

RAPPORT DE SYNTHÈSE

**Les comparaisons indirectes
Méthodes et validité**

Juillet 2009

Ce rapport de synthèse est téléchargeable sur
www.has-sante.fr

Haute Autorité de Santé
Service communication
2 avenue du Stade de France - F 93218 Saint-Denis La Plaine CEDEX
Tél. :+33 (0)1 55 93 70 00 - Fax :+33 (0)1 55 93 74 00

Sommaire

Résumé.....	6
Définition	6
Contexte et intérêts potentiels	6
Méthodes statistiques.....	6
Validité des comparaisons indirectes	7
L'essai de comparaison directe est-il un « étalon-or » ?.....	7
Utilisation actuelle et acceptation par les agences	7
Lecture critique	7
Conclusions - Points clés	7
1. Introduction	9
1.1 Définitions et généralités.....	9
1.2 Objectifs	10
2. Contexte	11
2.1 Manque de comparaisons directes entre traitements actifs	11
2.2 Prédilection pour les essais contre placebo	12
2.3 Non transitivité des tests statistiques	13
2.4 Essais cliniques de non infériorité.....	13
3. Méthodes de la recherche documentaire et critères de sélection.....	14
3.1 Sources d'informations.....	14
3.2 Stratégie de la recherche bibliographique.....	14
3.3 Critères de sélection des publications.....	15
4. Les différentes techniques permettant de réaliser des comparaisons indirectes...16	
4.1 Les techniques inappropriées	16
4.2 Les comparaisons indirectes ajustées	18
4.3 L'approche mixte des comparaisons indirectes et directes	22
4.4 Les méthodes basées sur une estimation globale d'un réseau d'essais.....	24
4.5 « Bayesian network meta-analysis »	27
4.6 Estimation à l'aide d'un modèle linéaire mixte	30
4.7 Méta-régression	33
4.8 Méthodes pour les critères de jugement continu	35
4.9 Comparaison des différentes méthodes.....	35
5. Utilisation actuelle des comparaisons indirectes	37
5.1 Données bibliographiques.....	37
5.2 Quelques publications récentes	37
5.3 Utilisation dans des rapports d'évaluation technologique ou des recommandations	39
6. Validité des comparaisons indirectes	40
6.1 Qu'est-ce que la validité des comparaisons indirectes ?.....	40
6.2 Les études de validité	40
6.3 La comparaison directe reste-t-elle l'étalon-or ?	42

6.4 Perspectives	44
7. Positionnement des instances réglementaires	45
8. Interprétation des résultats	46
9. Guide de lecture critique	47
9.1 Situation optimale.....	47
9.2 Guide de lecture critique	47
9.3 Études de cas	50
10. Conclusion.....	55
11. Points clés	56
Annexe 1. Liste des participants.....	57
Annexe 2. Bibliographie.....	58
Annexe 3. Recherche de recommandations, rapports d'évaluation technologique, procédures méthodologiques sur Internet : liste des sites consultés.....	64
Index.....	65

Liste des abréviations

AMM	Autorisation de mise sur le marché
HAS	Haute Autorité de Santé
FDA	<i>Federal Drug Administration</i>
EMA	<i>European Medicines Agency</i>
AFSSAPS	Agence française de sécurité sanitaire des produits de santé
NICE	<i>National Institute for Health and Clinical Excellence</i>

Résumé

Définition

Les techniques de comparaison indirecte ont pour objectif d'estimer l'effet d'un traitement A par rapport à un traitement B à partir des résultats des essais de A et de B *versus* un même contrôle (placebo ou traitement actif). Ces techniques réalisent donc une extrapolation se basant sur l'hypothèse que les effets qu'auraient A et B *versus* contrôle dans les conditions d'un essai « face à face » sont identiques à ceux observés dans les essais à la base de la comparaison indirecte.

La comparaison directe consiste à comparer A et B directement dans un essai « face à face ».

Les résultats de comparaisons directes peuvent être combinés avec ceux de comparaisons indirectes à l'aide d'une approche mixte, « Mixed Treatment Comparison (MTC) ».

Contexte et intérêts potentiels

Les comparaisons indirectes sont de plus en plus utilisées car elles permettent d'aborder des questions fréquentes en évaluation concernant la hiérarchie d'efficacité et/ou de sécurité des traitements concurrents.

Dans de nombreux domaines thérapeutiques, les traitements concurrents sont évalués préférentiellement *versus* placebo et les comparaisons directes sont rares. Les comparaisons indirectes peuvent permettre d'estimer l'efficacité et/ou la sécurité relatives des traitements entre eux dans l'attente d'éventuels essais de comparaisons directes.

Dans d'autres domaines thérapeutiques, les comparaisons directes sont impossibles à effectuer, pour des raisons éthiques (pathologies engageant le pronostic vital ou fonctionnel du patient à court terme par exemple) ou de faisabilité dans le cas des maladies rares. Seules les comparaisons indirectes permettent alors d'estimer l'efficacité et/ou la sécurité relatives des traitements.

De plus, il est possible d'émettre certaines réserves sur le caractère « étalon-or » des essais de comparaison directe qui sont souvent de moindre qualité méthodologique que les essais *versus* placebo (insuffisance d'aveugle, manque de sensibilité,...).

L'approche par comparaison directe peut aussi faire apparaître les problèmes de non transitivité de certains tests statistiques lors des comparaisons deux à deux.

Méthodes statistiques

Des méthodes adaptées de comparaison indirecte ont été développées ces dernières années. De nombreux travaux ayant un objectif de comparaison indirecte utilisent encore des méthodes inadaptées comme les comparaisons naïves d'estimations ponctuelles ou de bras actifs issus d'essais contrôlés distincts. Ces méthodes ne sont pas licites et les travaux les employant ne méritent pas l'appellation de comparaison indirecte.

Les méthodes adaptées de comparaisons indirectes se regroupent en deux catégories :

- les méthodes qui ne comparent que deux traitements simultanément, comme les comparaisons indirectes ajustées. La comparaison de plusieurs traitements implique la réalisation de plusieurs comparaisons « deux à deux » ;
- les méthodes qui intègrent simultanément plusieurs traitements à l'aide d'un modèle statistique complexe. Elles modélisent le réseau des comparaisons possibles entre tous les traitements concurrents. Ces méthodes intègrent les essais comparant deux traitements et les essais multibras. Différentes méthodes d'estimation des paramètres de ces modèles ont été proposées : méthodes Bayésiennes, modèles linéaires mixtes, méta-régression.

Les performances relatives de ces différentes méthodes ont été peu étudiées. Un travail de simulation a montré que toutes ces méthodes étaient non biaisées.

En l'état actuel des connaissances, la méta-analyse Bayésienne en réseau semble être la méthode la plus intéressante car flexible et séduisante sur un plan théorique.

Pour éviter des choix arbitraires dans les données traitées, les calculs de comparaisons indirectes, quelle que soit la méthode utilisée, doivent porter sur les résultats de méta-analyses qui supposent une recherche exhaustive des données et une sélection des essais sur leur aptitude à faire une estimation non biaisée.

Validité des comparaisons indirectes

Il n'existe qu'un seul travail de grande ampleur de comparaisons empiriques des résultats issus de comparaisons indirectes à ceux issus de comparaisons directes¹. Cette étude de validité portait sur la méthode des comparaisons indirectes ajustées. Le critère de discordance utilisé était l'existence d'une différence statistiquement significative entre les estimations de l'effet traitement des deux approches. Parmi 44 cas étudiés (utilisation de deux traitements pour une pathologie donnée), seulement 3 cas de discordances ont été mis en évidence. Les différences allaient dans les deux sens. Aucun facteur pronostique de la discordance n'a été identifié.

Les autres études de validité disponibles n'abordent qu'un seul champ thérapeutique et, de ce fait, ont plus valeur d'études de cas que d'études empiriques de validation.

Au final, les premières données disponibles sont plutôt en faveur de la validité des comparaisons indirectes mais d'autres études empiriques seraient souhaitables avant de conclure définitivement sur ce point.

L'essai de comparaison directe est-il un « étalon-or » ?

Les études empiriques de validation des comparaisons indirectes partent du postulat couramment admis que l'essai « face à face » est un étalon-or et qu'il donne une estimation à l'abri des biais. Cependant, aucune étude empirique n'a validé ce postulat. Ce point mérite réflexion car les essais de comparaison directe entre traitements actifs sont fréquemment réalisés en ouvert et il existe des exemples de non reproductibilité des résultats en fonction du promoteur de l'essai.

Utilisation actuelle et acceptation par les agences

L'utilisation des techniques de comparaisons indirectes n'est plus rare actuellement. Plus d'une centaine de publications (articles scientifiques, recommandations, rapports d'évaluation de technologies de santé) recourant à cette approche sont maintenant disponibles, apportant une expérience empirique substantielle. Le retour d'expérience reste cependant faible.

Au niveau de l'enregistrement, aucune agence (FDA, EMEA, AFSSAPS) ne s'est formellement positionnée sur l'acceptabilité des argumentations basées sur des résultats de comparaisons indirectes.

Au niveau des organismes d'évaluation ou d'élaboration des recommandations, seul le NICE mentionne dans ses procédures la possibilité de recourir aux comparaisons indirectes, de manière exploratoire et au cas par cas. Il est explicitement fait mention que les comparaisons directes restent l'approche de base.

Lecture critique

L'étude théorique des méthodes proposées et l'expérience acquise avec les premières utilisations de ces techniques permettent de lister les points cruciaux à analyser dans la lecture critique des résultats de comparaisons indirectes. Un guide de lecture opérationnel est proposé à l'issue du travail bibliographique.

Conclusions - Points clés

¹ Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326(7387):472.

Pour pallier le manque d'essais comparatifs des traitements actifs entre eux dans de très nombreux champs de la médecine, le recours aux comparaisons indirectes est envisageable chaque fois que la question de la hiérarchie d'efficacité ou de sécurité des traitements concurrents se pose.

Cette pratique peut permettre, dès l'enregistrement, d'évaluer le positionnement d'un nouveau traitement par rapport aux traitements existants et cela sans devoir attendre plusieurs années après la commercialisation les données de comparaisons plus conventionnelles issues d'essais « face à face ».

Les résultats de comparaison directe ne devront pas reléguer au second plan ceux des comparaisons indirectes ; ces deux estimations devront être soigneusement confrontées et si possible combinées à l'aide d'une approche mixte (*mixed treatment comparison*).

La mise en œuvre de cette approche devrait s'effectuer par l'intermédiaire de méta-analyses en réseau, les mieux à même de dessiner la hiérarchie complète d'un champ thérapeutique donné (envisageant si possible toutes les ressources thérapeutiques disponibles). Les autres techniques valides sont aussi utilisables en l'absence pour l'instant de comparaisons approfondies des performances relatives des différentes méthodes.

Une seule grande étude sur la validité des comparaisons indirectes étant pour l'instant disponible, de nouvelles vérifications empiriques devront être entreprises en utilisant les méthodes modernes de comparaison indirecte. Des efforts de recherche devront être soutenus pour étudier des questions concernant la fiabilité des comparaisons directes et les performances des différentes méthodes de comparaison indirecte proposées.

La diffusion des méthodes de comparaisons indirectes pourrait permettre de mieux définir les stratégies de hiérarchisation des traitements entre eux à partir de données factuelles.

Conséquences pour la pratique

- Les comparaisons indirectes sont réalisables et les bases d'une évaluation critique de leur qualité existent.
- Le recours aux comparaisons indirectes est envisageable chaque fois que la question de la hiérarchie d'efficacité ou de sécurité des traitements concurrents se pose.
- L'acceptabilité des comparaisons indirectes doit être envisagée au cas par cas en prenant en compte le domaine thérapeutique et les conséquences de l'absence d'essais de comparaison directe.

Conséquences pour la recherche

- Des études comparant les performances des différentes techniques statistiques doivent être réalisées pour identifier la ou les méthode(s) préférentielle(s).
- Il convient aussi de réaliser des études empiriques de validation des essais de comparaison directe afin de mieux cerner leur aptitude à être le « gold standard » de la comparaison de deux traitements actifs.
- De nouvelles études empiriques de validation des comparaisons indirectes en réseau seraient aussi nécessaires.

1. Introduction

1.1 Définitions et généralités

Les comparaisons directes et indirectes sont deux approches visant à comparer l'efficacité (ou la sécurité) de deux traitements actifs A et B, afin de répondre à la question A est-il supérieur ou non-inférieur à B ?

Une comparaison directe est effectuée quand les deux traitements A et B sont comparés directement dans un essai thérapeutique randomisé. Cet essai est aussi appelé essai contrôlé *versus* traitement actif (« active-control »), ou comparaison face à face (« head to head comparison »). Le résultat d'un essai de ce type est donc une mesure de l'effet de A par rapport à B (risque relatif, odds ratio, différence des risques, hazard ratio, différence absolue ou relative des moyennes, etc.) et un test statistique de la réalité statistique d'une différence d'effet entre A et B. Un essai comparant directement un nouveau traitement à un traitement standard concluant montre ainsi la supériorité ou la non infériorité (en fonction de l'objectif de l'essai) du nouveau sur le standard.

Dans l'approche par comparaison indirecte, l'efficacité des deux traitements A et B est comparée par l'intermédiaire de l'efficacité respective des deux traitements par rapport à un contrôle commun, le plus souvent un placebo². L'idée est de répondre à la question A est-il supérieur ou inférieur à B à l'aide du résultat de la comparaison de A au contrôle et du résultat de la comparaison de B au même contrôle (Figure 1). Cette comparaison de A et B est indirecte car elle passe par un troisième traitement, le contrôle commun. Les essais A *versus* contrôle et B *versus* contrôle fournissent les comparaisons de base à partir desquelles sera effectuée la comparaison indirecte de A et de B.

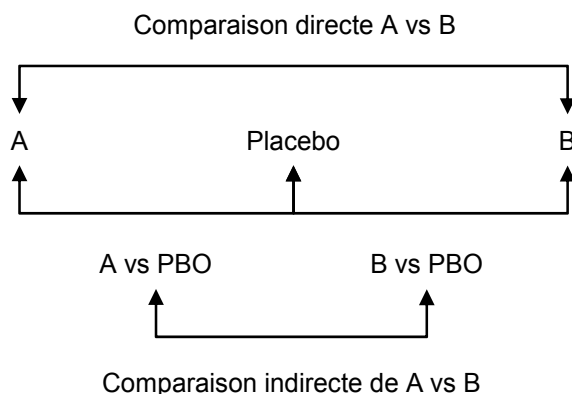


Figure 1 - Illustration des principes des comparaisons directes et indirectes. Dans cet exemple, le traitement contrôlé des comparaisons de base est le placebo (PBO)

A titre d'exemple, lorsque le critère de jugement est binaire, les résultats issus d'une comparaison directe peuvent être exprimés sous forme d'un risque relatif. Ce risque relatif mesure la réduction supplémentaire de fréquence de l'événement, critère de jugement, qu'apporte le traitement A par rapport à celle qu'apporte le traitement B (B étant efficace et réduisant déjà la fréquence de l'événement).

