

Méta-analyse : les bases méthodologiques

PARTIE I

Ludwig Serge Aho Glélé¹, Simon Aho²

1- Service d'épidémiologie et hygiène hospitalière – Centre hospitalier universitaire (CHU) – Dijon – France

2- Institut de cancérologie de Lorraine – Vandœuvre-lès-Nancy – France

✉ Dr Ludwig Serge Aho Glélé – Service d'épidémiologie et hygiène hospitalière – CHU de Dijon – 5 boulevard Jeanne d'Arc – 21000 Dijon – France – E-mail : ludwig.aho@chu-dijon.fr

Ce qu'est une méta-analyse et ce qu'elle n'est pas

Une méta-analyse (MA) est une synthèse, quantitative, des résultats de l'ensemble des essais¹ étudiant une question similaire. Elle permet souvent d'aboutir à des conclusions, alors que les résultats des études qui sont incluses dans cette MA peuvent ne pas être concluants, par manque de puissance. Une revue systématique de la littérature (*systematic review*) correspond à une recherche exhaustive de la littérature, associée à une analyse systématique de la qualité des études répondant à une question donnée. Mais elle ne comporte pas de synthèse quantitative comme une MA. La revue systématique est une étape préalable à la réalisation de MA. La MA n'est pas juste la réalisation d'une « moyenne simple » de la taille des effets procurés par chaque étude (mesurés par un risque relatif, un *odds ratio* [OR], un *hazard ratio*...) pour aboutir à la taille de l'effet commun. En agissant de la sorte, le groupe « intervention » de chaque étude ne disposerait plus de son groupe témoin spécifique. On aboutirait à des résultats paradoxaux, en particulier par le paradoxe de Simpson² [5,6]. Ainsi, lorsqu'un ou plusieurs des essais inclus dans la MA n'utilisent pas un schéma de randomisation de type « un-pour-un » (i.e. à chaque individu randomisé dans le groupe « intervention » correspond un individu randomisé dans le groupe témoin), la simple

mise en commun des données des essais (« moyenne simple ») peut conduire au paradoxe de Simpson.

Dans ce paradoxe (en fait un biais), l'association entre deux variables binaires (par exemple exposition et survenue de maladie) est similaire à l'intérieur de sous-groupes de la population, mais change de signe, si les individus des sous-groupes sont fusionnés et analysés sans stratification. Ceci peut être résumé par le titre de l'article de Baker et Kramer *Good for women, good for men, bad for people* [7]. Lorsqu'il y a de grandes disparités de taille d'échantillon parmi les essais, l'impact dudit paradoxe peut être assez important. L'exemple du **Tableau I** illustre un cas de paradoxe de Simpson. Il s'agit d'estimer la relation entre exposition aux lignes électriques à hautes tensions et la survenue de leucémie. Cinq études cas-témoins sont analysées et conduisent à des OR allant dans le même sens, i.e. OR supérieur à 1. En combinant simplement les données de ces cinq études, on aboutit à un OR égal à 0,67, donc inférieur à 1 (en revanche une MA conduit à un OR de 1,28). L'effet commun mesuré par la MA est basé sur une « moyenne pondérée » (stratification). La pondération usuelle est l'inverse de la variance de chaque étude.

La méta-analyse: une méthode récente

La première MA aurait été effectuée en 1904 par Karl Pearson³ [9] et concerne les maladies infectieuses. L'objectif était d'analyser des données comparant les taux d'infec-

1- Dans le présent texte, essai est assimilé à étude. L'essai peut être comparatif (randomisé ou non) ou non comparatif (par exemple une étude de prévalence).

2- Le paradoxe de Simpson (ou paradoxe de Yule-Simpson) a été initialement décrit par Karl Pearson en 1899. En 1903, Undy Yule le redécouvre. En 1951 Edward Simpson publie un article détaillé sur le sujet [1]. Ce paradoxe est lié au fait qu'une association entre deux variables peut disparaître ou même s'inverser suivant que l'on considère les données dans leur ensemble, ou bien segmentées par groupes. Ce paradoxe se produit en présence de facteur de confusion [2]. Dans le cas des essais randomisés, lorsqu'un ou plusieurs des essais inclus dans la MA n'utilisent

pas un schéma de randomisation de type « un-pour-un » (i.e. à chaque individu randomisé dans le groupe « intervention » correspond un individu randomisé dans le groupe témoin), la simple mise en commun des données des essais (« moyenne simple ») peut conduire au paradoxe de Simpson. Lorsqu'il y a de grandes disparités de taille d'échantillon parmi les essais, l'impact dudit paradoxe peut être assez important. Voir également : Goltz et al. pour le paradoxe de Simpson dans la recherche scientifique [3] et Delaye JP pour les paradoxes mathématiques, dont celui de Simpson [4] (p. 115-118, pour le paradoxe de Simpson).

3- Karl Pearson est aussi l'auteur d'un des tests statistiques les plus utilisés, le test du Chi-2 [8].

Tableau I – Illustration du paradoxe de Simpson
(Source : adapté de Hanley, 2000) [5].

	Étude	Distance <100 m	Distance >100 m	Odds ratio
1	Cas	18	162	1,12
	Témoins	25	252	
2	Cas	12	26	1,62
	Témoins	123	431	
3	Cas	27	121	1,02
	Témoins	104	475	
4	Cas	7	21	2,78
	Témoins	3	25	
5	Cas	14	353	2,03
	Témoins	7	359	
Total	Cas	78	683	0,67
	Témoins	262	1 542	

tion et mortalité parmi les soldats volontaires, vaccinés contre la fièvre typhoïde, dans divers endroits de l'Empire britannique, à ceux des soldats non vaccinés. Pearson a présenté les résultats de ses analyses dans un tableau dans lequel chaque ligne correspondait à une étude, ligne dans laquelle était notée la mesure de l'effet, ainsi qu'une mesure de l'incertitude. La dernière ligne donnait une estimation globale de l'effet (sans toutefois estimer l'incertitude globale associée à cette estimation). L'une des premières MA relative à un traitement médical curatif est plus tardive, publiée en 1955 par Beecher [10] et le terme « MA », dans son sens statistique, a été proposée par Glass en 1976 [11]. Dans les années 1970, à partir des travaux menés en sciences humaines par Glass [11], Schmidt et Hunter [12], des techniques analytiques plus élaborées apparaissent. D'autres détails historiques figurent dans l'article de O'Rourke [13] et le livre de Hunt et al. [14].

