

Notions utiles

La puissance d'une étude

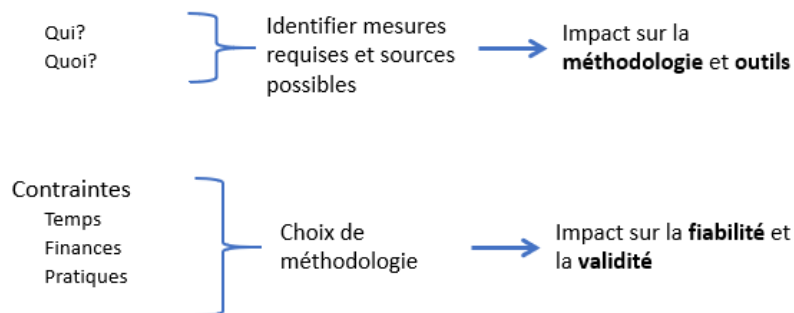
Pour obtenir un résultat statistiquement significatif l'échantillon doit être suffisamment grand pour mesurer un effet, et de nombreux outils existent pour calculer automatiquement le nombre de participants nécessaires.

Le fonctionnement est peu accessible si on n'est pas statisticien, mais voici le processus de calcul de la taille d'échantillon selon le site *The Analysis Factor*

1. Etablir un test d'hypothèse (null/alternative)
2. Etablir le niveau de significativité (ex $p < 0.05$)
3. Etablir la taille d'effet minimal ayant un intérêt scientifique.
4. Estimer les valeurs des autres paramètres nécessaires au calcul (écart type à partir de données pré-existantes)
5. Spécifier la puissance visée (0,8 ou 0,9)
6. Calculer

Le recueil des données

Deux éléments d'une étude



Pour choisir leur méthodologie les auteurs vont prendre en compte d'abord l'influence des deux éléments constitutifs ; l'échantillon et le sujet. Ceci va permettre de décider quoi et comment mesurer, et aider dans le choix des outils. Ensuite les aspects pratiques seront pris en compte pour déterminer ce choix, qui aura un impact sur la fiabilité et la validité de l'étude.

La fiabilité

On cherche toujours à éliminer les erreurs de mesure, qu'elle soit d'origine matérielle ou humaine. Une mesure doit d'abord être constante dans le temps. En anglais on parle de *reliability*, ou *consistency*. La fiabilité peut être évaluée de quatre manières différentes.

En répétant un test à une intervalle donnée lorsqu'on s'attend à ce qu'une mesure reste stable dans le temps.

En utilisant en parallèle deux tests équivalents mais non identiques

En incluant plusieurs façons de mesurer une même chose

En ayant recours à une vérification du degré d'accord entre deux évaluateurs

La validité

Il est important de faire la distinction entre la **validité d'une étude** et la **validité d'un test**.

Rappelons d'abord les deux versants de la validité d'une étude : La validité interne est atteinte en l'absence de biais, et la validité externe lorsque les conclusions peuvent être étendues à l'ensemble de la population. En pratique c'est un jeu de vases communicants entre ces deux.

La validité d'un test détermine à quel point une mesure doit mesurer ce qu'elle prétend mesurer. Par exemple, tester la rétention de quelque chose ne teste pas la capacité à l'appliquer, car dans le premier cas il s'agit de savoir, et pour le deuxième de savoir-faire.

On peut distinguer plusieurs types de validité, et les principaux sont listés ci-dessous

Surface (*face validity*) – atteinte lorsqu'un test semble répondre aux besoins

Contenu (*content validity*) – lorsque les items sont représentatifs de l'ensemble du domaine

Critère (*criterion validity*) – lorsque les résultats d'un test sont en adéquation avec ceux d'un autre. Il y a deux sous-catégories :

Prévision (*predictive validity*) – le test donné peut prévoir B à partir de A

Comparaison (*concurrent validity*) – le test est concordant avec test déjà éprouvé

Lorsqu'on souhaite vérifier la fiabilité et la validité de ce qui a été mesuré en lisant un article, on peut se poser quatre questions de base.

Est-ce que les détails de méthodes de recueil sont fournis ?

Est-ce que toutes les variables ont été clairement définies ?

Est-ce que les détails des mesures sont identifiables ?

Est-ce que les auteurs ont expliqué les intervalles entre mesures ?

Distinguer la fiabilité et la validité des résultats :

	Fiabilité	Validité
Qu'est ce qu'elle indique?	Le degré auquel on peut reproduire les résultats lorsque l'étude est menée une seconde fois dans les mêmes conditions.	Le degré auquel les résultats mesurent vraiment ce qu'ils sont censés mesurer.
Comment on peut l'évaluer ?	En vérifiant la cohérence des résultats dans le temps, avec différents évaluateurs, et entre les différentes parties du test.	En vérifiant le degré de concordance avec des théories éprouvées et d'autres mesures de la même notion.
Comment les différencier?	Une mesure fiable n'est pas toujours valide. Les résultats pourront être reproductibles, mais ils ne sont pas nécessairement justes.	Une mesure valide est généralement fiable également: si un test produit des résultats justes, ceux-ci devraient être reproductibles.