Le Tableau 1 illustre les résultats typiques d'un essai de comparaison directe de A vs B.

Tableau 1 - Résultats d'un essai de comparaison directe de A *versus* B

	Traitement A n=1 000	Traitement B n=1 000	Risque relatif [IC 95 %]	Valeur de p
Décès	127 (12,7 %)	234 (23,4 %)	0,54 [0,45 ; 0,66]	<0,001

² Les comparaisons indirectes sont aussi réalisables avec un contrôle commun actif.

Les résultats de comparaisons directes figurant au Tableau 2, nous permettent de disposer de deux risques relatifs, celui de A vs PBO et celui de B vs PBO. Ces risques relatifs mesurent l'efficacité « absolue » du traitement considéré (A ou B), c'est à dire la réduction relative de fréquence de l'événement qu'apporte ce traitement par rapport à l'absence de son utilisation³.

Intuitivement, si le risque relatif de A *versus* PBO est plus petit que celui de B *versus* PBO, on a tendance à considérer que A est plus efficace que B, bien que, comme nous le verrons par la suite, cette méthode de comparaison soit inappropriée.

Tableau 2 - Résultats des essais de comparaisons directes de base nécessaires à la comparaison indirecte de A et de B par l'intermédiaire du placebo

A <i>versus</i> Placebo				
	Traitement A n=1 000	Placebo n=1 000	Risque relatif [IC 95 %]	Valeur de p
Décès	134 (13,4 %)	274 (27,4 %)	0,49 [0,41 ; 0,59]	<0,001
B <i>versus</i> Placebo				
	Traitement B n=1 000	Placebo n=1 000	Risque relatif [IC 95 %]	Valeur de p
Décès	56 (5,6 %)	103 (10,3 %)	0,50 [0,4 ; 0,74]	<0,001

Jusqu'à récemment, la comparaison directe était considérée comme l'approche de référence, tandis que les « comparaisons indirectes » étaient l'objet de nombreuses réserves, notamment en raison des méthodes simplistes utilisées. Des méthodes appropriées ont été développées. Des études théoriques et empiriques de validation réalisées ces dernières années sur le sujet laissent entrevoir une place en évaluation pour les comparaisons indirectes correctement réalisées et interprétées avec rigueur.

1.2 Objectifs

Ce travail bibliographique a plusieurs objectifs :

1. présenter le contexte dans lequel les techniques de comparaisons indirectes se sont développées ;
2. identifier les situations où les comparaisons indirectes ont un réel apport ;
3. présenter les différentes techniques fiables de comparaisons indirectes ;
4. établir les points forts et les limites de chaque méthode ;
5. rechercher quelle est la validité des comparaisons indirectes ;
6. proposer une méthode d'évaluation et de lecture critique d'un travail de comparaison indirecte.

³ Dans le cas le plus courant où le critère de jugement a valeur d'échec du traitement, un RR < 1 indique que le traitement testé est supérieur au traitement contrôle. Plus le risque relatif est proche de zéro, plus le bénéfice apporté par le traitement testé est important. Un RR > 1 témoigne d'un effet délétère du traitement testé ou d'une plus grande efficacité du traitement contrôle. L'interprétation est inversée si le critère de jugement est un événement favorable. La réduction relative de risque (RRR) est assez fréquemment utilisée à la place du risque relatif : $RRR = (1-RR) \times 100 \%$.

2. Contexte

Actuellement, il est extrêmement fréquent de disposer de plusieurs traitements concurrents pour une même situation pathologique. Il s'agit de traitements, médicamenteux ou non, qui s'adressent aux mêmes patients et qui implique un choix thérapeutique. Par exemple, à la phase aiguë de l'infarctus du myocarde, de nombreux fibrinolytiques sont disponibles pour la repermeabilisation de l'artère coronaire obstruée ; l'angioplastie primaire a le même objectif. Le traitement de la crise aiguë de migraine peut faire appel à des traitements non spécifiques (antalgiques, anti-inflammatoires non stéroïdiens) ou spécifiques tels que les triptans. Plusieurs anticholinestérasiques sont indiqués dans la maladie d'Alzheimer.

En évaluation, se pose la question de l'efficacité et de la tolérance relatives des traitements concurrents les uns par rapport aux autres. Dans le but de faire des choix ou des recommandations rationnels, il serait souhaitable de savoir si un traitement est plus efficace et/ou mieux toléré que les autres et, dans ce cas, comment se situe son efficacité et/ou sa tolérance par rapport aux autres. L'idée est de pouvoir proposer au patient le meilleur traitement en première intention et de réserver les autres pour une deuxième ou troisième intention.

Malheureusement, les comparaisons directes des traitements concurrents entre eux sont rarement disponibles, empêchant souvent d'aller bien loin dans la comparaison relative de la balance bénéfice/risque des traitements. L'absence de ces informations engendre un blocage du processus rationnel d'évaluation des traitements sur des faits prouvés. Le choix se base alors sur les propriétés pharmacologiques des traitements, sur des résultats portant sur des critères intermédiaires, sur des considérations spéculatives diverses et variées, et c'est bien souvent des facteurs externes (attrait de la nouveauté, pression marketing, etc.) qui décident du traitement préférentiel.

L'idée de faire des comparaisons indirectes apparaît dans ce contexte. Il s'agit d'exploiter au mieux l'information issue de comparaisons directes disponible (principalement issue d'essais contre placebo, éventuellement de quelques essais épars *versus* traitements actifs) pour estimer l'efficacité et la tolérance relatives des différents traitements concurrents (1-4).

2.1 Manque de comparaisons directes entre traitements actifs

Pourquoi ne dispose-t-on généralement pas de tous les essais de comparaisons directes entre traitements actifs nécessaires à l'évaluation ?

Dans un schéma idéal, la question ne devrait pas se poser. A partir du moment où un premier traitement a montré son efficacité, les nouveaux arrivants devraient se comparer à lui pour montrer leur supériorité (d'efficacité ou de tolérance). Ce schéma idéal permettrait ainsi d'établir la hiérarchie de tous les traitements.

Le premier traitement T1 montre son efficacité par rapport au placebo (sauf impossibilité pour raison éthique). Il devient le traitement standard. Le deuxième traitement T2 se compare à T1 et montre sa supériorité (d'efficacité et/ou de tolérance). Il devient le nouveau traitement standard. Un troisième traitement se compare à T2, etc.

Cependant, même le parfait respect de ce processus peut conduire à des manques de comparaisons directes pourtant utiles. Cela se produit, par exemple, lorsque le troisième traitement échoue à montrer sa supériorité sur T2. Que devient ce traitement ? Faut-il l'abandonner définitivement ? Ne pourrait-il pas représenter une alternative de seconde intention (en cas d'intolérance d'un patient à T2 par exemple). En effet, il est peut être plus efficace et/ou mieux toléré que T1 (bien que non supérieur à T2). Un essai T3 vs T1 serait nécessaire pour répondre à cette interrogation. Il impliquerait de donner aux patients un traitement, T1, connu comme non optimal au moment de la réalisation de l'essai. Dans le cas d'une pathologie engageant le pronostic vital ou fonctionnel, un tel essai est irréalisable pour des raisons éthiques.

Une autre situation conduisant à un manque de comparaison directe peut survenir dans ce schéma idéal. Dans les domaines à haut potentiel, l'activité de développement est intense et deux nouveaux traitements (ou plus) peuvent débiter leurs essais simultanément. Dans ce cas les nouveaux

traitements T3 et T4 vont se comparer à T2, traitement de référence au moment de l'initiation de leurs essais. A l'issue des essais, T3 et T4 se révèlent supérieurs à T2. Se pose alors la question de l'efficacité et de la tolérance relative de T3 et T4. Dans l'absolu, rien ne s'oppose alors à la réalisation d'un essai T3 vs T4 (car il n'entraînera pas de perte de chance connue pour les patients). En pratique, cet essai inter promoteurs est rarement réalisé.

La situation peut être encore plus complexe avec le recours aux essais de non infériorité. Les nouveaux traitements peuvent éviter de se positionner en supériorité par rapport au traitement de référence pour éviter la situation où T3 échoue à montrer sa supériorité par rapport à T2 et se retrouve ainsi dans une situation de quasi éviction. Un essai visant à montrer que T3 est non inférieur (« équivalent ») à T2 est entrepris à la place de l'essai de supériorité T2 vs T3, dans l'espoir de pouvoir conclure que T3 est aussi efficace que T2 et donc interchangeable. On sait que cette conclusion est erronée car la méthodologie de l'essai de non infériorité introduit une perte d'efficacité consentie pour permettre de conclure à la non infériorité. A l'issue d'un essai de non infériorité concluant, T3 est potentiellement inférieur à T2 et, en pratique, T3 n'obtient qu'une indication de seconde intention⁴. La question qui se pose alors est de savoir comment se positionne T3 par rapport à T1, en sachant qu'il s'agit d'envisager une comparaison directe avec un traitement d'efficacité non optimale (T1). En théorie, on devrait être assuré que T3 est au moins aussi efficace que T1 car la limite de non infériorité aurait dû être choisie afin de le garantir (cf. ICH E10), mais l'étude empirique du domaine des essais de non infériorité montre que cette occurrence est très rare (5,6).

2.2 Prédilection pour les essais contre placebo

Dans de nombreux domaines, le recours exclusif au placebo reste acceptable au cours du développement d'un nouveau traitement même après la mise à disposition d'un premier traitement efficace (7). C'est par exemple le cas des traitements symptomatiques de pathologies n'engageant pas le pronostic vital ou fonctionnel du patient. Dans ces domaines, le schéma idéal évoqué précédemment n'est pas suivi.

A l'issue de ce type de développement, l'évaluation peut se retrouver avec n comparaisons *versus* placebo sans aucune notion de l'efficacité relative des traitements entre eux (si l'on s'en tient aux essais réalisés). Il s'agit d'une situation de fait, très courante.

Ce manque chronique d'essais comparant les traitements actifs entre eux a régulièrement été dénoncé aussi bien en France qu'Outre Atlantique.

Un éditorial du *New York Times* du 16 novembre 2003 intitulé « Head-to-Head Drug Combat »⁵ relève que « ...For the most part, drugs in this country are not tested against other drugs in the same class. Instead they are tested against a placebo, and if shown to be comparatively safe and effective are approved for marketing. That leaves both patients and their doctors uncertain which approved drugs are better than their competitors and whether high-priced drugs warrant their added cost compared with lower-cost alternatives ».

En France, ce manque est aussi relevé dans un rapport de la Cour des Comptes sur l'application des lois de financement de la sécurité sociale du 12 septembre 2007, chapitre 9 (8) :

« Par ailleurs, le caractère non obligatoire et donc non systématique des essais cliniques contre comparateurs est problématique. La Commission de la transparence ne dispose pas du pourcentage exact de dossiers présentant des essais cliniques contre comparateurs, mais une estimation réalisée à partir d'un échantillon laisse à penser que moins de la moitié des dossiers dispose de ces données. »

⁴ Dans ce contexte, traitement de seconde intention désigne par exemple un traitement indiqué en cas de contre indication ou de mauvaise tolérance à un traitement de première intention.

⁵ <http://query.nytimes.com/gst/fullpage.html?res=9A0CE1DE1338F935A25752C1A9659C8B63>, consulté le 2 nov. 2007

2.3 Non transitivité des tests statistiques

Une autre difficulté que l'on est susceptible de rencontrer dans la recherche de la hiérarchie d'efficacité (ou de sécurité) de plusieurs traitements est la non transitivité de certains tests statistiques employés dans les essais cliniques (9-11).

Ce problème fait que, même si toutes les comparaisons directes nécessaires sont réalisées, les conclusions des essais s'avèrent paradoxales, empêchant le classement des traitements.

Par exemple, dans un même essai à 3 bras comparant les traitements A, B et C, il est possible de trouver A meilleur que B, B meilleur que C et C meilleur que A en raison de la non transitivité de certains tests⁶.

La non transitivité est bien connue pour le test de Wilcoxon, mais se retrouve aussi pour toute la famille des tests basés sur les rangs comme le Logrank et le Cox, très fréquemment utilisés dans les essais cliniques.

Ce phénomène représente donc aussi une limite à l'approche basée uniquement sur des essais « face à face » conduisant à des comparaisons 2 à 2 des traitements. Les paradoxes induits par la non transitivité des tests sont résolus par l'utilisation des méthodes d'estimation globale du réseau d'essais décrites dans la section 4.4.

2.4 Essais cliniques de non infériorité

Les essais de non infériorité sont de plus en plus courants à l'heure actuelle. Ils cherchent à montrer qu'un nouveau traitement n'est pas inférieur à un traitement standard. Derrière cette conclusion se cache une perte d'efficacité consentie potentielle. Ce point, constamment méconnu dans les conclusions de ces essais, fait qu'un essai de non infériorité probant ne permet pas de conclure que le nouveau traitement est aussi efficace que le traitement de référence. On peut seulement conclure, avec un risque d'erreur de 5 %⁷, qu'il n'entraîne pas une perte d'efficacité supérieure à la limite choisie. Le risque d'erreur limité à 5 % est celui de faire cette conclusion alors que le nouveau traitement est nettement inférieur au traitement standard. Etant donnée cette perte d'efficacité consentie, il faut que le nouveau traitement présente des avantages par rapport au traitement standard sur d'autres plans que l'efficacité (tolérance, modalités d'utilisation, etc.).

En pratique, les limites de non infériorité sont encore choisies de manière arbitraire et ne répondent pas aux exigences réglementaires imposées par ICH E9 (§ 3.3.2). La limite choisie doit garantir que la perte d'efficacité entraînée par le nouveau traitement ne sera pas supérieure à l'efficacité du traitement standard, sinon l'essai de non infériorité pourrait s'avérer concluant avec un nouveau traitement inférieur au traitement standard : *An equivalence margin should be specified in the protocol; this margin is the largest difference that can be judged as being clinically acceptable and should be smaller than differences observed in superiority trials of the active comparator.*

L'interprétation des résultats des essais de non infériorité n'étant pas intuitive et la vérification de l'acceptabilité de la limite choisie nécessitant de disposer des résultats des essais réalisés avec le traitement standard, il a été proposé à plusieurs reprises de transformer les résultats de l'essai de non infériorité en une comparaison par rapport à un placebo putatif à l'aide d'un processus d'extrapolation similaire à celui mis en œuvre dans les comparaisons indirectes.

Une approche basée sur le réseau d'essais permet d'intégrer les essais de non infériorité et de supériorité et ainsi d'unifier l'interprétation des résultats de ces différents types d'études, sans tenir compte des seuils de non infériorité utilisés et très souvent non justifiés.