La méta-analyse : une méthode nécessaire

La synthèse quantitative des connaissances scientifiques est nécessaire. Selon Rayleigh, « *si comme on le soupçonne parfois, la science n'était rien d'autre que l'accumulation laborieuse de faits, elle s'immobiliserait bientôt écrasée sous son propre poids... Deux processus sont donc à l'œuvre, côte à côte, la réception de nouvelles données et la digestion et assimilation des anciennes* » [15]. Selon Chan et al. [16], en accord avec le paradigme de Kuhn⁴ [19], la MA, en aidant à accroître la précision et la

4- Dans son ouvrage *La structure des révolutions scientifiques*, publié en 1962 [17], Thomas S. Kuhn insiste sur les bouleversements de la pensée scientifique liés aux travaux de Copernic, Newton, Einstein, etc. Kuhn analyse ces moments de crise que traverse la science. Il mentionne également les conditions requises pour l'apparition d'une telle crise. Pour Kuhn, il y a révolution scientifique lorsqu'une théorie scientifique consacrée par le temps est rejetée au profit d'une nouvelle théorie. L'ouvrage de Kuhn a été traduit en français, sous le titre *La structure des révolutions scientifiques* [18].

portée d'un paradigme, contribue au progrès de la science dite normale. La MA facilite la prise de conscience par les chercheurs des anomalies d'un domaine et, ce faisant, elle joue un rôle clé dans le déclenchement de crises et de révolutions scientifiques, contribuant ainsi également au progrès de la science (selon le paradigme de Kuhn). Dans le même esprit, Shadish et Lecy⁵ [20], dans leur article intitulé *The meta-analytic big bang* examinent l'impact de la MA, puis explorent les raisons pour lesquelles la MA a été élaborée, dans les années 1970, par les chercheurs en sciences sociales, grâce aux travaux concomitants de Glass [11], de Rosenthal [21] et de Schmidt [12]. Ils notent que la MA a impacté toutes les « sciences » (sciences sociales, sciences, droit, médecine...), ainsi que de nombreux domaines tels que la modélisation statistique (modèles multiniveaux...), les statistiques médicales, l'évaluation des programmes. Aujourd'hui les principaux objectifs sont les suivants :

- augmenter la puissance du test statistique, en augmentant la taille de l'échantillon, ou dans le même esprit, d'améliorer la précision de l'estimation de l'effet de taille ;
- lever le doute en cas de résultats apparemment discordants, entre études ou entre revues de la littérature. Dans ce cas, la MA tente d'expliquer la variabilité des résultats disponibles ;
- formuler des hypothèses pouvant aboutir à la réalisation de nouvelles études.

Deux grands types de méta-analyse

La première est la MA sur données individuelles. La seconde est la MA sur données agrégées (résumées), i.e. généralement extraites de publications. La MA sur données individuelles demeure la méthode de référence. Elle permet une vérification et une analyse approfondie des données. Elle est néanmoins plus chronophage que la MA de données agrégées.

Les quatre étapes d'une méta-analyse

Ces étapes doivent être suivies par tout protocole de MA :

- 1- identification des études à inclure ;
- 2- recueil et vérification des données ;
- 3- analyse statistique principale et représentation graphique ;
- 4- exploration et analyse de l'hétérogénéité.

Identification des études à inclure

La stratégie de la recherche de la littérature répondant aux questions doit être systématique [22] : les publications

5- Ces auteurs analysent enfin pourquoi la MA a vu le jour dans les années 1970 dans le domaine des sciences sociales grâce aux travaux de Gene Glass, Robert Rosenthal et Frank Schmidt, qui ont chacun élaboré des théories similaires de MA à peu près au même moment. L'article se termine en expliquant comment la théorie de la « configuration aléatoire » (« *chance-configuration* ») de Simonton et l'épistémologie évolutionniste de Campbell peuvent expliquer pourquoi la MA a eu lieu avec ces chercheurs et non en sciences médicales.

indexées dans les bases de données bibliographiques telles que *Pubmed* [23], *ScienceDirect* [24], *Cochrane Library* [25], *Istex*, via *Inist-CNRS* [26], etc., mais aussi les bases de données d'essais cliniques telles que *clinical-trials.gov* [27] ou *EU ClinicalTrials Register* [28] et enfin la littérature grise (rapports d'études ou de recherches, actes de congrès, thèses, brevets... [29-31]). Cette dernière est disponible dans des bases telles que *OpenGrey* [32]⁶.

Recueil et vérification des données

Une étape importante de la MA est de vérifier la qualité de chaque étude et, le cas échéant, de décider sur des critères objectifs son inclusion ou non dans la MA. Des grilles d'analyse des biais des études sont disponibles et utiles : RoB (*Risk of bias*) pour les études randomisées [34]⁷ et Robins pour les études d'observation [35]. Il existe aussi des grilles spécifiques au type d'étude, par exemple Quadas-2 pour la MA d'études diagnostiques [36].