Evaluer la validité d'une étude

Menaces pour la validité interne

1. **Maturation** – Dans une étude plus longue on peut constater un changement des comportements ou attitudes des participants au fil du temps
2. **Sélection** – Les biais peuvent se produire à l'inclusion, et ensuite à l'attribution
3. **Histoire** – Parfois des événements socio-politiques peuvent influencer sur un résultat qui ne serait pas le même à une autre période
4. **Instrumentation** – sans calibration et contrôle des outils de mesure, le résultat sera peu fiable

5. **Régression** – c'est effet bien connu dû à un phénomène statistique qui fait tendre un résultat vers une moyenne
6. **Attrition** – Cela peut provenir des données manquantes, où par des abandons, surtout si cela révèle des différences fondamentales entre ceux qui persévèrent et les autres.

Menaces pour la validité externe

1. **Effet de test réactif** – répondre à une enquête avant une intervention peut influencer sur le comportement
2. **Effet de sélection interactif** – les participants ont des caractéristiques qu'on ne trouverait pas ailleurs
3. **Effet d'innovation réactif** – la nature artificielle de la situation change l'attitude des participants
4. **Interférence externe** – la participation à des activités en dehors de l'étude peut avoir une influence sur ce qui se passe lors des interventions ou observations.

Le rôle de l'échantillon dans la validité

L'échantillon doit être un **sous-groupe** d'une **population cible** que l'on a identifié lors de la formulation de la question de recherche. Seulement ainsi il devient possible d'extrapoler les résultats.

Il doit être suffisamment bien défini qu'on sélectionne à partir de **critères d'inclusion et d'exclusion**

Les méthodes d'attribution choisies vont souvent avoir une influence sur la validité, car toutes ne sont pas égales. Dans la liste ci-dessous on voit une présentation hiérarchique en fonction du risque vis-à-vis de la validité :

1. Aléatoire
2. Systématique (tous les x)
3. Stratification (tiré de sous-groupes)
4. Communautés
5. Convenance (biais quasi systématique)

Review articles

Nous avons déjà vu que la recherche se base sur la connaissance des recherches précédentes, et que n'importe quel type d'étude commence par une revue de la littérature. Dans un article de recherche originale on se servira d'une revue pour établir l'intérêt et originalité de l'étude dans l'introduction, et on confrontera ses propres résultats à ceux de la littérature dans la discussion. Parfois, on va publier un article dont le point focal est une revue de la littérature dans une démarche d'exploration ou de questionnement.

Le *review article* est fondé sur des recherches déjà publiées, et cherche à synthétiser les connaissances, organiser ces dernières, évaluer ce qui est déjà paru, faire ressortir des tendances dans un domaine donné, et/ou mettre en évidence des besoins.

Theory/model – **présentation d'une nouvelle théorie ou schéma**

Issue – **enquête sur une question ou débat**

Narrative – **comparaison et résumé, approche qualitative**

Status quo – **état des connaissances sur un sujet**

History – **exploration chronologique d'un domaine**

Best evidence – **études ciblées, approche méthodique**

Systematic – **analyse statistique d'études, méta-analyse**

Il y a différents genres d'article de revue, classifiés selon leurs objectifs. Ci-dessus une liste non-limitative.

L'objectif de la *review article* va influencer sur la structure du corps de l'article, entre l'introduction et la conclusion. Pour les cinq premiers le contenu ne suit généralement pas le format IMRaD

Bien qu'il y eût la toute première étude que l'on pourrait qualifier de revue systématique en 1753, les articles de type *review* avaient toujours eu une structure libre. Dans les années 70 et 80 on commençait à reprocher aux auteurs de ne pas avoir repéré des éléments importants, ou de diverger dans les conclusions d'un article à l'autre. Pour la recherche scientifique, il était important d'éviter les défauts induits par trop de liberté, et en 1992 Antmann, Lau *et al.* ont publié un article¹ dans lequel ils affirme que les *review articles* avaient besoin d'une rigueur identique à la recherche primaire, et c'est à partir de ce moment que la revue systématique est véritablement née et formalisée. Cela correspond également à la période pendant laquelle des consortiums se constituaient pour établir des exigences communes pour les études de recherche primaire.

Dans le tableau ci-dessus on voit les principales différences entre la revue dite narrative, et l'approche méthodologique de la revue systématique. Elle utilise une méthodologie rigoureuse pour faire face à l'explosion de publications, faire ressortir de nouvelles pistes de recherche, et pour évaluer des publications discordantes.