⁶ Le lecteur intéressé par cette curiosité statistique pourra consulter sur Internet : http://en.wikipedia.org/wiki/Nontransitive_dice, <http://mathworld.wolfram.com/EfronsDice.html>.

⁷ En fait 2,5 %, mais comparable au risque d'erreur d'un essai de supériorité de conclure à la supériorité en raison de la latéralité des hypothèses conduisant à la recommandation du nouveau traitement (cf. www.spc.univ-lyon1.fr/polycoq sections : test unilatéraux et essais de non infériorité)

3. Méthodes de la recherche documentaire et critères de sélection

3.1 Sources d'informations

3.1.1 Bases de données bibliographiques interrogées

- Biosis Previews (*BioScience Information Service*, Etats Unis) ;
- Current Contents Search (*Institute for Scientific Information* - Thomson Scientific, Etats Unis) ;
- Embase (Elsevier, Pays Bas) ;
- Medline (*National Library of Medicine*, Etats Unis) ;
- Pascal (Institut national de l'information scientifique et technique, France).

3.1.2 Autres sources

- *The Cochrane Library* (Wiley Interscience, Etats Unis) ;
- *The Cochrane Methodology Register* ;
- Sites d'organismes publiant des recommandations et des rapports d'évaluation technologique (voir Annexe 3) ;
- Google Scholar ;
- Bibliographie des documents sélectionnés.

3.2 Stratégie de la recherche bibliographique

La recherche dans les bases de données bibliographiques a été réalisée sur les périodes suivantes :

- Embase, Medline et Pascal : de 01/1985 à 04/2009 ;
- Biosis Previews : de 01/1993 à 09/2007 ;
- Current Contents Search : de 01/1990 à 09/2007 ;
- The Cochrane Methodology Register : issue 2, 2007.

Aucun descripteur issu de thesaurus des bases bibliographiques interrogées n'existe à l'heure actuelle pour désigner le concept de « comparaison indirecte ». De ce fait, la recherche a été effectuée exclusivement en texte libre. Les termes en texte libre ont été combinés en autant d'étapes que nécessaire à l'aide des opérateurs « ET » « OU » « SAUF ».

La Figure 2 présente la stratégie de recherche dans les bases bibliographiques. Seules les publications en français et en anglais ont été recherchées.

Figure 2 - Stratégie de recherche dans les bases Biosis Previews, Current Content Search, Embase, Medline et Pascal

indirect comparison*/titre OU comparaison*indirecte*/titre OU indirect analysis/titre OU analyse* indirecte*/titre OU
[(indirect evaluation*/titre ET direct/titre) OU evaluation* indirecte*/titre ET direct*/titre]

OU

[indirect comparison*/titre, résumé OU comparaison indirecte/titre, résumé OU indirect analysis/titre, résumé OU analyse indirecte/titre, résumé OU indirect evaluation*/titre OU evaluation indirecte/titre, résumé OU (indirect evidence*/titre ET direct/titre) OU indirect treatment* comparison*/titre, résumé OU mixed treatment*/titre OU multiple treatment*/titre OU several treatment*/titre OU mixed treatment* comparison*/titre, résumé OU multiple treatment* comparison*/titre, résumé OU several treatment* comparison*/titre, résumé] ET [randomized controlled trial*/titre, résumé OU essai randomisé/titre, résumé OU randomized controlled trial*/descripteur OU essai randomisé/descripteur OU meta-analys*/titre, résumé OU metaanalys*/titre, résumé OU meta-analysis/descripteur OU metaanalyse/descripteur]

* troncature

3.3 Critères de sélection des publications

Ont été retenus les articles proposant des méthodes de comparaisons indirectes, effectuant des études de validité des comparaisons indirectes ou discutant l'intérêt et les limites de ces approches ainsi que les éditoriaux et les textes de procédure ou de position émanant des agences sanitaires nationales et internationales. Ont également été sélectionnés quelques exemples d'utilisation de comparaisons indirectes.

4. Les différentes techniques permettant de réaliser des comparaisons indirectes

De manière générale, toutes techniques produisant une estimation de l'effet (risque relatif, odds ratio, différence de risque, hazard ratio, différence de moyenne) de A par rapport à B à partir des estimations de l'effet de A *versus* contrôle et de B *versus* contrôle permet de réaliser des comparaisons indirectes.

Trois grands types de méthodes sont décrits dans ce chapitre :

- les méthodes non valides, inappropriées ;
- la méthode des comparaisons indirectes ajustées, comparant les traitements 2 à 2 ;
- les méthodes modélisant simultanément tous les traitements concurrents dans la situation pathologique envisagée.

4.1 Les techniques inappropriées

Deux méthodes régulièrement utilisées dans la littérature sont impropres à réaliser des comparaisons indirectes valides. Aucune terminologie consacrée n'est disponible pour désigner ces méthodes. Nous proposons de les identifier par la description de leur principe : comparaisons naïves des estimations ponctuelles ou des bras actifs issus d'essais contrôlés distincts.

4.1.1 Comparaison naïve des estimations ponctuelles

L'approche consistant à comparer l'estimation ponctuelle de l'efficacité de A vs placebo⁸ à celle de B *versus* placebo est inappropriée. Elle vient spontanément à l'esprit mais malheureusement elle présente de très nombreuses limites.

Il s'agit, par exemple, de comparer le risque relatif obtenu sur la mortalité lors de l'évaluation de A *versus* placebo à celui obtenu lors de l'évaluation de B *versus* placebo. Tout autre indice d'efficacité (réduction relative de risque, odds ratio, réduction relative des odds, différence de risque, nombre de patients à traiter (NNT, *number needed to treat*), différence absolue ou relative de moyennes,...) peut être théoriquement employé à la place du risque relatif. Le traitement entraînant la plus grande réduction relative de risque sera qualifié plus efficace que l'autre.

Par exemple, les deux traitements A et B ont donné lieu à un essai de mortalité chacun dont les résultats sont représentés dans le Tableau 3.

Tableau 3 - Résultats des estimations ponctuelles de l'essai A vs placebo et de l'essai B vs placebo

	Risque relatif Décès de toute cause
Traitement A (<i>versus</i> placebo)	0,76
Traitement B (<i>versus</i> placebo)	0,87

Le traitement A entraîne une réduction relative de mortalité de 24 %, plus importante que le traitement B (13 %). A sera donc qualifié de « plus efficace ».

En dehors de toutes considérations sur la méthodologie des essais, sur les populations incluses et les traitements délivrés, la principale limite de cette approche est de ne pas tenir compte de la précision statistique des estimations. En fait, il se peut très bien que l'estimation la plus favorable soit aussi celle qui a l'intervalle de confiance le plus large, et donc une borne péjorative de l'intervalle de confiance compatible avec une faible efficacité. Dans ce cas, la conclusion effectuée à partir des estimations ponctuelles, privilégiera le traitement dont le bénéfice minimal garanti⁹ est le plus faible,

⁸ Pour alléger l'exposé nous nous référons à des comparaisons *versus* placebo. Néanmoins, tout ce qui est écrit est valable quelque soit le comparateur utilisé : placebo, soins standards, traitement actif.

⁹ Le bénéfice minimal garanti (à 95 %) est l'interprétation de la borne péjorative de l'intervalle de confiance.

au détriment de celui qui nous assure qu'il apporte un bénéfice supérieur, même dans le pire des cas d'erreur statistique sur l'estimation. Cette situation est illustrée dans le Tableau 4.

Tableau 4 - Résultats des estimations ponctuelles et intervalles de confiance de l'essai A vs placebo et de l'essai B vs placebo

	Risque relatif Décès de toute cause [IC 95 %]
Traitement A (<i>versus</i> placebo)	0,76 [0,53 ; 0,99]
Traitement B (<i>versus</i> placebo)	0,86 [0,84 ; 0,90]

Avec le traitement B, une réduction de mortalité d'au moins 10 % est garantie avec un degré de certitude de 95 % tandis qu'avec A nous n'avons l'assurance que d'une réduction de 1 % (avec un degré de certitude de 95 %). Un exemple réel de l'utilisation de cette approche est donné dans l'étude de cas de la section 9.3.1.

La prise en considération de la borne péjorative de l'intervalle de confiance montre que la comparaison des estimations ponctuelles est inappropriée. Elle ne permet pas d'obtenir une méthode satisfaisante de comparaison indirecte. Le raisonnement se focalise sur une seule possibilité bien particulière (la borne péjorative de l'intervalle de confiance) et néglige l'ensemble des autres possibilités offertes par l'intervalle de confiance. Les méthodes développées contournent ce problème.

De plus, la réalisation de ce type de comparaisons avec des différences de risques ou des NNT est faussée par les éventuelles différences de risques de base entre les essais.

4.1.2 Comparaison naïve des bras actifs

Une autre approche inappropriée consiste à considérer les bras traitements actifs des essais sans tenir compte de la randomisation, c'est à dire de leur appariement au sein de l'essai.

A partir des deux essais A *versus* placebo et B *versus* placebo, seules sont utilisées les données du bras A du premier essai et celles du bras B du second essai. Les deux bras placebo sont en quelque sorte mis de côté. L'effet de A *versus* B est alors estimé en comparant les résultats du bras A à celui du bras B. De ce fait, plus aucun contrôle des facteurs de confusion propres à chaque étude n'est effectué. Il y a perte totale du caractère contrôlé des deux essais. Le résultat obtenu n'est plus à l'abri d'un biais de confusion. De même, il y a perte totale du bénéfice de la randomisation. La comparaison effectuée n'est pas à l'abri d'un biais de sélection. En conclusion, cette approche, bien qu'effectuée à partir de deux essais contrôlés randomisés, n'a plus aucune fiabilité méthodologique et devient purement observationnelle.

A partir des résultats présentés dans les Tableau 5 et Tableau 6, cette approche comparerait 45/1 567 à 36/1 251 soit 2,87 % à 2,88 %. Cette comparaison inciterait à conclure qu'il n'est pas possible de considérer A comme plus efficace que B. Cette conclusion est paradoxale puisque l'examen des deux essais montre que A est efficace par rapport au placebo tandis que B n'a pas montré son efficacité. La comparaison A *versus* B proposée est confondue par le risque de base.

Tableau 5 - Résultats de l'essai comparant A au placebo

	Traitement A	Placebo	Risque relatif [IC 95 %] Valeur de p
Décès	45/1 567 (2,87 %)	103/1 546 (6,66 %)	0,43 [0,31 ; 0,61] p<0,001

Tableau 6 - Résultats de l'essai comparant B au placebo

	Traitement B	Placebo	Risque relatif [IC 95 %] Valeur de p
Décès	36/1 251 (2,88 %)	42/1 298 (3,24 %)	0,89 [0,57 ; 1,38] p=0,6

En plus des arguments théoriques contre la validité de cette approche (cf. supra), le caractère biaisé de cette approche a été montré de manière empirique (12). L'étude de cas de la section 9.3.2 rapporte un exemple réel de l'utilisation de cette approche.

Cette approche, complètement invalide sur le plan méthodologique, est à l'origine des réticences actuelles contre les comparaisons indirectes. Ce fut la première méthode utilisée, mais le problème qu'elle pose n'est pas consubstantiel aux comparaisons indirectes. Les méthodes spécialement développées solutionnent cette difficulté comme il sera démontré dans la section 4.2.1.

4.2 Les comparaisons indirectes ajustées

L'estimation indirecte ajustée (« adjusted indirect comparison ») est la première méthode valide disponible permettant de réaliser des comparaisons indirectes.

4.2.1 Principe

Cette méthode de comparaison indirecte deux à deux compose avec l'estimation de l'effet de A *versus* placebo et celle de B *versus* placebo. De ce fait, elle ne détruit pas la randomisation et conserve les propriétés méthodologiques de l'essai contrôlé randomisé.

Le calcul nécessaire est très simple. Considérons par exemple le cas où le critère de jugement d'intérêt est binaire et où l'effet est mesuré par un risque relatif.

Soit RR_a le risque relatif issu de la comparaison du traitement A au placebo et RR_b celui trouvé lors de la comparaison du traitement B au placebo. Le résultat de la comparaison indirecte A *versus* B est le risque relatif $RR_{a/b}$, obtenu par

$$RR_{a/b} = \frac{RR_a}{RR_b} \quad (0.1)$$

Le calcul de l'intervalle de confiance de ce risque relatif passe par le calcul de l'intervalle de confiance de son logarithme, en utilisant comme variance la somme des variances des logarithmes des deux risques relatifs.

$$\text{var}(\log RR_{a/b}) = \text{var}(\log RR_a) + \text{var}(\log RR_b)$$

À partir de cette variance, l'intervalle de confiance du risque relatif $RR_{a/b}$ s'obtient de façon standard.

Un calcul similaire autorise l'extrapolation du bénéfice par rapport au placebo d'un nouveau traitement à partir du RR de ce traitement par rapport au traitement de référence et du RR du traitement de référence par rapport au placebo. Ce calcul est utile avec l'essai de non infériorité.

La démonstration de l'équation (0.1) est très simple. Dans un essai de comparaison directe, les patients ont un risque d'événement sans traitement de r_0 . Dans le bras recevant A, le risque sous traitement r_a devient

$$r_a = r_0 \times RR_a \quad (0.2)$$

où RR_a désigne comme précédemment l'effet de A par rapport à « pas de traitement » (mesuré dans une comparaison A *versus* placebo).

De manière identique, le risque sous traitement r_b dans le bras du traitement B est

$$r_b = r_0 \times RR_b \quad (0.3)$$

L'effet de A par rapport à B dans cet essai est mesuré par le risque relatif $RR_{a/b}$ qui est par définition

$$RR_{a/b} = \frac{r_a}{r_b}$$

qui devient en utilisant les équations (0.2) et (0.3)

$$\begin{aligned} RR_{a/b} &= \frac{r_a}{r_b} = \frac{r_0 RR_a}{r_0 RR_b} \\ &= \frac{RR_a}{RR_b} \end{aligned}$$

Ce qui justifie le calcul de la comparaison indirecte ajustée.

Dans ces calculs, le risque de base r_0 disparaît complètement. Cette extrapolation, basée sur les mesures des effets est donc totalement insensible au risque de base des essais de A et de B. La comparaison indirecte est donc valide même si les patients étudiés dans les essais de A et de B ne sont pas les mêmes. La condition de validité se trouve ailleurs, comme cela est détaillé dans la section qui suit.

4.2.2 Condition de validité : la stabilité des effets

Pour que cette approche donne un résultat qui a un sens médical, il faut qu'il y ait une certaine cohérence entre les deux risques relatifs RR_a et RR_b . C'est à dire que, dans les conditions (critères de gravité de la pathologie, caractéristiques des patients, traitements concomitants, etc.) où le traitement B a été évalué contre placebo, l'évaluation du traitement A aurait donné le même risque relatif que celui qui a été observé dans l'essai de A *versus* placebo effectivement réalisé (la réciproque devant aussi être vérifiée).