Analyse statistique principale et représentation graphique

Estimation de la taille de l'effet, article par article

L'estimation de la taille de l'effet⁸, article par article, prend en compte la nature du critère de jugement. Si le critère de jugement est binaire, on pourra estimer un OR, un risque relatif, une différence du risque relatif. Si le critère de jugement est la survie⁹, on pourra estimer des hazard ratio. Pour les critères de jugement quantitatif (moyenne, médiane, coefficients de corrélation), on pourra les standardiser : moyenne standardisée [6] ; « z-score » pour les coefficients de corrélation.

En ce qui concerne la moyenne standardisée¹⁰, le d de Cohen est souvent utilisé [38]. Il conduit à des biais (surestimation de l'effet) en cas de petits échantillons. L'alternative, non biaisée, au d de Cohen est le g de Hedges [39]. La formule de Hedges et Olkin [40], permet de passer de d à g¹¹. Notons que g est le standard utilisé dans le logiciel RevMan de la *Cochrane Library* [41].

Notons qu'il est possible de convertir un effet en un autre

(dans un sens donné...). On peut ainsi convertir un log (OR) (logarithme de l'odds ratio) en d (différence standardisée) et un coefficient de corrélation r, en d. Les différentes formules de conversion sont disponibles dans l'ouvrage de Borenstein et al. (chapitre VII) [6]. Les détails statistiques relatifs concernant la taille de l'effet figurent dans les ouvrages de Ellis [42] et de Grissom et Kim [43].

Estimation de la taille de l'effet global

La taille de l'effet global est estimée via une pondération des effets individuels (i.e. effet estimé par chaque article) par leur précision (inverse de la variance). Deux types de modèles statistiques peuvent être utilisés pour réaliser cette estimation : les modèles à effet fixe et les modèles à effet aléatoire [6,44]. Dans un modèle à effet fixe, on considère que chaque essai i représente une estimation d'un unique « vrai » effet du traitement¹².

$$\theta_i = \theta$$

L'estimation de l'effet commun peut être alors obtenue en utilisant la moyenne des estimations de chaque essai, pondérée par l'inverse de leur variance. Cette pondération est nécessaire du fait que les différentes estimations de i ne sont pas égales en termes de précision ou de variance. L'effet du traitement commun s'écrit alors :

$$\hat{\theta} = \frac{\sum \theta_i \omega_i}{\sum \omega_i}$$

Si l'hétérogénéité est « importante » ou, pour certains, si le test d'hétérogénéité est statistiquement significatif, il faut alors utiliser un modèle à effet aléatoire. Le modèle à effet aléatoire permet aux effets du traitement de varier, en faisant l'hypothèse que chaque essai représente une estimation d'un réel effet du traitement, lui-même étant une variable aléatoire normalement distribuée autour d'un effet global constant de moyenne et de variance. Ce modèle permet de décomposer la variance totale en une variabilité inter-essai et une variabilité intra-essai. En pratique, la recherche d'hétérogénéité statistique doit être systématique. En l'absence d'hétérogénéité, les modèles à effets fixes et aléatoires conduisent au même résultat.

MODÈLE À EFFETS FIXES OU ALÉATOIRES ?

Hunter et Schmidt [45] ont montré que les modèles à effets fixes entraînaient une erreur de type I¹³ importante, par rapport aux modèles à effets aléatoires. De plus, ils

6- Voir la note méthodologique concernant la méthode Grade (parue dans *Hygiènes* en 2018 [33]).

7- Une mise à jour sera bientôt disponible : Sterne JAC et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* (in press) [34].

8- Il arrive que cet effet figure dans les articles. Parfois, la taille de l'effet n'est pas disponible dans l'article et il faut l'estimer, ainsi que son écart-type, a posteriori.

9- D'une manière générale, le terme survie est relatif à un événement binaire (décès, pathologie...) survenant au cours du temps, plus précisément le temps au bout duquel l'événement se produit (*time to event*). Voir l'ouvrage de Hill et al. [37] pour plus de détails.

10- Moyenne standardisée = (moyenne du « groupe 1 » – moyenne du « groupe 2 ») / écart-type commun des deux « groupes ». Les groupes peuvent respectivement être un groupe « témoin » et un groupe « traitement ».

11- Formule de Hedges et Olkin : $g \approx d * \left(1 - \frac{3}{4 * (n1 + n2) - 9} \right)$

12- Dans le présent texte, un traitement est assimilé à une intervention, pharmacologique ou non (par exemple un dispositif médical tel que le masque...).

13- L'erreur de type I correspond au « fameux » p ou α . L'erreur de type I survient, dans un test d'hypothèse, statistique, lorsque l'hypothèse nulle, qui est en réalité vraie, est rejetée par erreur. L'erreur de type II survient lorsque l'hypothèse nulle est acceptée par erreur. Selon JF Bach, « *Mieux vaut se tromper quelques fois que de ne pas entreprendre une démarche scientifique originale par peur de commettre une erreur* » [46].

conduisent à des intervalles de confiance de l'effet global qui sont biaisés (étant plus étroits que leur valeur nominale). Ils recommandent de n'utiliser que les modèles à effets aléatoires en routine. Cela ne signifie pas que ces modèles sont la panacée ! Leurs limites ont été récemment (2019) rappelées et résumées par Serghiou et Goodman [47]. Premièrement, ces modèles n'expliquent pas l'hétérogénéité ! Il est recommandé de réduire l'hétérogénéité par des analyses en sous-groupe ou des méta-régressions. Deuxièmement, il existe de nombreuses méthodes pour calculer les estimations des effets. La méthode de DerSimonian-Laird [44] est la plus souvent utilisée¹⁴. Elle conduit à des intervalles de confiance trop étroits et des valeurs de p trop petits lorsqu'il y a peu d'études (i.e., inférieures à 10-15) incluses dans la MA et que l'hétérogénéité est importante. Troisièmement, les études de petite taille influencent plus fortement les estimations, qu'en cas d'utilisation des modèles à effets fixes. Le modèle de Doi et Barendregt (*inverse variance heterogeneity model* ou *Ivhet model*) est une alternative aux modèles à effets aléatoires [49]. Sa mise en œuvre dans les publications est pour l'instant très limitée¹⁵.