TABLE 1 Differences between a systematic review and a narrative review (based on Cook *et al.*⁸)

Feature	Narrative review	Systematic review
question	often broad in scope	often a focused clinical question
sources and search	not usually specified, potentially biased	comprehensive sources and explicit search strategy
selection	not usually specified, potentially biased	criterion-based selection, uniformly applied
appraisal	variable	rigorous critical appraisal
synthesis	often a qualitative summary	quantitative summary
inferences	sometimes evidence-based	usually evidence-based

Le processus

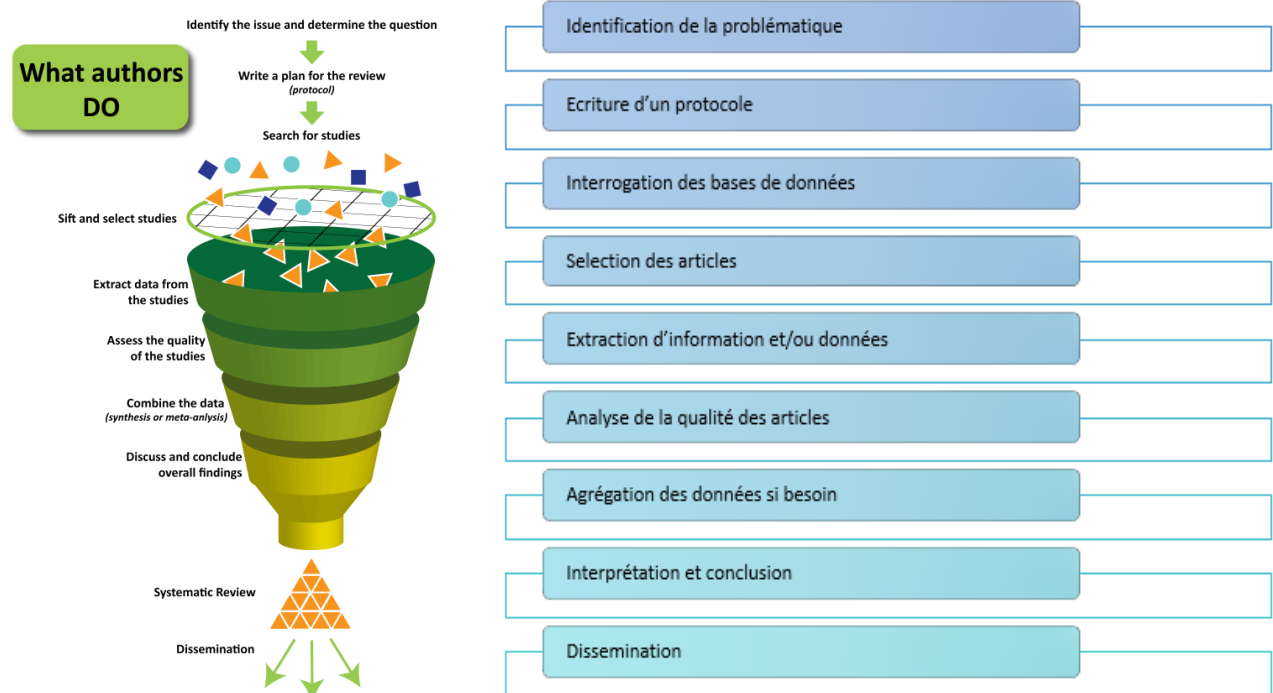


Figure 1: By Centre for Health Communication and Participation
La Trobe University, Australasian Cochrane Centre

L'illustration ci-dessus montre le processus de recherche, de tri, d'analyse et de synthèse qui est suivi dans la construction d'une revue systématique.

Les ressources utiles

La **Cochrane Library** est une base de données internationale de plus de 7500 revues systématiques de haute qualité dans le domaine de la santé. Ces revues sont régulièrement mises à jour, et on peut se fier à ces articles qui sont de très bonne qualité. Elles suivent les directives publiées dans leur livre de référence, consultable à <http://handbook.cochrane.org/>. Il y a également une courte vidéo sur qu'est-ce qu'une revue systématique accessible sur leur site à cochrane.org, dans l'onglet *about us*.

Nous avons déjà évoqué les outils répertoriés sur le site du *Equator Network*. Pour les revues systématiques c'est **PRISMA** (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) qui est l'outil indiqué. Celui-ci comporte deux parties, une liste de 27 items ainsi qu'un diagramme de flux pour le processus de sélection d'articles. La version actuelle est celle de 2009, et vous y trouverez aussi un certain nombre d'extensions qui enrichissent l'outil de base. <http://www.prisma-statement.org/>

GRADE – analyse et comparaison des études

Le groupe de travail international *Grading of Recommendations Assessment, Development and Evaluation* existe depuis 2000 et réunit des personnes qui souhaitent rectifier les problèmes des systèmes de classification dans le domaine de la santé. Il a développé une approche commune et transparente au classement des preuves et recommandations <http://www.gradeworkinggroup.org/>. Puisque la méthodologie de la revue systématique comporte une phase où on va analyser la qualité des articles, et même à appliquer des coefficients dans certaines circonstances, cette ressource est fondamentale.