En d'autres termes, cette condition implique que les essais *versus* placebo de A et B soient similaires en termes de covariables d'interaction¹⁰ avec les traitements, c'est à dire qu'il y ait stabilité des effets pour les différents contextes dans lesquels se sont déroulés les essais pris en considération par les calculs (13).

Une covariable d'interaction est une variable rattachée aux patients, au contexte de l'essai ou aux modalités d'administration qui module l'efficacité du traitement, ce qui se traduit par une variation du risque relatif en fonction de la valeur de cette covariable. Ces interactions sont suspectées à l'aide des analyses en sous groupes et plus particulièrement à l'aide du test d'interaction.

Le Tableau 7 donne un exemple d'interaction. L'effet du traitement A n'est vraisemblablement pas le même chez les diabétiques et chez les non diabétiques. Le risque relatif global de l'essai dépend alors de la répartition des patients entre diabétiques et non diabétiques.

Si, dans l'essai où est évalué le traitement B, la proportion de diabétiques est plus importante que celle existant dans l'essai du traitement A, la comparaison de A *versus* placebo dans les conditions de l'essai de B ne donnera pas le même risque relatif que celui dont on dispose à partir de l'essai de A réalisé.

Tableau 7 - Exemple d'une interaction entre l'effet du traitement A (mesuré ici par le risque relatif) et une caractéristique patient (diabétique ou non diabétique)

	Effectif	Risque relatif Traitement A vs placebo	Interaction
Essai global	1 230	0,90	
Diabétiques	430	0,76	p=0,001
Non diabétiques	800	0,95	

Faire le rapport des deux risques relatifs pose alors le problème de savoir de quels types de patients le $RR_{a/b}$ obtenu est alors représentatif : certainement pas représentatif des patients inclus dans l'essai B, vu que A n'aurait pas donné le risque relatif pris en compte pour le calcul chez ces patients (il y a moins de diabétiques dans l'essai à l'origine de l'estimation de RR_a).

Si B n'est sujet à aucune interaction (il n'existe pas de variable modifiant son effet), sa comparaison contre placebo dans les conditions de l'essai de A aurait donné le même risque que celui dont on dispose. Le rapport RR_a/RR_b est alors valide pour la population étudiée dans l'essai de A et non pas pour la population des patients inclus dans l'essai de B.

Si les deux traitements sont sujets aux interactions, le rapport n'a plus de sens médical.

En pratique, écarter un problème de ce type nécessite de vérifier pour les deux traitements qu'aucune covariable ne modifie substantiellement l'effet des traitements. Cela s'effectue en regardant les tests d'interaction des analyses en sous groupes. Mais le diagnostic peut être rendu difficile par le manque de puissance des tests d'interaction et par la possibilité que les variables pertinentes n'aient pas été analysées.

¹⁰ Ici le terme interaction a son sens des essais thérapeutiques et non pas celui des interactions pharmacologiques dans le cadre des associations de traitement. Il désigne les variables qui modifient la valeur de l'effet traitement, du risque relatif par exemple. Cette interaction est explorée dans les essais à l'aide des analyses en sous groupes par exemple.

Par contre, les patients des essais *versus* placebo de A et de B ne sont pas nécessairement identiques en dehors de covariables d'interaction. En effet, s'ils diffèrent par leurs risques d'événements sous placebo, cette différence n'entraînera pas de modification de RR_a ou RR_b , puisque le risque relatif fait abstraction du niveau de risque de base en calculant un rapport.

La situation problématique est bien représentée par les variables modulant le risque relatif. Il faut donc que les essais de A et de B soient similaires au niveau de ces covariables d'interaction. La difficulté vient du fait que les covariables d'interaction sont mal connues. Il faudra compter sur les analyses en sous groupes réalisées dans les essais pour identifier les covariables d'interaction si elles existent et, le cas échéant, vérifier si la répartition de ces covariables est identique dans tous les essais à l'aide des tableaux des caractéristiques de base.

Cependant, il s'avère que l'existence de covariables d'interaction forte est assez peu fréquente. *A priori*, il est relativement vraisemblable que les conditions de validité sont satisfaites à partir du moment où les essais de A et de B ont été effectués dans des contextes proches (essais contemporains les uns des autres, disponibilité de traitements concomitants similaires) et chez des patients dont les caractéristiques sont proches.

4.2.3 Utilisation de la méta-analyse

En pratique, on se retrouve confronté à l'existence de plusieurs essais de A *versus* placebo et de B *versus* placebo. Les calculs de comparaisons indirectes ajustées nécessitent un seul risque relatif pour A et un seul pour B. Sélectionner un seul essai de A et un seul de B introduirait un arbitraire inacceptable dans la démarche. La solution consiste donc à effectuer en premier la méta-analyse de tous les essais de A *versus* placebo et celle des essais de B *versus* placebo. La méta-analyse synthétisant toute l'information apportée par plusieurs essais en un seul résultat, on se retrouvera à l'issue de cette étape préliminaire avec deux seuls risques relatifs (Figure 3).

De plus, le caractère généralisable des résultats issus d'une méta-analyse est *a priori* supérieur à celui des résultats issus d'une seule étude, puisque cette technique regroupe des études réalisées sur des populations et avec des protocoles quelque peu différents. Cette synthèse des données contribue à augmenter la stabilité de l'effet des traitements et surtout permet de tester cette stabilité.

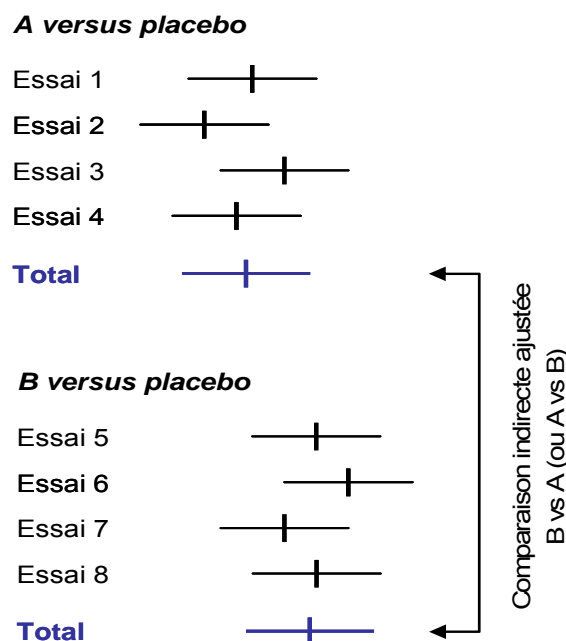


Figure 3 - Utilisation de la méta-analyse pour le calcul des comparaisons indirectes ajustées.

4.2.4 Existence de plusieurs chemins et évaluation de l'incohérence

Dans certains cas, il s'avère que des comparaisons indirectes ajustées pourraient être réalisées à l'aide de plusieurs chemins et non pas seulement en passant par le placebo.

Le terme chemin désigne en fait le comparateur commun des comparaisons de base sur lesquelles s'appuie l'extrapolation effectuée par la comparaison indirecte ajustée.

Dans le cas simple que nous avons évoqué jusqu'à présent, la comparaison indirecte de A *versus* B s'effectue à l'aide du chemin A → placebo → B puisque les deux comparaisons de bases utilisées le sont *versus* placebo.

La comparaison indirecte de A et B peut aussi passer par un autre chemin si des essais de A et de B *versus* un autre comparateur commun sont disponibles. Par exemple, un ancien traitement actif C. La disponibilité d'essais A *versus* C et B *versus* C va permettre d'estimer A *versus* B par le chemin A → C → B.

Les calculs vont donc donner deux estimations indépendantes de la comparaison indirecte de A *versus* B (celle passant par le placebo et celle passant par C). L'existence de ces deux résultats va permettre de tester s'il y a une incohérence entre ces deux chemins, c'est à dire si le résultat de la comparaison indirecte varie en fonction du chemin (Figure 4). Ce test se base sur l'hétérogénéité entre les deux estimations produites du risque relatif A vs B : $RR_{A/B}$ et $RR'_{A/B}$.

En cas d'hétérogénéité (c'est à dire s'il existe une différence statistiquement significative entre $RR_{A/B}$ et $RR'_{A/B}$), il y a incohérence : le résultat varie en fonction du chemin. Cette situation est gênante car, quelque soit le chemin, on devrait obtenir le même résultat. Il conviendra d'émettre des réserves sur les estimations, sauf en cas d'explication claire et non arbitraire de l'incohérence.

En l'absence d'hétérogénéité, le résultat final sera la méta-analyse des deux estimations $RR_{A/B}$ et $RR'_{A/B}$ donnant la meilleure précision possible. L'absence d'incohérence détectée renforce la confiance que l'on peut avoir dans le résultat.

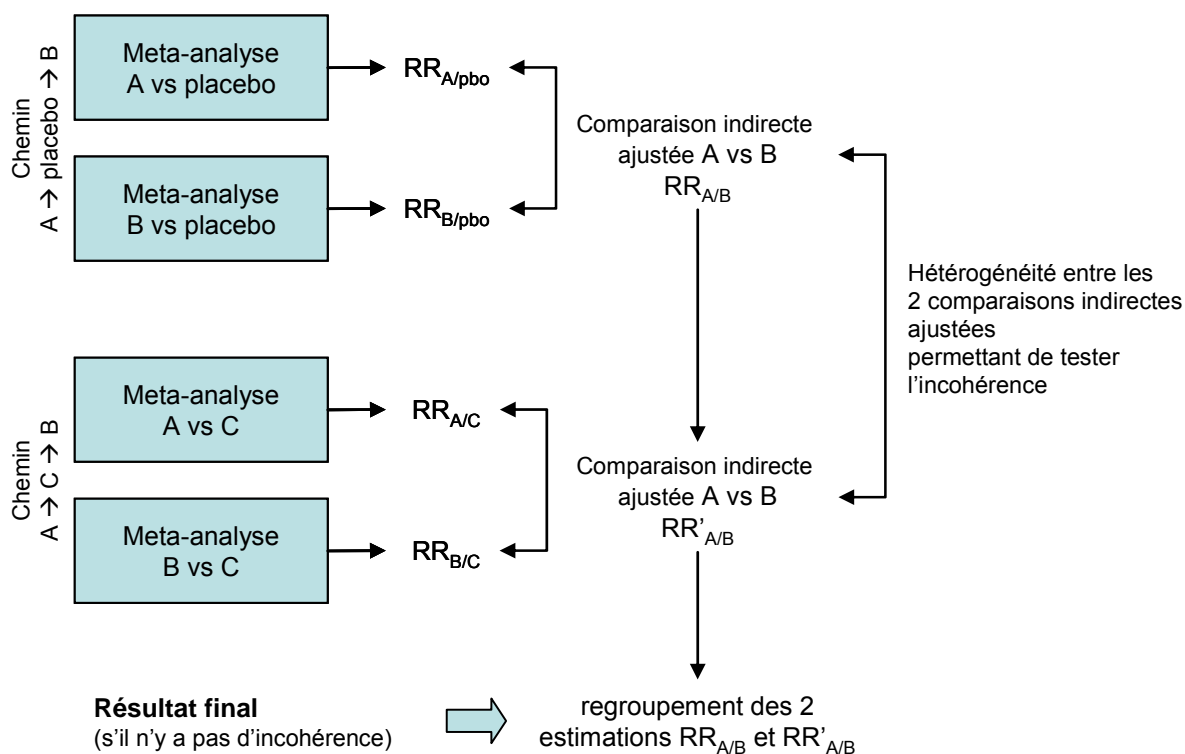


Figure 4 - Illustration du processus de comparaisons indirectes ajustées en cas d'existence de plusieurs chemins de comparaisons de base

4.2.5 Comparaisons de plus de deux traitements

En évaluation, il arrive fréquemment que la question posée soit de comparer simultanément plus de deux traitements concurrents.

A l'aide de la méthode des comparaisons indirectes ajustées, la comparaison de ces n traitements va nécessiter de procéder 2 à 2, étant donné que les comparaisons indirectes ajustées n'envisagent simultanément que deux traitements.

Pour 4 traitements T1, T2, T3 et T4, cela revient à faire 6 comparaisons :

- T1 vs T2 ;
- T1 vs T3 ;
- T1 vs T4 ;
- T2 vs T3 ;
- T2 vs T4 ;
- T3 vs T4.

Un problème de comparaisons multiples survient alors, entraînant une inflation du risque alpha. En effet, sous l'hypothèse nulle que tous les traitements ont la même efficacité (c'est à dire qu'il n'existe aucune différence entre les traitements pris 2 à 2), il y a pourtant 6 façons d'en trouver une du fait du seul hasard (les 6 comparaisons 2 à 2). Si chaque comparaison unitaire est réalisée avec un risque alpha de 5 %, le risque global de trouver à tort au moins une différence entre les 4 traitements est fortement majoré et s'élève à 26 %. Nous avons une chance sur quatre de conclure à l'existence d'un traitement plus efficace que les autres alors qu'en réalité ils sont tous aussi efficaces.

Ainsi, les comparaisons indirectes ajustées ont l'avantage d'un principe simple, facile à mettre en œuvre et facilement compréhensible. Par contre, cette méthode présente assez rapidement des limites lorsque la question posée englobe plus de deux traitements. D'autres méthodes n'ayant pas de limitations dans ce cas ont été développées. Il s'agit des méthodes basées sur une modélisation comme les méta-analyses en réseaux ou la méta-régression.

4.3 L'approche mixte des comparaisons indirectes et directes

4.3.1 Principes

Les comparaisons indirectes peuvent conserver de l'intérêt même en cas de disponibilité d'essais de comparaisons directes.

En effet, l'essai de comparaison directe est souvent unique. Assez fréquemment, il a été conçu avec un manque manifeste de puissance. Dans d'autres cas, les modalités d'utilisation du comparateur sont discutables. Dans ces situations, l'approche mixte, appelée « mixed treatment comparison »¹¹, où l'on confronte les résultats des comparaisons directes avec ceux des comparaisons indirectes présente un réel intérêt, celui de lever ou de confirmer les réserves que l'on s'apprête à faire sur l'essai de comparaison directe (Figure 5).

Cette approche mixte permet aussi d'apprécier la validité des résultats de la comparaison indirecte ajustée.

En cas de concordance des estimations issues des comparaisons directes et indirectes, les résultats sont combinés pour donner une estimation globale de la différence d'effet entre A et B tenant compte de la totalité de l'information disponible.

¹¹ Aucune terminologie consacrée n'existe en français pour l'instant.