14- DerSimonian et Laird ont publié, près de 20 ans après leur article princeps, une mise à jour de leur méthode, avec entre autres, l'utilisation d'un estimateur robuste de variance [48].

15- Moins de trente articles mentionnant cette méthode dans Pubmed, et 122 citations dans Google Scholar, au 22 juillet 2019; vs 25 762 citations dans Google Scholar pour l'article de DerSimonian et Laird [44]. Il faut bien entendu pondérer par l'ancienneté de la publication (et non pas par l'inverse de la variance...).

Représentation graphique des effets individuels et de l'effet global

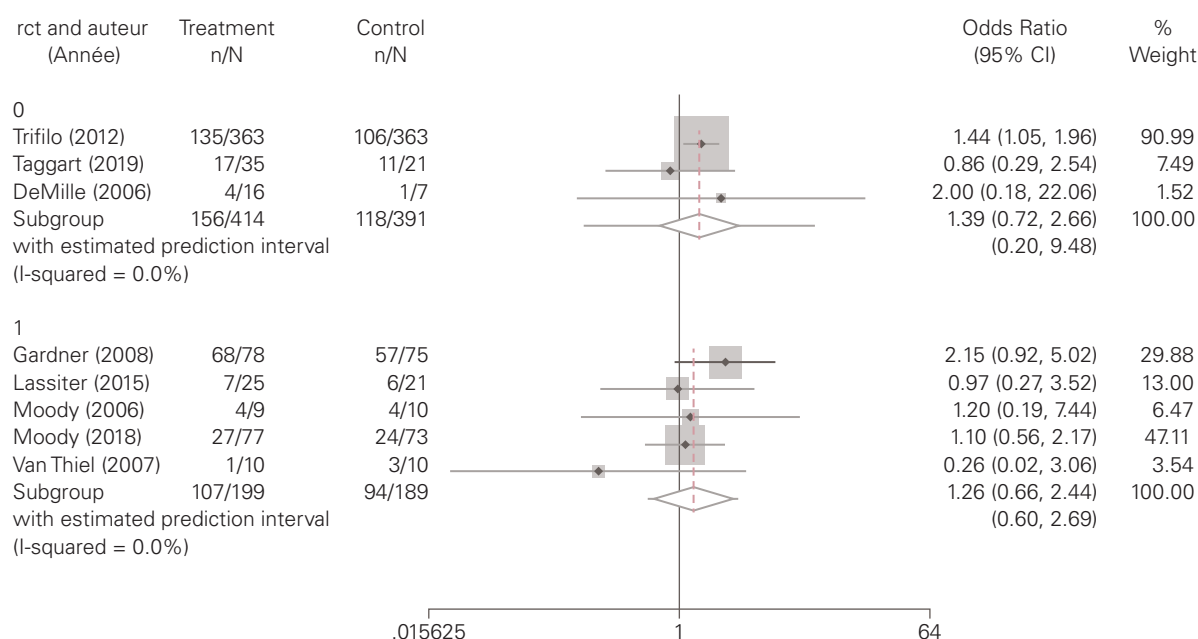
Elle est réalisée à l'aide du *forest plot* (graphique en forêt). L'effet du traitement dans chaque essai et l'effet global du traitement sont positionnés sur le même graphique (Figure 1). Nous présentons l'exemple, mis à jour, de la relation entre type alimentation (faible inoculum bactérien ou non) et survenue d'infection chez les patients neutropéniques⁶ selon deux types d'étude : trois études randomisées (groupe 0) vs cinq études d'observation (groupe 1).

L'OR de chaque essai (carré), est représenté avec son intervalle de confiance à 95% (ligne horizontale). L'OR global (losange et ligne verticale en pointillé) et son intervalle de confiance à 95% (extrémité du losange) sont également représentés. D'autres informations figurent sur le graphique, tels que le I^2 qui quantifie l'hétérogénéité, les effectifs de chaque groupe (intervention vs control) et le poids de chaque étude dans la MA. Une pondération utilisée en MA est celle de l'inverse de la variance [50]. On peut noter que pour le sous-groupe des études d'observation, l'étude de Trifilio [51] a un poids de 90,99%.

Les « moustaches » de part et d'autre du losange représentent l'intervalle de confiance de la distribution prédictive approximative d'un futur essai réalisé dans un contexte similaire à celui de la MA¹⁶ [52,53]. On note qu'ici, cet intervalle est très large pour les études d'observation, allant de 0,20 à 9,48 (vs 0,60 à 2,69 pour les

16- Celui-ci n'est pas estimable s'il y a moins de trois études.

Figure 1 – Méta-analyse de la relation entre type alimentation et survenue d'infection, chez les patients neutropéniques.



Note: Weights are from random-effects model

N : nombre de patients ; n : nombre de patients infectés ; weight : poids relatif de chaque étude.

études randomisées). Le trait vertical passant par 1 correspond à l'égalité entre les deux interventions. Un OR de part ou d'autre de cette ligne est en faveur ou en défaveur de l'intervention (ici, il semble être en faveur d'une alimentation à faible inoculum bactérien). La MA est réalisée en utilisant un modèle à effet aléatoire avec une estimation de τ^2 (variance de la taille de l'effet) via la méthode de Hartung-Knapp-Sidik-Jonkman [54] (qui est plus adaptée aux MA incluant peu d'études que la méthode classique de DerSimonian et Laird). Pour les études randomisées, l'OR global est de 1,39, avec un intervalle de confiance à 95% allant de 0,72 à 2,66. Pour les études d'observation, l'OR global est de 1,26, avec un intervalle de confiance à 95% allant de 0,66 à 2,44. Donc quel que soit le type d'étude, randomisé ou non, l'alimentation à faible inoculum bactérien n'a pas montré son efficacité dans la prévention des infections chez les patients neutropéniques.