Éléments pour l'analyse

Le lecteur des revues systématiques sera attentif à évaluer l'adhésion à la méthodologie, la cohérence entre les études retenues, la présence de résultats qui sont généralisables, et les conclusions qui sont tirées, notamment lorsqu'il s'agit d'une étude sur l'efficacité d'un ou plusieurs traitements.

Voici une suite de questions que l'on peut se poser :

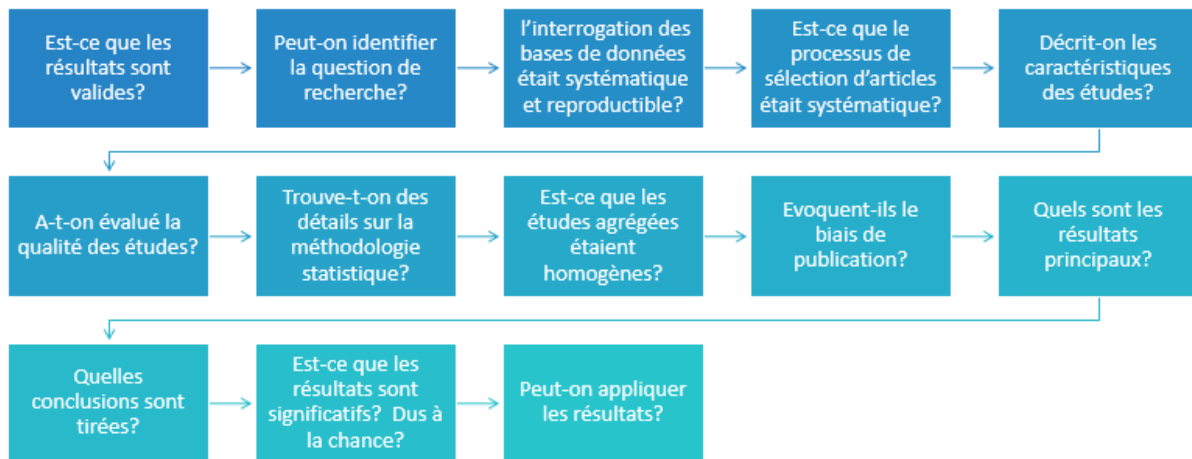


Figure 2

Et d'autres questions complémentaires :

1. Peut-on identifier les éléments clés PICO ?
2. Quelle stratégie de recherche a été mise en œuvre ?
3. Quels sont les critères de sélection de sources ?
4. Y a-t-il eu recours à des tiers indépendants/aveugles ?
5. Le protocole est-il clair ?
6. Comment ont été gérés les données manquantes ?
7. Y a-t-il pris en compte de résultats discordants ?
8. Prennent-ils en compte des exclusions, ou de biais de publication ?
9. Les recommandations ou conclusions, sont-elles logiques ?
10. Est-il question de mise à jour ?

Figure 3

Ce qui faut faire et ne pas faire lors d'une analyse d'une revue systématique

A FAIRE	A ÉVITER
Vérifier la focalisation de la question de recherche	Supposer qu'il y ait toujours une pertinence clinique
Vérifier la présence des critères d'inclusion et d'exclusion	Accepter les conclusions sans analyser la méthodologie
Vérifier la méthodologie de l'interrogation des bases de données	Faire l'impasse sur la vérification de la méthodologie d'analyse statistique
S'assurer que les auteurs limitent le risque de biais	Accepter les conclusions comme le dernier mot
Vérifier que la qualité des articles est évaluée	
Vérifier la prise en compte d'hétérogénéité des articles	
Evaluer la pertinence clinique	

Figure 4

Travail personnel

En puisant dans tous les réflexes de lecture critique que vous avez acquis jusqu'à présent, et en prenant des éléments de ce qu'on a vu aujourd'hui (PRISMA notamment) et les figures 2, 3 & 4, procédez à une analyse et commentaire de la revue systématique :

Li, T., Jiang, S., Han, M., Yang, Z., Lv, J., Deng, C., Reiter, R.J. and Yang, Y., 2019. Exogenous melatonin as a treatment for secondary sleep disorders: A systematic review and meta-analysis. *Frontiers in neuroendocrinology*, 52, pp.22-28.

ⁱ Antman, E.M., Lau, J., Kupelnick, B., Mosteller, F. and Chalmers, T.C., 1992. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: treatments for myocardial infarction. *Jama*, 268(2), pp.240-248.