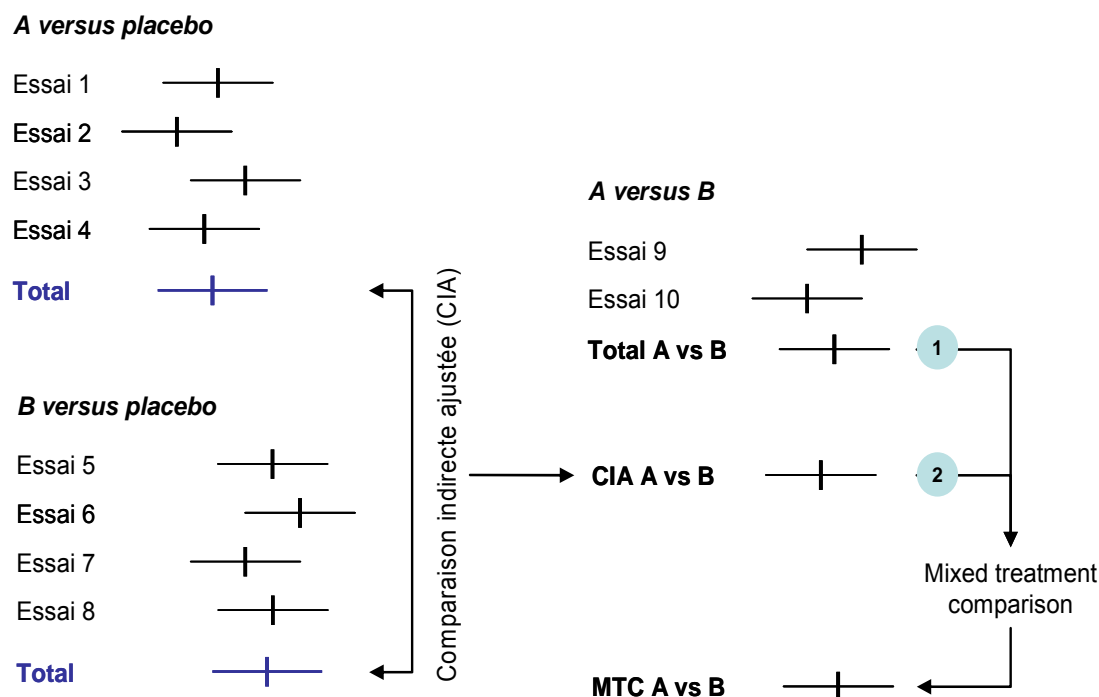
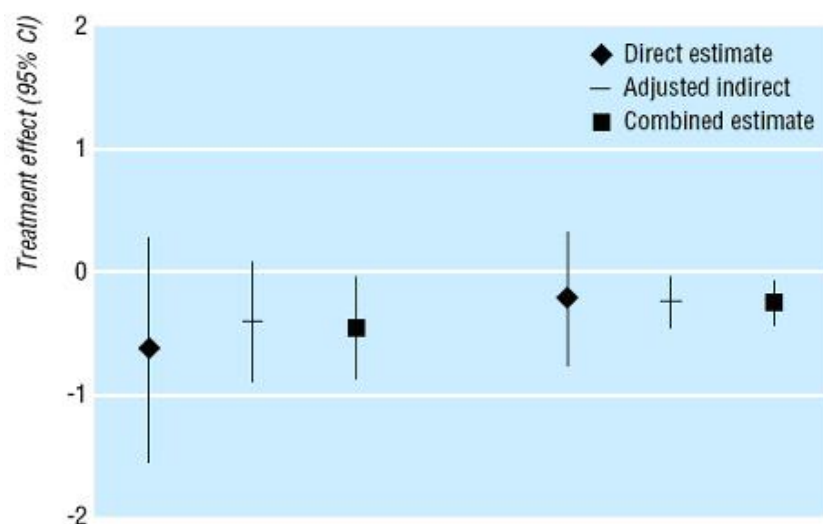


Figure 5 - Illustration du processus de « Mixed Treatment Comparison » (MTC). Le résultat de la méta-analyse des essais de comparaison directe : 1. est combiné avec celui de la comparaison indirecte : 2. la concordance de ces 2 estimations est testée par un test d'hétérogénéité.

4.3.2 Exemple

Le regroupement des estimations issues des comparaisons indirectes ajustées avec celles des comparaisons directes s'effectue simplement par méta-analyse. Song *et al.* (14) donnent deux exemples de l'application de cette technique issus des publications faites par Soo *et al.* (15) et Trindade et Menon (16) où, l'estimation des comparaisons directes étant en parfaite phase avec celle issue des comparaisons indirectes, la meilleure estimation possible est celle qui combine les résultats des deux approches (Figure 6).



Soo *et al.* (15) Trindade et Menon (16)

Figure 6 - Deux exemples d'intégration par méta-analyse des résultats des comparaisons directes avec ceux issus des comparaisons indirectes ajustées

Reproduced from Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326(7387):472. Copyright 2003, with permission from BMJ Publishing Group Ltd.

4.4 Les méthodes basées sur une estimation globale d'un réseau d'essais

4.4.1 Principe du réseau d'essais

L'idée générale de cette approche est de représenter l'ensemble de l'information apportée par les divers essais d'un domaine thérapeutique sous la forme d'un réseau (« Network of evidence ») (17). Dans ce réseau, les sommets (les points) représentent les traitements et les flèches reliant les sommets représentent les comparaisons disponibles (effectuées par des essais comparant les traitements situés de part et d'autre de la flèche, le sens de la flèche donnant le sens de la comparaison et des risques relatifs). La Figure 7 donne un exemple d'un tel réseau issu de la méta-analyse des traitements antihypertenseurs publiée en 2007 par Elliott et Meyer (18). Cette méta-analyse s'intéressait à l'apparition du diabète avec les traitements antihypertenseurs.

Les méthodes d'estimation globale proposées (cf. section 4.4.4) permettent alors d'évaluer l'efficacité de tous les traitements par rapport à un traitement de référence choisi arbitrairement ou non. Les résultats sont représentés sous forme de graphiques classiques de méta-analyse, directement compréhensibles.

L'estimation est globale, utilisant les comparaisons directes existantes et les comparaisons indirectes réalisables. La hiérarchie est établie de manière quantitative à partir de l'estimation des tailles d'effet, ce qui la met à l'abri des problèmes liés à la non transitivité des tests statistiques (cf. § 2.3).

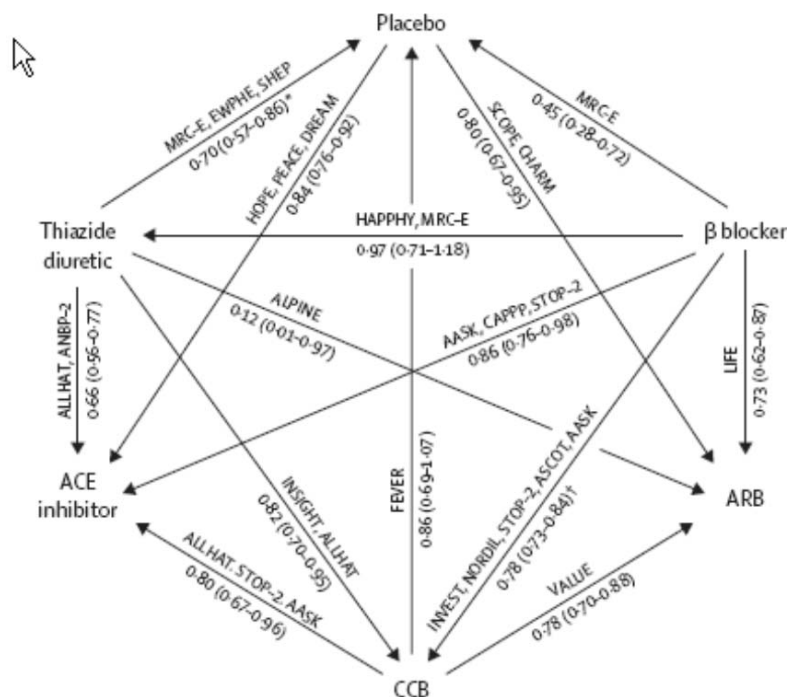


Figure 7 - Réseaux des essais d'antihypertenseurs dans lesquels l'incidence du diabète est rapportée. Les comparaisons directes disponibles sont représentées par les flèches avec indication du nom de l'essai, de l'odds ratio et de son intervalle de confiance d'après Elliott et Meyer, 2007 (18)

Reprinted from Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet* 2007;369(9557):201-7. Copyright 2007, with permission from Elsevier

Attention, l'interprétation des résultats à la recherche du traitement associé à l'incidence de diabète la plus faible revient à faire une comparaison indirecte naïve des odds ratio. A l'issue de cette analyse, les sartans (ARB) seraient considérés comme étant supérieurs aux IEC (*ACE inhibitor*) parce que l'odds ratio sartan *versus* diurétique est plus intéressant que l'odds ratio IEC *versus* diurétique. Or l'interprétation correct de ces résultats doit se limiter à : « Cette méta-analyse en réseau montre que le placebo, les antagonistes calciques (CCB), les IEC et les sartans (ARB) entraînent moins de cas incidents de diabète que les diurétiques ». Aucune conclusion n'est possible pour les bêtabloquants.

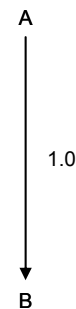
4.4.2 La notion d'incohérence

Le réseau le plus simple est celui qui concerne seulement deux traitements A et B (Figure 8). La valeur figurant à côté de la flèche est l'estimation de l'effet de B par rapport à A. Dans la suite de la présentation, nous considérerons que l'estimation de l'effet est effectuée par la différence (de risque par exemple) désignée par δ_{AB} :

$$\delta_{AB} = x_B - x_A$$

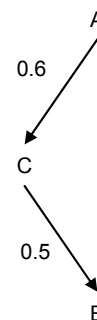
où x_A et x_B désignent les valeurs obtenues dans les bras A et B de la comparaison A *versus* B. x peut être un risque (dans ce cas δ_{AB} est la différence des risques) ou une moyenne (et δ_{AB} est une différence de moyenne).

Figure 8 - Réseau impliquant 2 traitements A et B



Dans une situation où deux comparaisons sont disponibles, celle de C par rapport à A et celle de B par rapport à C, on obtient le réseau représenté sur la Figure 9.

Figure 9 - Réseau comprenant 2 types de comparaisons C *versus* A et B *versus* C



Un moyen simple d'obtenir l'effet de B par rapport à A est de faire la somme $\delta_{AB} = \delta_{AC} + \delta_{CB}$. Soit, avec les valeurs numériques de l'exemple de la Figure 10, $\delta_{AB} = 0,6+0,5 = 1,1$.

Si 4 types de comparaisons sont disponibles, le graphe du réseau de ces résultats prend la forme représentée Figure 10.

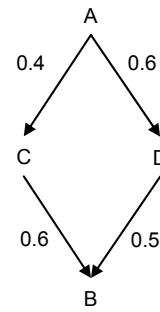


Figure 10 - Réseau impliquant 4 traitements A, B, C, D

La même différence δ_{AB} peut maintenant être obtenue par deux chemins différents : $A \rightarrow C \rightarrow B$ et $A \rightarrow D \rightarrow B$. δ_{AB} est obtenu par $\delta_{AB} = \delta_{AC} + \delta_{CB}$ pour le premier chemin et par $\delta_{AB} = \delta_{AD} + \delta_{DB}$ pour le second, soit avec les valeurs numériques : $0,4 + 0,6 = 1,0$ et $0,6 + 0,5 = 1,1$. Ainsi, on obtient deux estimations de δ_{AB} très proches l'une de l'autre quel que soit le chemin de comparaison suivi. Une certaine cohérence existe dans le réseau de comparaisons indirectes possibles.

Dans la situation illustrée par la Figure 11, le premier chemin donne comme estimation de δ_{AB} $-0,6 - 0,4 = -1,0$ et le second $0,5 + 0,6 = 1,1$. Ici, il y a une complète discordance entre les deux estimations traduisant une incohérence dans le réseau de comparaisons. Aucune estimation indirecte de A *versus* B ne sera possible. Cette notion d'incohérence est donc un élément important de la validité des comparaisons indirectes (réalisées à travers un réseau de comparaisons). Si elle est testable, c'est à dire si plusieurs chemins existent pour effectuer la même extrapolation, son absence est un élément confortant la validité des extrapolations obtenues : quelque soit la façon de procéder, la même estimation est obtenue.

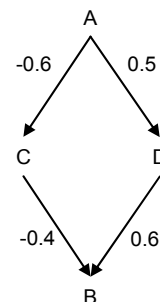


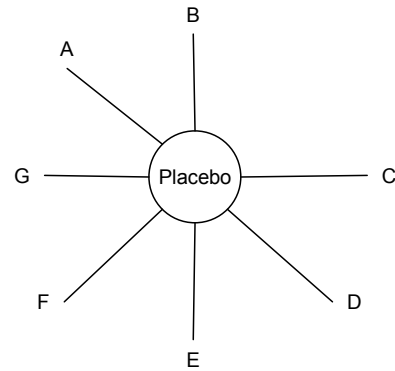
Figure 11 - Exemple d'incohérence dans un réseau de comparaisons

4.4.3 Les réseaux de comparaisons correspondant aux différents schémas de développement

Les différents schémas de développement évoqués dans les sections 2.1 et 2.2 peuvent être illustrés à l'aide de réseaux de comparaisons.

Le réseau présenté Figure 12 correspond à une situation où tous les traitements concurrents ont été développés à l'aide d'essais *versus* placebo.

Figure 12 - Réseau d'évaluation de tous les traitements concurrents *versus* placebo



Le réseau de la Figure 13 correspond à une situation où le recours au placebo est impossible à partir du moment où un traitement efficace est disponible (évaluation de traitements visant à réduire la mortalité par exemple). Les essais utilisent comme contrôle le dernier traitement ayant montré sa supériorité.

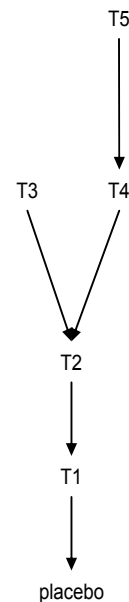


Figure 13 - Réseau où le recours au placebo est impossible à partir du moment où un traitement efficace est disponible

4.4.4 Les méthodes d'estimation globale proposées

Plusieurs modèles ont été proposés dans la littérature pour représenter des réseaux. Plusieurs méthodes statistiques pour l'estimation des paramètres de ces modèles (en particulier ceux caractérisant les comparaisons de traitements) ont aussi été proposées. Cette diversité débouche sur la classification suivante :

- Estimation à l'aide de méthodes Bayésiennes : *Bayesian network meta-analysis*
 - Modèle de Lu et Ades (19)
 - Modèle de Caldwell *et al.* (20)
- Estimation à l'aide d'un modèle linéaire mixte
 - *Network meta-analysis* de Lumley (21)

4.5 « Bayesian network meta-analysis »

Parmi les différents modèles proposés pour réaliser des méta-analyses bayésiennes en réseau (*Bayesian network meta-analysis*), le plus abouti semble être celui proposé par Lu et Ades (19). Il découle des précédents modèles proposés par Caldwell *et al.* (20) (cf. infra), et Smith *et al.* (22). Tous

ces modèles reposent sur la proposition de méta-analyses Bayésiennes faite par Higgins et Whitehead (23).