Un point important est qu'ici, uniquement par simplification, nous avons traité les études d'observation comme si elles étaient du même type, ce qui n'est pas le cas. Une des études est une pseudo cohorte rétrospective [51]; une autre est de type « avant-après » [55]; une autre est une étude pilote, prospective [56]. Il est recommandé de réaliser l'analyse pour chacun des sous-groupes d'étude d'observation [57].

Exploration et analyse de l'hétérogénéité

La recherche d'une hétérogénéité entre les études incluses dans une MA est un point important. En effet, le résultat global obtenu ne peut s'interpréter qu'en l'absence d'hétérogénéité entre les différentes études. Il existe une hétérogénéité lorsque la variation des résultats des essais dépasse la simple fluctuation d'échantillonnage. L'hétérogénéité peut être clinique, avec une différence relative, à la définition de la maladie; aux critères d'inclusion des patients; aux méthodes diagnostiques; à la durée de suivi; aux doses médicamenteuses; aux critères de jugement (par exemple, colonisation vs infection), etc. L'hétérogénéité peut être statistique, avec une discordance des résultats des différentes études. Elle peut être liée à un biais de publication, à une méthodologie inappropriée...

L'hétérogénéité statistique peut être détectée graphiquement, via le *forest plot*, avec des intervalles de confiance de l'effet observé dans les différentes études qui ne se chevauchent pas suffisamment. L'hétérogénéité statistique peut être formellement testée, par exemple via le Q de Cochran ou le I^2 de Higgins [50,51]. Ce dernier estime la non-concordance (*inconsistency*) entre les études. Il prend en compte le test Q de Cochran et le nombre d'études incluses dans la MA. Schématiquement, il représente la proportion de variation de l'effet traitement entre les études incluses dans la MA. Des seuils ont été proposés: $I^2 < 0,25$ = hétérogénéité faible; I^2 compris entre 0,25 et 0,5 = hétérogénéité modérée; $I^2 > 0,5$ = hétérogénéité importante. Ces tests d'hétérogénéité

étant peu puissants, il faut s'assurer, même dans le cas de résultats non significatifs, qu'aucun des essais ne présente de résultats extrêmes. Dans le cas de résultats extrêmes, une analyse de sensibilité peut permettre de conforter le résultat observé.

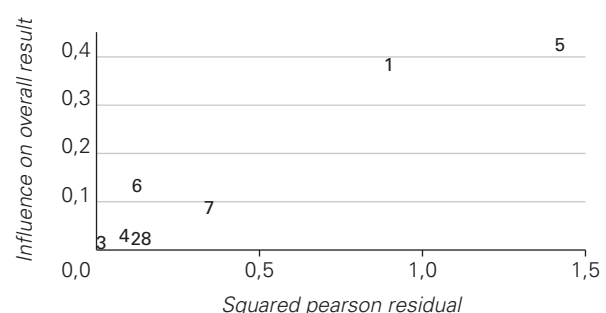
Que faire en cas d'hétérogénéité? La démarche classique est d'utiliser un modèle à effet aléatoire (par exemple, utilisant la méthode de DerSimonian et Laird). En l'absence d'hétérogénéité, on pourra utiliser un modèle à effet fixe (modèle de Mantel et Haenszel ou de Peto par exemple) plus robuste pour la mesure d'événements rares. En présence d'hétérogénéité, il faut essayer de l'expliquer. Les sources d'hétérogénéité sont identifiées par l'analyse qualitative des études incluses. Les approches peuvent graphiques, statistiques (via des tests) ou mixtes. La recherche d'études pouvant entraîner une hétérogénéité est possible en inspectant le graphique des OR (ou autre mesure d'effet) [58]. Baujat et al. [59] ont ainsi proposé une méthode graphique qui permet de visualiser facilement les essais les plus hétérogènes et les plus influents de la MA sur un graphique en deux dimensions (Figure 2). On représente: (i) en abscisse, la contribution de chaque étude à la statistique globale d'hétérogénéité; (ii) en ordonnée, la différence standardisée de l'effet global du traitement, avec et sans chaque étude (cette quantité décrit l'influence de chaque étude sur l'effet global du traitement). Les études qui s'éloignent de la masse des observations sont suspectes d'être source d'hétérogénéité.

Dans notre exemple, les études numérotées « 1 » et « 5 » sont suspectes d'être source d'hétérogénéité.

L'hétérogénéité peut aussi être explorée à l'aide d'assez nombreuses autres méthodes qui sont présentées par l'Annexe I – Autres méthodes d'étude de l'hétérogénéité¹⁷. Il s'agit entre autres du graphique radial de Galbraith [60], du graphique de Labbe [61] des analyses en sous-groupes avec méta-régression prenant en compte par exemple l'année de publication, du *funnel plot* qui

17- L'annexe est disponible sur le site internet de la revue pour les abonnés numériques (www.hygienes.net) ou sur simple demande à l'auteur correspondant.

Figure 2 – Graphique de Baujat (Méta-analyse de la relation entre type alimentation et survenue d'infection, chez les patients neutropéniques).



suggère un biais de publication, des tests d'Enger et de Begg, de la méthode *trim-and-fill*, etc.

Analyse de sensibilité: méta-analyse *leave-one-out*

Cette procédure peut être vue comme une analyse de sensibilité¹⁸. L'objectif est de montrer comment chaque étude affecte l'estimation de l'effet global procuré par la MA. La MA *leave-one-out* consiste à effectuer une MA en omettant une étude. Le processus est répété jusqu'à ce que toutes les études soient omises. Il est possible d'omettre plus d'une étude à la fois, mais les calculs deviennent volumineux, en raison du nombre élevé de combinaisons possibles¹⁹.

