets sur les résultats souhaitables et indésirables, et de l'importance de ces résultats (valeurs et préférences typiques estimées)
- confiance dans l'ampleur des estimations des effets des interventions sur les résultats importants (voir GRADE dans une ligne précédente)
- confiance dans les valeurs et les préférences, et leur variabilité dans l'utilisation des ressources



Types de données probantes pour lesquelles il n'existe pas encore d'outils d'évaluation de la qualité

Modélisation

Il n'existe pas encore d'outil largement accepté pour la plupart des types de modélisation. Cependant, certaines questions générales peuvent être posées (tout comme celles répertoriées dans la liste de contrôle Philips ci-dessus), telles que :

- structure du modèle (par exemple, justification explicite, hypothèses justifiées et horizon temporel approprié)
- les données utilisées (par exemple, les probabilités de base d'études d'observation, les effets d'intervention à partir d'un éventail de sources* et les évaluations de quatre types d'incertitude, à savoir la structure du modèle, les étapes méthodologiques suivies, l'hétérogénéité de la population étudiée et les paramètres utilisés)
- cohérence (interne et externe)
- disponibilité du logiciel ou de l'outil afin qu'il puisse être évalué par d'autres

* L'un des défis avec la COVID-19 était que les conceptions d'étude généralement utilisées pour déterminer les effets des interventions, tels que les essais contrôlés randomisés, étaient difficiles sur le plan éthique ou logistique et/ou prenaient du temps avant d'être complétées. Par conséquent, d'autres conceptions d'étude devaient être utilisées et il fallait demander l'avis d'experts (et il existe des approches qui permettent de le faire de manière systématique et transparente, comme SHELF – voir bit.ly/30nteC4)

Approches utilisées avec certains types de données probantes pour lesquelles il n'existe pas encore d'outils d'évaluation de la qualité

Intelligence artificielle

Il n'existe pas encore d'outil largement accepté