La mise en œuvre de ces méthodes fait appel à des techniques d'intégration par Monte Carlo des chaînes de Markov à l'aide du logiciel Winbugs.

4.5.1 Modèle de Lu et Ades

■ Méthode statistique

Lu et Ades (19) proposent une extension à I traitements du modèle de Smith *et al.* (22) pour combiner comparaisons directes et indirectes afin de réaliser des « mixed treatment comparisons ». Leur modèle prend aussi en compte naturellement les essais multibras qui posent le problème de la non indépendance (corrélacion) des comparaisons réalisées avec un contrôle commun. L'estimation est réalisée par une approche Bayésienne implémentée à l'aide des techniques de Monte Carlo sur Chaîne de Markov.

Le modèle de base à deux traitements ne s'applique qu'aux critères de jugement binaires et se construit de la façon suivante.

Les exposants T et C désignent respectivement le traitement étudié T et le traitement contrôle C. L'indice k désigne le $k^{\text{ème}}$ essais. r est le nombre d'événements observés dans un groupe de taille n , p

est la probabilité de survenue de l'événement. L'effet du traitement δ_k est autorisé à varier d'un essai à l'autre. Il s'agit donc d'un modèle de méta-analyse aléatoire. L'effet global du traitement (en temps

que méta-paramètre) est estimé par d . Au niveau de chaque essai, μ_k désigne la fréquence moyenne d'événements dans les groupes T et C qui, n'étant pas un paramètre d'intérêt, est traitée comme un paramètre de nuisance.

Avec ces conventions, le modèle est

$$r_k^T \approx \text{bin}(p_k^T, n_k^T)$$

$$r_k^C \approx \text{bin}(p_k^C, n_k^C)$$

$$\text{logit}(p_k^T) = \mu_k + \delta_k / 2$$

$$\text{logit}(p_k^C) = \mu_k - \delta_k / 2$$

$$\delta_k \approx \text{normal}(d, \sigma^2)$$

L'intérêt de ce modèle est qu'il peut être facilement généralisé à des essais multibras. L'extension à I traitements i , s'effectue par :

$$r_{ki} \approx \text{bin}(p_{ki}, n_{ki})$$

$$\text{logit}(p_{k1}) = \mu_i - \delta_{k2} / I - \delta_{k3} / I - \dots - \delta_{ki} / I$$

$$\text{logit}(p_{k2}) = \mu_i + (I-1)\delta_{k2} / I - \delta_{k3} / I - \dots - \delta_{ki} / I$$

$$\text{logit}(p_{ki}) = \mu_i - \delta_{k2} / I - \delta_{k3} / I - \dots + (I-1)\delta_{ki} / I$$

Encadré - Méthodes de Monte Carlo par chaînes de Markov (MCMC)

Les méthodes de Monte Carlo par chaînes de Markov (MCMC) permettent, pour un paramètre d'intérêt, d'obtenir sa distribution à partir des données observées. D'une manière simplifiée, la distribution d'une variable aléatoire donne la probabilité que le paramètre prenne une valeur particulière. L'observation de la distribution permet de déterminer la valeur « la plus probable », c'est à dire le sommet de la distribution. Ainsi la distribution peut être assimilée à un histogramme.

Si l'on connaissait parfaitement la valeur du paramètre que l'on cherche à estimer, la distribution serait un trait vertical. Une seule valeur à une probabilité de 100 %, toutes les autres une probabilité de 0 % d'être la vraie valeur du paramètre estimé. En pratique, cela n'est jamais le cas, et l'étalement de la distribution traduit l'imprécision avec laquelle la vraie valeur est connue.

La distribution du paramètre que l'on cherche à estimer est déterminée à partir des données observées (comme tout procédé d'estimation). Cependant, les méthodes Bayésiennes ont un prix à payer qui est celui de nécessiter une hypothèse *a priori* sur la distribution du paramètre. Le résultat produit, la distribution estimée du paramètre, dépend donc à la fois des données et de l'*a priori* qui a été choisi. Ces méthodes donnent une distribution *a posteriori* (c'est à dire après l'observation des données) à partir d'une distribution *a priori* (telle qu'on l'imagine *a priori*, avant d'obtenir toute information par les données). La nécessité de donner une distribution *a priori* de l'élément que l'on ignore et que l'on cherche à connaître à travers l'observation est la principale difficulté de cette approche. Ce point est important, car le résultat (la distribution *a posteriori* estimée) peut changer fortement en fonction de l'*a priori* qui est fixé.

Heureusement, il est possible de trouver des *a priori* « non informatifs », c'est à dire qui n'introduisent que très peu d'information. Ainsi, ces *a priori* ne conditionnent pas le résultat. Ces distributions *a priori* non informatives disent en quelque sorte, qu'*a priori* (c'est à dire avant d'observer la première donnée), toute valeur est possible pour la vraie valeur du paramètre et qu'il n'est donc pas justifié d'en favoriser une plus qu'une autre.

L'obtention de la distribution *a posteriori* à partir de la distribution *a priori* s'effectue à l'aide du théorème de Bayes, d'où le nom de méthode Bayésienne. Le théorème de Bayes spécifie que la

distribution *a posteriori* $f(\theta|d)$ du paramètre θ compte tenu des données d observées est

$$f(\theta|d) = \frac{l(d|\theta) f_0(\theta)}{g(d)}$$

Avec $f_0(\theta)$ qui désigne la distribution *a priori* du paramètre et $l(d|\theta)$ qui est la vraisemblance des données en fonction de la valeur du paramètre. La vraisemblance est la probabilité d'obtenir

les données qui ont été effectivement observées pour une valeur donnée du paramètre. $g(d)$ est la distribution inconditionnelle des données (aussi appelée distribution marginale des données ou distribution des prédictions *a priori*).

Avec les techniques de MCMC, la transformation à l'aide de l'information apportée par les données de la distribution *a priori* pour obtenir la distribution *a posteriori* ne va pas se faire de façon analytique (à l'aide de formule) mais en recourant au calcul intensif grâce à un ordinateur. L'estimation de la valeur du paramètre θ du modèle s'effectue en répétant en boucle le même processus un très grand nombre de fois jusqu'à l'obtention d'une stabilisation de l'estimation de la distribution.

Le processus d'estimation consiste à produire à l'aide de génération de nombres aléatoires une distribution *a posteriori* à partir d'une distribution *a priori* des données. La distribution *a posteriori* obtenue devient la distribution *a priori* de la boucle suivante, et ainsi de suite. Un théorème assure la convergence du processus vers la vraie distribution *a posteriori* du paramètre à estimer.

Le processus produit donc une séquence (chaîne) de nombres qui après exclusion des premières valeurs (*burn-in*) représente un échantillon aléatoire de la distribution *a posteriori*. L'analyse

descriptive de cette séquence (moyenne, médiane, quantile, représentation de l'histogramme) permet d'estimer de manière empirique la distribution du paramètre d'intérêt.

En conclusion, l'utilisation de ces techniques utilise en entrée : les données observées, une distribution *a priori* et la première valeur du paramètre. En sortie est produite une longue séquence de nombre (chaîne de Markov) dont la partie finale est une estimation empirique de la distribution. L'assurance que la séquence obtenue est un échantillon aléatoire correct de la distribution *a posteriori* du paramètre dépend de la valeur initiale et de la distribution *a priori* choisie ainsi que de la convergence (relative stabilité) de la séquence obtenue. Une absence de convergence non diagnostiquée et/ou un choix malheureux de l'*a priori* ou de la valeur initiale conduit à des estimations erronées.

4.5.2 Modèle de Caldwell *et al.*

■ Méthode statistique

Le modèle proposé par Caldwell *et al.* (20) est une généralisation de celui proposé par Higgins et Whitehead (23).

Soit une méta-analyse où k traitements ont été comparés entre eux dans les essais. Pour chaque essai j , soit r_{jk} le nombre d'événements observés dans le bras du traitement k dont l'effectif est n_{jk} . Le nombre d'événements suit une distribution binomiale

$$r_{jk} \approx B(p_{jk}, n_{jk})$$

où p_{jk} est la fréquence (le risque) de survenue de l'événement, critère de jugement.

L'utilisation d'un odds ratio pour représenter les effets traitement revient à modéliser des logit de la façon suivante :

$$\text{logit}(p_{jk}) = \mu_{jb} \text{ lorsque } k=b, b \text{ désignant le traitement de référence des comparaisons}$$

$$\text{et } \text{logit}(p_{jk}) = \mu_{jb} + \delta_{jkb} \text{ quand } k \text{ est différent de } b.$$

μ_{jb} représente en quelque sorte une valeur de base de l'odds du critère de jugement dans l'essai j .

Ce modèle peut être décliné en deux variantes reposant soit sur un modèle fixe des effets, soit sur un modèle d'effets aléatoires. L'estimation s'effectue par MCMC à l'aide du logiciel Winbugs par exemple et nécessite l'introduction, dans le processus d'estimation, de distributions *a priori* non informatives.

Cette approche a par exemple été utilisée pour comparer les fibrinolytiques à la phase aiguë de l'infarctus du myocarde (20).

4.6 Estimation à l'aide d'un modèle linéaire mixte

■ Méthode statistique

L'utilisation d'un modèle linéaire mixte pour la réalisation des comparaisons indirectes à travers un réseau de comparaisons a été proposée en 2002 par Lumley (21).

L'effet traitement estimé à partir de l'essai k comparant les traitements i et j est noté Y_{ijk} avec comme variance estimée σ_{ijk}^2 .

Le modèle des Y_{ijk} comporte trois composantes en plus de l'erreur d'échantillonnage :

1. le méta vrai effet des traitements i et j qui est désigné par μ_i et μ_j respectivement. Le méta vrai effet est la valeur autour de laquelle fluctuent les vrais effets traitement développés dans

chaque essai. L'effet observé provient du vrai effet plus ou moins entaché de l'erreur d'échantillonnage (liée à la fluctuation d'échantillonnage).

- l'effet aléatoire η_{ik} et η_{jk} qui représente la différence entre le vrai effet traitement de l'essai k et les méta-valeurs correspondantes. Les 2 effets η_{ik} et η_{jk} ont comme variance τ^2 . Le vrai effet du traitement i dans l'essai k est donc $\mu_i + \eta_{ik}$. Y_{ijk} est l'estimation de la différence entre les traitements i et j dans l'essai k , c'est donc une estimation de la différence entre les vrais effets $\mu_i + \eta_{ik}$ et $\mu_j + \eta_{jk}$.
- un second effet aléatoire ξ_{ij} représentant le changement de l'effet du traitement i quand il est comparé à j . ξ_{ij} permet ainsi de capter l'incohérence éventuelle du réseau d'évidence. Sa variance ω permet de mesurer l'incohérence du réseau.

En conclusion, le modèle est donc

$$Y_{ijk} = (\mu_i + \eta_{ik}) - (\mu_j + \eta_{jk}) + \xi_{ij} + \varepsilon_{ijk}$$

$$\varepsilon_{ijk} = N(0, \sigma_{ijk}^2)$$

$$\eta_{ik} \approx N(0, \tau^2)$$

$$\xi_{ij} \approx N(0, \omega^2)$$

Où ε_{ijk} désigne l'erreur aléatoire d'échantillonnage qui entache au niveau de l'essai k l'estimation de la différence entre le traitement i et le traitement j .

Encart - Modèle mixte, hiérarchique

En statistique, un effet est la modification qu'induit un facteur dans la valeur d'une variable aléatoire. Dans un essai, le facteur traitement entraîne une modification dans la valeur de la variable utilisée comme critère de jugement. La plupart du temps, les facteurs sont considérés comme étant fixes. L'effet traitement est donc la modification constante qu'induit en moyenne le traitement sur la variable aléatoire utilisée comme critère de jugement.

Les modèles à effets mixtes, appelés aussi modèles à effets aléatoires, permettent de prendre en compte plusieurs niveaux de fluctuation/erreurs aléatoires (les effets aléatoires) et ainsi de considérer que des effets peuvent varier aussi de manière aléatoire.

La qualification aléatoire de l'effet signifie que cet effet varie d'une unité statistique à l'autre, de manière aléatoire suivant une certaine distribution.

En méta-analyse, la modélisation à l'aide d'un modèle fixe représente la réalité en considérant que l'effet Y observé dans l'essai k est égal au vrai effet traitement, plus les fluctuations aléatoires d'échantillonnage (c'est à dire l'erreur aléatoire se produisant au niveau de l'essai entre la valeur observée et la vraie valeur).

Dans ce modèle, il n'y a donc qu'un seul effet (le vrai effet du traitement) et cet effet est fixe pour tous les essais (d'où le nom de modèle à effet fixe).

Un modèle aléatoire de méta-analyse stipulera que le vrai effet du traitement est variable d'un essai à l'autre en fluctuant de manière aléatoire autour d'une valeur qui est un méta-paramètre. L'effet qui est recherché au niveau de l'essai n'est donc plus fixe mais aléatoire, d'où le nom de modèle à effet aléatoire. La méta-analyse cherche alors à estimer le méta-paramètre. Ce modèle stipule donc une hétérogénéité de l'effet du traitement (l'effet varie d'un essai à l'autre de manière aléatoire).

Dans ce modèle aléatoire, nous avons donc deux niveaux de variabilité due au hasard :

- les fluctuations d'échantillonnage (comme avec le modèle fixe) décrivant la différence entre la

- vraie valeur de l'effet traitement dans un essai et l'effet observé
- les fluctuations du vrai effet entre les essais.

■ Exemple d'application

Cette méthode a été employée dans la méta-analyse de Psaty *et al.* sur les antihypertenseurs de première intention (24). L'objectif était de compléter et d'estimer toutes les comparaisons possibles du réseau. Par exemple, aucun essai ne compare directement les *Low-Dose Diuretics* aux *Angiotensin Receptor Blockers*. La comparaison indirecte réalisée de manière implicite par la méthode a permis de documenter l'efficacité relative de ces deux types de traitement (Figure 14 et Figure 15).