Les résultats concernant notre exemple sont dans la **Figure 3**. On note pour les études d'observation, que l'omission de l'étude de Trifilio [51] dans la MA conduit à une estimation très imprécise de la taille de l'effet, avec un OR de 0,99, mais surtout un intervalle de confiance allant de 0,002 à 585,8²⁰. Pour les études randomisées, l'omis-

sion d'une des études, quelles qu'elles soient, ne semble pas beaucoup influencer l'estimation de l'effet global.

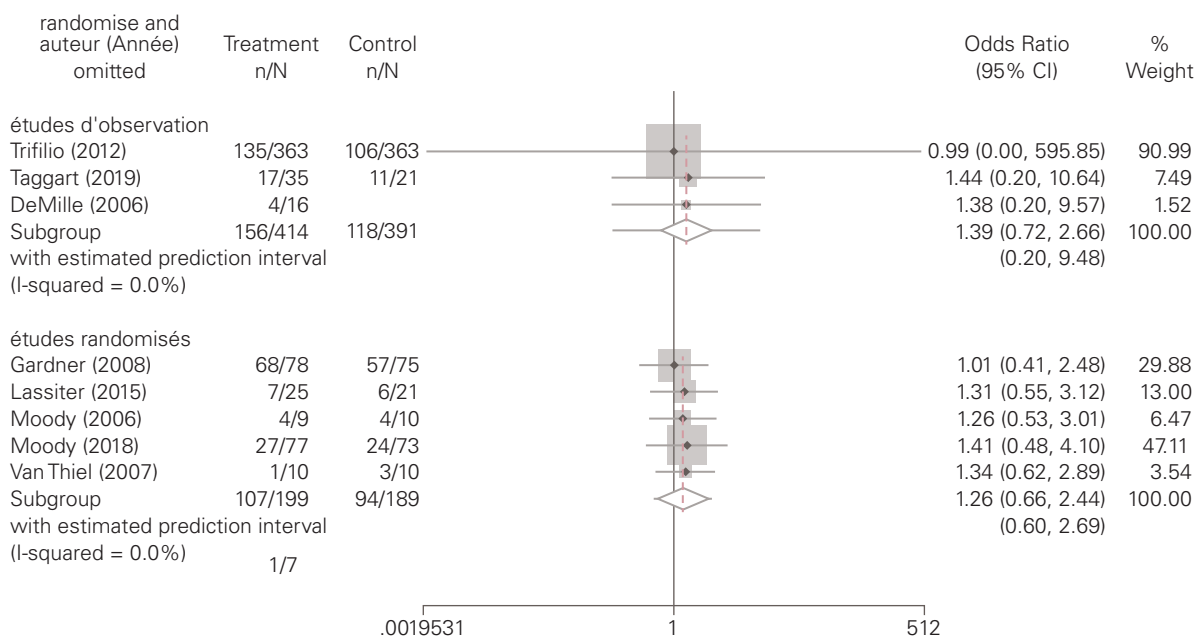
Méta-analyse cumulative

La MA cumulative consiste à effectuer une MA en introduisant les études les unes après les autres, afin de suivre l'évolution de l'effet global. Le processus est répété jusqu'à ce que toutes les études soient incluses. L'ordre d'introduction le plus fréquent est chronologique (par date de publication), afin de détecter une tendance temporelle, mais d'autres ordres sont possibles tels que la précision (inverse de la variance) [63]. Cette dernière variante de MA cumulative permet aussi de détecter les biais de publication [63,64]. La MA cumulative n'est pas une réactualisation de la MA. La réactualisation d'une MA consiste à estimer un nouvel effet global, chaque fois qu'une nouvelle étude est disponible. Des analyses statistiques appropriées telles que la MA séquentielle, doivent être utilisés.

Les résultats de notre exemple figurent dans le **Tableau II**. On note, malgré l'imprécision des intervalles de confiance liée au modèle utilisé, que les estimations sont similaires pour les études randomisées et pour les études d'observation, à la dernière étude publiée. Il n'y a pas assez d'études dans la MA pour voir se dégager une éventuelle tendance temporelle.

Nous avons illustré cette première partie à l'aide d'un exemple de MA incluant peu d'études, ce qui reflète la réalité. En effet, selon l'étude Davey et al. [65], conduite

Figure 3 – Analyse *leave-one-out* de la méta-analyse estimant la relation entre type d'alimentation et survenue d'infection, chez les patients neutropéniques.



Note : Weights are from random-effects model, expressed relative to the total weight in the overall model

N : nombre de patients ; n : nombre de patients infectés ; weight : poids relatif de chaque étude.

sur la base de données *Cochrane*, le nombre médian d'études par méta-analyse n'est que de trois sur un total de 22 453 méta-analyses concernant la santé.

La seconde partie de cette mise au point méthodologique sera consacrée aux aspects particuliers de la méta-analyse. On abordera entre autres :

- l'agrégation ou non des données : MA sur données individuelles [66]. C'est la référence. Elle est plus chronophage et un peu plus compliquée à mettre en œuvre que la MA sur données agrégées ;
- la MA en réseau [67], est utilisée lorsque les interventions n'ont pas été comparées, mais ont été étudiées par rapport à un comparateur commun ;
- la MA bayésienne [68] qui s'applique à tous les aspects statistiques de la MA. ■

Tableau II – Méta-analyse cumulative : la relation entre type alimentation et survenue d'infection, chez les patients neutropéniques.