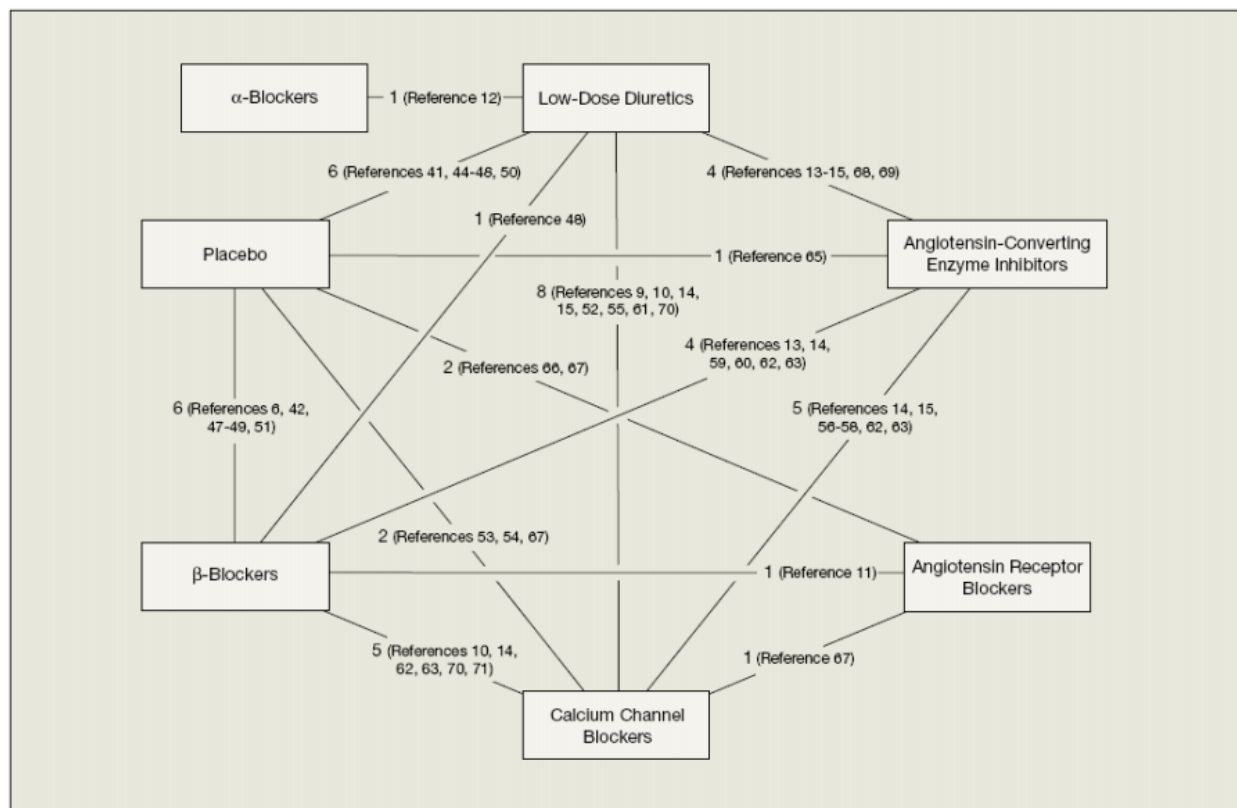


Figure 14 - Réseau de comparaisons pris en compte

Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, *et al.* Health outcomes associated with various antihypertensive therapies used as first-line agents. A network meta-analysis. *JAMA* 2003;289(19):2534-44. Copyright © 2003 American Medical Association. All rights reserved

Figure 2. Network Meta-analysis of First-Line Treatment Strategies in Randomized Controlled Clinical Trials in Hypertension

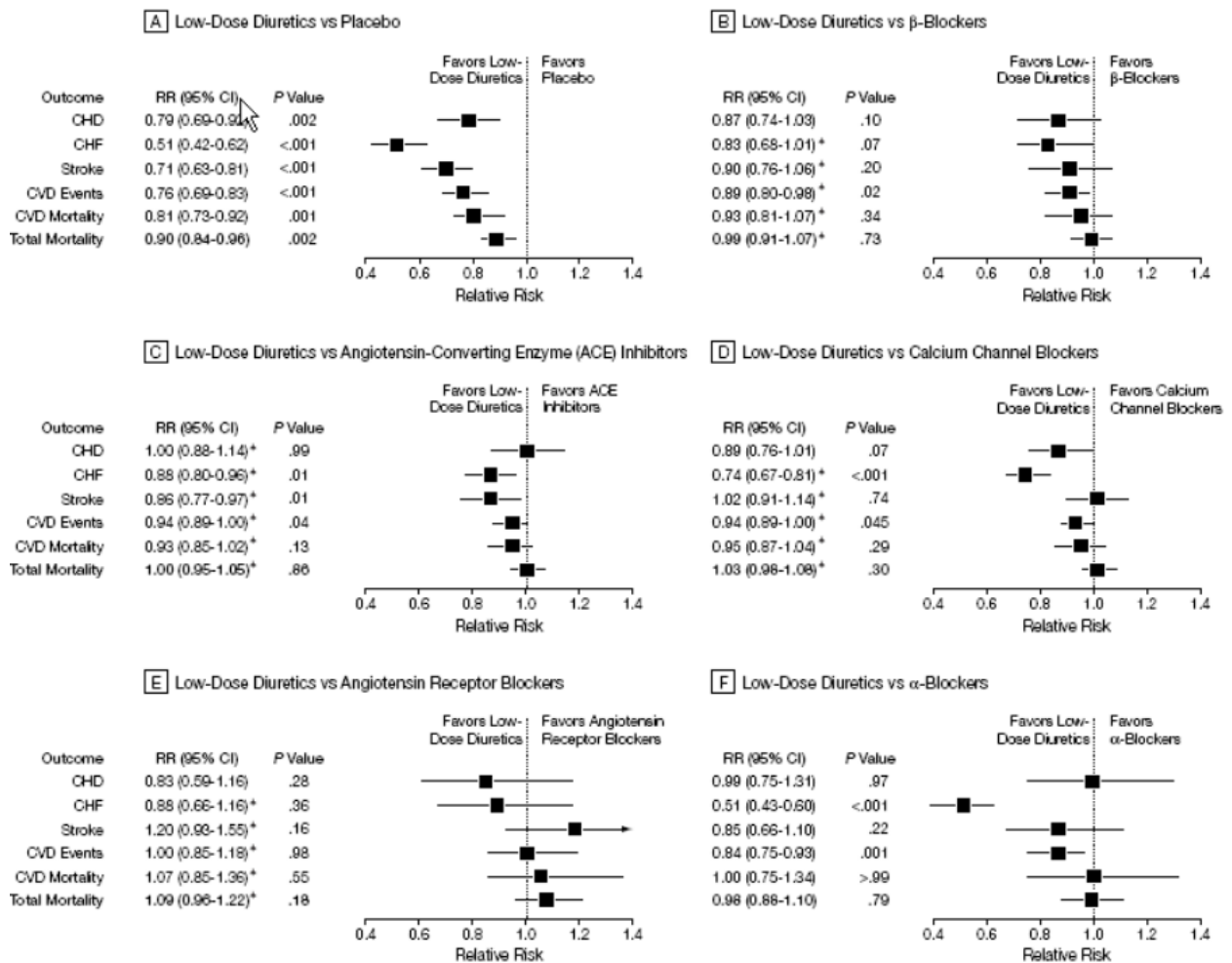


Figure 15 - Résultats de la méta-analyse en réseau

Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, *et al.* Health outcomes associated with various antihypertensive therapies used as first-line agents. A network meta-analysis. *JAMA* 2003;289(19):2534-44. Copyright © 2003 American Medical Association. All rights reserved

4.7 Méta-régression

4.7.1 Généralités

Une autre méthode utilisable pour réaliser des comparaisons indirectes est la méta-régression.

La méta-régression est une technique standard de méta-analyse qui consiste à modéliser l'effet observé dans les essais à l'aide de covariables (25-29). Il s'agit souvent d'une régression linéaire avec comme variables explicatives des covariables décrivant les traitements évalués (dose, molécules,...), les patients inclus dans les essais, les caractéristiques méthodologiques des essais, etc.

La méta-régression a par exemple été utilisée pour rechercher s'il existait une relation entre la réduction de morbi mortalité apportée par les statines et l'intensité de la baisse du LDL obtenue (30).

Pour des raisons statistiques, il est préférable d'utiliser le logarithme de l'odds ratio comme mesure de l'effet du traitement. Le modèle de méta-régression s'écrit alors :

$$\log(OR_i) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \varepsilon_i$$

où OR_i désigne l'odds ratio mesuré dans l'essai i , $x_{1,i}$ la valeur de la covariable 1 dans cet essai i , etc. pour toutes covariables prises en considération. ε_i représente le résidu aléatoire qui est considéré comme distribué normalement avec comme variance résiduelle σ^2 .

ε_i représente ainsi la part de la variance de l'effet entre études non expliquée par les covariables considérées.

L'estimation des coefficients β_1, β_2, \dots peut s'effectuer par différentes méthodes d'ajustement comme les moindres carrés, les moindres carrés pondérés, etc. Thompson et Sharp recommandent cependant d'utiliser la méthode du maximum de vraisemblance restreinte (*Restricted Maximum Likelihood Method*) (28).

Un modèle aléatoire peut aussi être utilisé. En estimant simultanément plusieurs comparaisons indirectes, la méta-régression s'apparente aussi aux méthodes basées sur la modélisation d'un réseau d'essais.

4.7.2 Méta-régression et effet traitement fixe

■ Méthode statistique

L'utilisation de la technique de méta-régression pour faire des comparaisons indirectes s'effectue en codant les traitements à comparer à l'aide des covariables.

Considérons une situation où 3 traitements A,B,C ont donné lieu à une série d'essais contre placebo. La réalisation des comparaisons indirectes de l'efficacité de ces 3 traitements s'effectuera en ajustant sur les données fournies par ces séries d'essais le modèle suivant :

$$\log(OR_i) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

Le codage du traitement évalué est effectué à l'aide de deux covariables X_1 et X_2 (technique des *dummy variable*) en prenant le traitement A comme référence :

Essais	X_1	X_2
A versus placebo	0	0
B versus placebo	1	0
C versus placebo	0	1

Ainsi, le coefficient β_1 estimera l'effet de B par rapport à A et le coefficient β_2 celui de C par rapport à A. Tout autre traitement aurait pu être pris comme référence bien entendu.

En effet, pour le traitement A (soit $x_1=0$ et $x_2=0$) on a

$$\begin{aligned} \text{Log}(OR_A) &= \beta_0 + \beta_1 0 + \beta_2 0 \\ &= \beta_0 \end{aligned}$$

Et pour le traitement B

$$\log(OR_B) = \beta_0 + \beta_1$$

Ainsi

$$\begin{aligned} \text{Log}(OR_B) - \text{Log}(OR_A) &= \beta_0 + \beta_1 - \beta_0 \\ &= \beta_1 \end{aligned}$$

Comme $\text{Log}(B) - \text{Log}(A) = \text{Log}\left(\frac{B}{A}\right)$, on a

$$\frac{OR_B}{OR_A} = \exp(\beta_1)$$

Or $OR_B / OR_A = OR_{B/A}$, l'odds ratio de B par rapport à A. En effet :

$$OR_B = OR_{B/PBO} = \frac{\text{odds}_B}{\text{odds}_{PBO}}$$

De même

$$OR_A = \frac{odds_A}{odds_{PBO}}$$

Ce qui donne

$$\begin{aligned}\frac{OR_B}{OR_A} &= \frac{odds_B}{odds_{PBO}} \frac{odds_{PBO}}{odds_A} \\ &= \frac{odds_B}{odds_A} \\ &= OR_{B/A}\end{aligned}$$

En conclusion on a donc

$$OR_{B/A} = \exp(\beta_1) \text{ et } OR_{C/A} = \exp(\beta_2)$$

La méta-régression permet aussi de faire des tests statistiques. Ainsi pour savoir si B est statistiquement différent de A il suffit de tester l'hypothèse $\beta_1=0$. De même, l'estimation de l'odds ratio de B/A est assortie de son intervalle de confiance.

■ Exemple d'application

La méta-régression avec un modèle à effet mixte a par exemple été utilisée par Eckert et Lancon pour comparer la duloxétine à la fluoxétine et à la venlafaxine (31).

4.7.3 Méta-régression et effet traitement aléatoire (méthode Bayésienne)

Nixon *et al.* proposent une implémentation de la méta-régression pour faire des « mixed treatment comparisons » à l'aide d'un modèle Bayésien estimé par MCMC (32). Par rapport à la méta-régression simple, leur modèle inclut un effet traitement aléatoire et améliore ses performances en ajustant sur des covariables.

4.8 Méthodes pour les critères de jugement continus

Les méthodes proposées jusqu'à présent ne s'appliquent qu'aux critères de jugement binaires. Parmi ces méthodes, une seule est directement transposable aux critères continus : celle des comparaisons indirectes ajustées. Il suffit de remplacer le risque relatif par une mesure adaptée à la méta-analyse des critères continus, la différence pondérée des moyennes (*weighted mean difference*, WMD) ou l'ampleur d'effet (*effect size*).

Par contre, l'approche par « evidence network » n'offre pas de transposition directe aux critères continus et va nécessiter des développements spécifiques.

La méta-analyse de Packer *et al.* comparant carvedilol et métoprolol sur la fraction d'éjection ventriculaire dans l'insuffisance cardiaque donne un exemple d'utilisation des comparaisons indirectes ajustées avec un critère de jugement continu (33).

4.9 Comparaison des différentes méthodes

Il n'existe qu'un travail ayant comparé les performances statistiques de ces différentes méthodes. Glenny *et al.* montrent (chapitre 5) en utilisant les données de l'essai IST (*International Stroke Trial*) pour créer des pseudo essais que ces différentes méthodes produisent des estimations très proches les unes des autres (12). Les méthodes qu'ils ont comparées sont les suivantes :

- comparaison indirecte ajustée à l'aide d'un modèle fixe ;
- comparaison indirecte ajustée à l'aide d'un modèle aléatoire (DerSimonian et Laird) ;
- régression logistique ;
- méta-régression à effet aléatoire ;
- méthode naïve par comparaison des bras.

Les résultats montrent qu'avec ces données, toutes les méthodes donnent des estimations ponctuelles très voisines les unes des autres. Cependant, pour l'estimation des erreurs standard (et

donc pour la réalisation des tests statistiques), étant donné l'existence d'au moins trois niveaux de variabilité dans la question posée par les comparaisons indirectes, les méthodes basées sur un modèle aléatoire semblent plus raisonnables.

Le Tableau 8 tente de comparer les avantages et les inconvénients théoriques des différentes méthodes. A côté des aspects statistiques, il convient aussi de noter que seules les comparaisons indirectes ajustées ont fait l'objet d'une étude empirique de leur validité par comparaison de leurs résultats à ceux des comparaisons directes (cf. chapitre 6). De plus, il s'agit de la seule méthode qui puisse donner des résultats en utilisant le risque relatif. Les autres approches conduisent à des odds ratio. Bien que l'odds ratio ait de bonnes propriétés mathématiques et statistiques, son interprétation reste difficile. Le risque relatif est d'un abord intuitif et semble de ce fait plus adapté pour les questions d'évaluation thérapeutique et les méta-analyses (34).