Études randomisées et auteurs (année)	Odds ratio	[95% intervalle de confiance]	% poids
Études d'observation			
DeMille (2006)	2,000	0,181	1,52
Trifilio (2012)	1,443	0,196	90,99
Taggart (2019)	1,388	0,724	7,49
Études randomisées			
Moody (2006)	1,200	0,194	6,47
Van Thiel (2007)	0,698	0,000	3,54
Gardner (2008)	1,389	0,154	29,88
Lassiter (2015)	1,408	0,483	13,00
Moody (2018)	1,265	0,655	47,11

Références

- 1- Simpson EH. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1951;13:238-241. Doi : 10.1111/j.2517-6161.1951.tb00088.x.
- 2- Altman DG, Deeks JJ. Meta-analysis, Simpson's paradox, and the number needed to treat. *BMC Med Res Methodol*. 2002;2:3. Doi : 10.1186/1471-2288-2-3.
- 3- Goltz HH, Smith ML. Yule-Simpson's paradox in research. *Practical Assessment, Research & Evaluation*. 2010;15. Doi : 10.7275/dgcc-jv81.
- 4- Delahaye JP. Au pays des paradoxes. Paris: Belin; 2008. 192 p.
- 5- Hanley JA, Thériault G. Simpson's paradox in meta-analysis. *Epidemiology*. 2000;11(5):613-614. Doi : 10.1097/00001648-200009000-00022.
- 6- Borenstein M, Hedges LV, Higgins JPT et al. Introduction to meta-analysis. Hoboken: Wiley; 2009. 452 p. Doi : 10.1002/9780470743386
- 7- Baker SG, Kramer BS. Good for women, good for men, bad for people: Simpson's paradox and the importance of sex-specific analysis in observational studies. *J Womens Health Gend Based Med*. 2001;10(9):867-872. Doi : 10.1089/152460901753285769.
- 8- Plackett RL. Karl Pearson and the Chi-squared test. *International Statistical Review / Revue Internationale de Statistique*. Non-Stationary Random Process for Large-Scale Failure and Recovery of Power Distribution. 1983;51(1):59-72. Doi : 10.2307/1402731.
- 9- Pearson K. Report on certain enteric fever inoculation statistics. *Br Med J*. 1904;2:1243-1246.
- 10- Beecher HK. The powerful placebo. *JAMA*. 1955;159(17):1602-1606. Doi : 10.1001/jama.1955.02960340022006.
- 11- Glass GV. Primary, secondary, and meta-analysis of research. *Educational Researcher*. 1976;5(10):3-8. Doi : 10.2307/1174772.
- 12- Schmidt FL, Hunter JE. Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*. 1977;62(5):529-540. Doi : 10.1037/0021-9010.62.5.529.
- 13- O'Rourke K. An historical perspective on meta-analysis: dealing quantitatively with varying study results. *J R Soc Med*. 2007;100(12):579-582. Doi : 10.1258/jrsm.100.12.579.
- 14- Hunt M. How science takes stock: the story of meta-analysis. New York: Fondation Russell Sage; 1998. 256 p.
- 15- Rayleigh J. Report of fifty-four meeting of the British Association for the Advancement of Science. London: John Murray; 1885. p. 3-23.
- 16- Chan ME, Arvey RD. Meta-analysis and the development of knowledge. *Perspect Psychol Sci*. 2012;7(1):79-92. Doi : 10.1177/1745691611429355.
- 17- Kuhn TS. The structure of scientific revolutions. Chicago: University of Chicago Press; 1962. 264 p.
- 18- Kuhn TS. La structure des révolutions scientifiques. Paris: Flammarion; 1983. 284 p.
- 19- Kuhn TS. The structure of scientific revolutions: 50th anniversary edition. Chicago: University of Chicago Press; 2012. 264 p.
- 20- Shadish WR, Lacy JD. The meta-analytic big bang. *Res Synth Methods*. 2015;6:246-264. Doi : 10.1002/jrsm.1132.
- 21- Rosenthal R. Combining results of independent studies. *Psychological Bulletin*. 1978;85(1):185-193. Doi : 10.1037/0033-2909.85.1.185.
- 22- Grewal A, Kataria H, Dhawan I. Literature search for research planning and identification of research problem. *Indian J Anaesth*. 2016;60(9):635-639. Doi : 10.4103/0019-5049.190618.
- 23- PubMed. NCBI. Accessible à : <https://www.ncbi.nlm.nih.gov/pubmed/> (Consulté le 10-03-20).
- 24- ScienceDirect. Accessible à : <https://www.sciencedirect.com/> (Consulté le 10-03-20).
- 25- Cochrane Library. Accessible à : <https://www.cochranelibrary.com/> (Consulté le 10-03-20).
- 26- Inist-CNRS. Accessible à : <https://www.istex.fr/rechercher/> (Consulté le 10-03-20).
- 27- US National Library of Medicine. ClinicalTrials.gov. Accessible à : <https://clinicaltrials.gov/> (Consulté le 10-03-20).
- 28- EU Clinical Trials Register. Accessible à : <https://www.clinicaltrials-register.eu/> (Consulté le 10-03-20).
- 29- Schöpfel J. Vers une nouvelle définition de la littérature grise. *Cahiers de la Documentation*. 2012;66: 14-24. Accessible à : https://archivesic.ccsd.cnrs.fr/sic_00794984/document. (Consulté 11-03-2020).
- 30- Adams J, Hillier-Brown FC, Moore HJ et al. Searching and synthesising "grey literature" and "grey information" in public health: critical reflections on three case studies. *Syst Rev*. 2016;5:164. Doi : 10.1186/s13643-016-0337-y.
- 31- Paez A. Gray literature: an important resource in systematic reviews. *J Evid Based Med*. 2017;10(3):233-240. Doi : 10.1111/jebm.12266.
- 32- OpenGrey. Accessible à : <http://www.opengrey.eu/> (Consulté le 10-03-20).
- 33- Aho Glélé LS, Aho S. Élaboration de recommandations : comment utiliser la méthode Grade ? *Hygiènes*. 2018;5:213-218. Doi : 10.25329/hy_xxvi_5-8.

- 34- Cochrane Library. RoB 2: a revised Cochrane risk-of-bias tool for randomized trials. Accessible à : <https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials> (Consulté le 10-03-20).
- 35- Sterne JA, Hernán MA, Reeves BC et al. Robins-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355: i4919. Doi : 10.1136/bmj.i4919.
- 36- Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155:529-536. Doi : 10.7326/0003-4819-155-8-201110180-00009.
- 37- Hill C, Com-Noué C, Kramar A. Analyse statistique des données de survie. Paris: INSERM : Médecine-sciences Flammarion; 1990. 190 p.
- 38- Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates; 1988. 567 p.
- 39- Hedges LV. Distribution Theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics*. 1981;6:107-128. Doi : 10.2307/1164588.
- 40- Hedges LV, Hedges LV, Olkin I. Statistical methods for meta-analysis. Elsevier Science; 1985. 369 p.
- 41- Cochrane Library. RevMan 5. Accessible à : <https://community.cochrane.org/help/tools-and-software/revman-5> (Consulté le 10-03-2020).
- 42- Ellis PD. The essential guide to effect sizes: statistical power, meta-analysis, and the interpretation of research results. 1st ed. Cambridge ; New York: Cambridge University Press; 2010. 192 p.
- 43- Grissom RJ, Kim JJ. Effect sizes for research: univariate and multivariate applications, Second Edition. 2nd ed. New York: Routledge; 2012. 41 p.
- 44- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177-188.
- 45- Hunter JE, Schmidt FL. Fixed effects vs. random effects meta-analysis models: implications for cumulative research knowledge. *International Journal of Selection and Assessment*. 2000;8: 275-292.
- 46- Bach J-F. L'erreur scientifique. Paris: Académie des sciences; 2011. Accessible à : https://www.academie-sciences.fr/archivage_site/academie/membre/s210611_bach.pdf (Consulté le 10-03-20).
- 47- Serghiou S, Goodman SN. Random-effects meta-analysis: summarizing evidence with caveats. *JAMA*. 2019;321:301-302. Doi : 10.1001/jama.2018.19684.
- 48- DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. *Contemp Clin Trials*. 2015;45:139-145. Doi : 10.1016/j.cct.2015.09.002.
- 49- Doi SAR, Barendregt JJ, Khan S, et al. Advances in the meta-analysis of heterogeneous clinical trials I: The inverse variance heterogeneity model. *Contemp Clin Trials*. 2015;45:130-138. Doi : 10.1016/j.cct.2015.05.009.
- 50- Hartung J, Knapp G, Sinha BK. Statistical meta-analysis with applications. 1st ed. Hoboken, N.J: Wiley-Blackwell; 2008. 248 p.
- 51- Trifilio S, Helenowski I, Giel M, et al. Questioning the role of a neutropenic diet following hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2012;18:1385-1390. Doi : 10.1016/j.bbmt.2012.02.015.
- 52- Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A Stat Soc*. 2009;172: 137-159.
- 53- Nagashima K, Noma H, Furukawa TA. Prediction intervals for random-effects meta-analysis: A confidence distribution approach. *Stat Methods Med Res*. 2019;28:1689-1702. Doi : 10.1177/0962280218773520.
- 54- Röver C, Knapp G, Friede T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Med Res Methodol*. 2015;15. Doi : 10.1186/s12874-015-0091-1.
- 55- Taggart C, Neumann N, Alonso PB, et al. Comparing a neutropenic diet to a food safety-based diet in pediatric patients undergoing hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2019;25:1382-1386. Doi : 10.1016/j.bbmt.2019.03.017.
- 56- DeMille D, Deming P, Lupinacci P, et al. The effect of the neutropenic diet in the outpatient setting: a pilot study. *Oncol Nurs Forum*. 2006;33:337-343. Doi : 10.1188/ONF06.337-343.
- 57- Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Hoboken, NJ: Wiley-Blackwell; 2019. 728 p.
- 58- Galbraith RF. A note on graphical presentation of estimated odds ratios from several clinical trials. *Stat Med*. 1988;7:889-894. Doi : 10.1002/sim.4780070807.
- 59- Baujat B, Mahé C, Pignon J-P, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Stat Med*. 2002;21:2641-2652. Doi : 10.1002/sim.1221.
- 60- Galbraith RF. Some applications of radial plots. *Journal of the American Statistical Association*. 1994;89: 1232-1242. Doi : 10.2307/2290987.
- 61- L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med*. 1987;107:224-233. Doi : 10.7326/0003-4819-107-2-224.
- 62- Martin GE. Counting: the art of enumerative combinatorics. Soft-cover reprint of hardcover 1st ed. 2001. New York, NY: Springer-Verlag New York Inc.; 2010. 252 p.
- 63- Atakpo P, Vassar M. Cumulative meta-analysis by precision as a method to evaluate publication bias. *Journal of Dermatological Science*. 2016;83: 251-253. Doi : 10.1016/j.jdermsci.2016.06.001.
- 64- Leimu R, Koricheva J. Cumulative meta-analysis: a new tool for detection of temporal trends and publication bias in ecology. *Proc Biol Sci*. 2004;271:1961-1966. Doi : 10.1098/rspb.2004.2828.
- 65- Davey J, Turner RM, Clarke MJ, et al. Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: a cross-sectional, descriptive analysis. *BMC Med Res Methodol*. 2011;11:160. Doi : 10.1186/1471-2288-11-160.
- 66- Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340. Doi : 10.1136/bmj.c221.
- 67- Dias S, Ades AE, Welton NJ, et al. Network meta-analysis for decision-making. John Wiley & Sons; 2018. 488 p.
- 68- Sutton AJ, Abrams KR. Bayesian methods in meta-analysis and evidence synthesis. *Stat Methods Med Res*. 2001;10: 277-303. Doi : 10.1177/096228020101000404.

La partie II de cette rubrique sera diffusée dans le numéro 3 de Hygiènes, volume XXVIII – Juin 2020.