En conclusion, en l'absence de différence connue dans les performances des différentes méthodes, il est difficile de recommander une méthode plus qu'une autre. Deux approches cependant peuvent être mises en avant :

1. les comparaisons indirectes ajustées, en raison de leur simplicité, de leur transparence (l'origine des résultats mis en avant est claire), de la possibilité d'exprimer les résultats en risque relatif et qu'il s'agit de la seule méthode disponible en standard en cas de critère continu ;
2. les méthodes Bayésiennes basées sur un réseau (*evidence network*) en raison de la grande flexibilité du modèle (permettant une modélisation fine et souple des données, adaptable à des cas particuliers), de l'estimation de l'incohérence, de la prise en compte des essais multibras.

Tableau 8 - Avantages et inconvénients théoriques des différentes méthodes

Méthode	Avantages	Inconvénients
Comparaison indirecte ajustée	<ul style="list-style-type: none"> • Simplicité des calculs • Transparence • Utilisable avec le risque relatif • Transposable aux critères continus • Une étude empirique de validité 	<ul style="list-style-type: none"> • Impossibilité de tester l'incohérence • Comparaisons deux à deux multiples (inflation du risque α) si comparaison indirecte de plus de 2 traitements
Méta-régression	<ul style="list-style-type: none"> • Ajustement sur d'autres covariables introduisant de l'hétérogénéité • Ne nécessite pas forcément de MCMC • Estimable aussi par MCMC 	<ul style="list-style-type: none"> • Pas d'estimation de l'incohérence • Donne des odds ratio • Pas d'étude empirique de validité
Méthodes basées sur les « evidence network »		
Lumley, 2002 (21)	<ul style="list-style-type: none"> • Test de l'incohérence • Ne nécessite pas de MCMC 	<ul style="list-style-type: none"> • Donne des odds ratio • Pas d'étude empirique de validité
Lu et Ades, 2004 (19)	<ul style="list-style-type: none"> • Test de l'incohérence • Prise en compte des essais multibras • Possibilité d'utiliser un modèle aléatoire 	<ul style="list-style-type: none"> • Méthode d'estimation délicate requérant un haut degré d'expertise (analyse de la convergence entre autres) • Approche Bayésienne donc nécessitant l'introduction d'information <i>a priori</i> • Donne des odds ratio • Pas d'étude empirique de validité
Caldwell <i>et al.</i> , 2005 (20)	<ul style="list-style-type: none"> • Test de l'incohérence • Prise en compte des essais multibras • Possibilité d'utiliser un modèle aléatoire 	<ul style="list-style-type: none"> • Méthode d'estimation délicate requérant un haut degré d'expertise (analyse de la convergence entre autres) • Approche Bayésienne donc nécessitant l'introduction d'information <i>a priori</i> • Donne des odds ratio • Pas d'étude empirique de validité

MCMC : *Monte Carlo Markov Chain*

5. Utilisation actuelle des comparaisons indirectes

5.1 Données bibliographiques

Une revue systématique de l'utilisation des méthodes de comparaisons indirectes a été publiée en 2005 par Glenny *et al.* (12). Dans ce travail, les revues systématiques antérieures à 1999 incluant des comparaisons indirectes ont été recherchées à l'aide de la *Database of Abstracts of Reviews of Effects* (DARE) et de la *Cochrane Database of Systematic Reviews* (CDSR).

Trente-six travaux incluant des comparaisons indirectes ont été retrouvés ; treize d'entre eux s'intéressaient simultanément aux comparaisons directes et indirectes. Dans 11 cas sur 36 (31 %), les comparaisons indirectes étaient réalisées de manière impropre par comparaison naïve des bras actifs. Dans les 25 autres cas (69 %), la méthode des comparaisons indirectes ajustées était employée.

5.2 Quelques publications récentes

La recherche de Glenny *et al.* étant maintenant un peu ancienne (recherche arrêtée en 1999), nous avons réalisé une recherche complémentaire non exhaustive de travaux plus récents (12) (cf. § 3). Cette recherche n'a pas été entreprise pour quantifier la fréquence d'utilisation de ces techniques, mais simplement pour brosser un panorama des techniques actuelles et des domaines d'utilisation.

La recherche bibliographique n'est pas simple en l'absence de terme unique permettant de désigner le concept de comparaisons indirectes. De plus, les comparaisons indirectes sont utilisées au sein de méta-analyses sans être mises en avant dans le titre ou dans l'abstract, ce qui rend encore plus difficile leur identification. Pour cette étude, la recherche bibliographique a employé les termes suivants :

- *Indirect comparison* ;
- *Mixed treatment comparison* ;
- *Multiple treatment comparison*.

Le résultat de ce survol est représenté dans le Tableau 9. Des méthodes naïves sont encore rencontrées. Ces travaux utilisent en majorité des comparaisons indirectes ajustées. De plus en plus de méthodes Bayésiennes sur réseau de comparaisons sont rencontrées.

Tableau 9 - Exemples récents de travaux comportant une approche de comparaison indirecte

Titre de la publication (réf.)	Année	Méthode
Gabapentin versus tricyclic antidepressants for diabetic neuropathy and post-herpetic neuralgia: discrepancies between direct and indirect meta-analyses of randomized controlled trials (35)	2009	Comparaison indirecte ajustée
Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis (36)	2009	Méta-analyse en réseau (méthode bayésienne)
Carbapenems versus other beta-lactams in the treatment of hospitalised patients with infection: a mixed treatment comparison (37)	2009	<i>Mixed Treatment Comparison</i>
An indirect comparison of the efficacy and safety of factor Xa inhibitors with thrombin inhibitors in preventing various thromboembolism after hip or knee surgery (38)	2008	<i>Mixed Treatment Comparison</i>
Meta-analysis of the combination of TNF inhibitors plus MTX compared to MTX monotherapy, and the adjusted indirect comparison of TNF inhibitors in patients suffering from active rheumatoid arthritis (39)	2008	Comparaison indirecte ajustée
Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer (40)	2008	Méta-analyse en réseau (méthode bayésienne)

Titre de la publication (réf.)	Année	Méthode
Direct and indirect comparison meta-analysis demonstrates the superiority of sirolimus- versus paclitaxel-eluting stents across 5854 patients (41)	2007	Comparaison indirecte ajustée
Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis (18)	2007	Méthode Bayésienne
Need for additional calcium to reduce the risk of hip fracture with vitamin D supplementation: evidence from a comparative metaanalysis of randomized controlled trials (42)	2007	Comparaison indirecte ajustée
Magnetic resonance colonography vs computed tomography colonography for the diagnosis of colorectal cancer: an indirect comparison (43)	2007	Comparaison indirecte ajustée des diagnostics odds ratios (DOR)
What is the optimal pharmacological prophylaxis for the prevention of deep-vein thrombosis and pulmonary embolism in patients with acute ischemic stroke? (44)	2007	Comparaison naïve des estimations ponctuelles
Adjusted indirect comparison of celecoxib versus rofecoxib on cardiovascular risk (45)	2006	Comparaison indirecte ajustée
Duloxetine compared with fluoxetine and venlafaxine: use of meta-regression analysis for indirect comparisons (31)	2006	Méta-régression
Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer (46)	2006	<i>Mixed Treatment Comparison</i>
Are statins created equal? Evidence from randomized trials of pravastatin, simvastatin, and atorvastatin for cardiovascular disease prevention (47)	2006	Comparaison indirecte ajustée
Efficacy of drug eluting stents in patients with and without diabetes mellitus: indirect comparison of controlled trials (48)	2006	Méta-régression et comparaisons indirectes ajustées
Efficacy of PDE-5-inhibitors for erectile dysfunction. A comparative meta-analysis of fixed-dose regimen randomized controlled trials administering the International Index of Erectile Function in broad-spectrum populations (49)	2006	Comparaison indirecte ajustée (avec Bonferroni)
Investigating heterogeneity in studies of resting energy expenditure in persons with HIV/AIDS: a meta-analysis (50)	2005	Méthode Bayésienne
Adjusted indirect comparison of intracoronary drug-eluting stents: evidence from a metaanalysis of randomized bare-metal-stent-controlled trials (51)	2005	Comparaison indirecte ajustée
Ximelagatran compared with warfarin for the prevention of systemic embolism and stroke. An imputed placebo analysis (52)	2005	Méthode inappropriée
Indirect comparison of interventions using published randomised trials: systematic review of PDE-5 inhibitors for erectile dysfunction (53)	2005	Comparaison indirecte ajustée
A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents (54)	2004	Méthode Bayésienne
Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials (55)	2004	Comparaison indirecte ajustée
Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation (56)	2004	Comparaison indirecte ajustée
Ultrasonic locating devices for central venous cannulation: meta-analysis (57)	2003	Méthode non identifiée
Indirect comparison meta-analysis of aspirin therapy after coronary surgery (58)	2003	Comparaison indirecte ajustée

Titre de la publication (réf.)	Année	Méthode
Effectiveness of nifedipine <i>versus</i> atosiban for tocolysis in preterm labour: a meta-analysis with an indirect comparison of randomised trials (59)	2003	Comparaison indirecte ajustée
Treatment of open fractures of the shaft of the tibia. A systematic overview and meta-analysis (60)	2001	Comparaison indirecte ajustée
Comparative effects of carvedilol and metoprolol on left ventricular ejection fraction in heart failure: results of a meta-analysis (33)	2001	Comparaison indirecte ajustée sur critères continus

5.3 Utilisation dans des rapports d'évaluation technologique ou des recommandations

L'approche de comparaison indirecte a été utilisée dans quelques travaux récents d'évaluation technologique (*Health Technology Assessment*) ou de recommandation (liste non exhaustive) :

- Manifestations and management of chronic insomnia in adults, *Agency for Healthcare Research and Quality*, 2005 (61)
- Metal-on-metal total hip resurfacing, TEC, *BlueCross BlueShield Association*, 2007 (62)
- Off-label uses of bevacizumab: renal cell carcinoma and other miscellaneous non-colorectal cancer indications, TEC, *BlueCross BlueShield Association*, 2006 (63)
- Botulinum toxin type A therapy for cervical dystonia, *Cochrane Review*, 2005 (64)
- Topical antibiotics without *steroids* for chronically discharging ears with underlying eardrum perforations, *Cochrane Review*, 2005 (65)
- Docetaxel for the adjuvant treatment of early node-positive breast cancer, *National Institute for Health and Clinical Excellence*, 2006 (66)
- Pemetrexed for the treatment of non-small-cell lung cancer, *National Institute for Health and Clinical Excellence*, 2007 (67)
- Adalimumab for the treatment of psoriatic arthritis, *National Institute for Health and Clinical Excellence*, 2007 (68)
- Rituximab for the treatment of rheumatoid arthritis, *National Institute for Health and Clinical Excellence*, 2007 (69)
- Comparative effectiveness and safety of oral diabetes medications for adults with type 2 diabetes, *Agency for Healthcare Research and Quality*, 2007 (70)
- Infliximab for the treatment of adults with psoriasis, *National Institute for Health and Clinical Excellence*, 2008 (71)
- Varenicline for smoking cessation, *National Institute for Health and Clinical Excellence*, 2007 (72)

6. Validité des comparaisons indirectes

6.1 Qu'est-ce que la validité des comparaisons indirectes ?

Classiquement, on considère que l'approche de référence pour comparer deux traitements actifs est celle des comparaisons directes. Par exemple, ICH E10 (*choice of control group in clinical trials*) mentionne (§ 2.1.7.4) : « Placebo-controlled trials lacking an active control give little useful information about comparative effectiveness, information that is of interest and importance in many circumstances. Such information cannot reliably be obtained from cross-study comparisons, as the conditions of the studies may have been quite different. »

En partant de ce postulat, pour être valides, les comparaisons indirectes doivent donner, pour une situation donnée, les mêmes résultats que ceux obtenus par comparaisons directes.

L'étude de la validité des comparaisons indirectes peut alors se ramener à l'étude des différences entre les résultats produits par les deux méthodes. Cette approche nécessite de disposer de domaines où les deux méthodes sont applicables, c'est à dire des domaines où il existe des comparaisons directes et du matériel permettant la réalisation des comparaisons indirectes.

6.2 Les études de validité

6.2.1 Etudes de Song *et al.*, 2003 et de Glenny *et al.*, 2005

L'étude la plus aboutie abordant la question de la validité des comparaisons indirectes est celle de Song *et al.* publiée dans le BMJ en 2003 (14). Ce travail est aussi rapporté avec plus de détails dans le document du NHS R&D HTA Programme coordonné par Glenny *et al.* (12).

Dans ce travail les comparaisons indirectes ont été réalisées avec la méthode « adjusted indirect comparison ». La mesure de discordance « discrepancy » utilisée est la différence $\Delta = T_{direct} - T_{indirect}$ où T_{direct} et $T_{indirect}$ représentent respectivement l'estimation directe et indirecte de la différence d'effet entre les deux traitements. Des domaines dans lesquels existaient à la fois des essais de comparaison directe et la possibilité de faire des comparaisons indirectes ont été recherchés. Les estimations T_{direct} et $T_{indirect}$ ont été obtenues par méta-analyses utilisant un modèle aléatoire. Pour les critères de jugement continus les estimations de l'effet traitement ont porté sur la différence des moyennes tandis que pour les critères binaires, le logarithme du risque relatif a été utilisé.

Il est possible de calculer un intervalle de confiance à 95 % de la différence Δ . En cas de parfaite concordance entre les estimations directe et indirecte, la différence Δ est nulle.

La Figure 16 représente les degrés de discordance et leur intervalle de confiance observés pour les 44 méta-analyses considérées appartenant à différents champs thérapeutiques.

L'intervalle de confiance de la différence Δ permet d'intégrer dans la comparaison des résultats des deux méthodes les incertitudes statistiques existant à leur niveau. Il convient en effet impérativement de tenir compte de ces incertitudes car la référence (T_{direct}) n'est pas connue avec exactitude mais estimée avec une marge d'erreur. Cet intervalle de confiance représente ainsi la plage de valeur raisonnablement compatible avec la vraie différence entre les deux méthodes compte tenu de la marge d'erreur possible sur T_{direct} et $T_{indirect}$. Lorsque cet intervalle n'englobe pas zéro, il est possible de conclure qu'il existe une réelle différence entre les deux estimations T_{direct} et $T_{indirect}$, même malgré les incertitudes pesant sur T_{direct} et $T_{indirect}$. Par contre, si l'intervalle contient zéro, la différence observée peut induire des erreurs des estimations T_{direct} et $T_{indirect}$ et il n'est donc pas possible de conclure à une réelle différence de résultats entre les deux estimations.

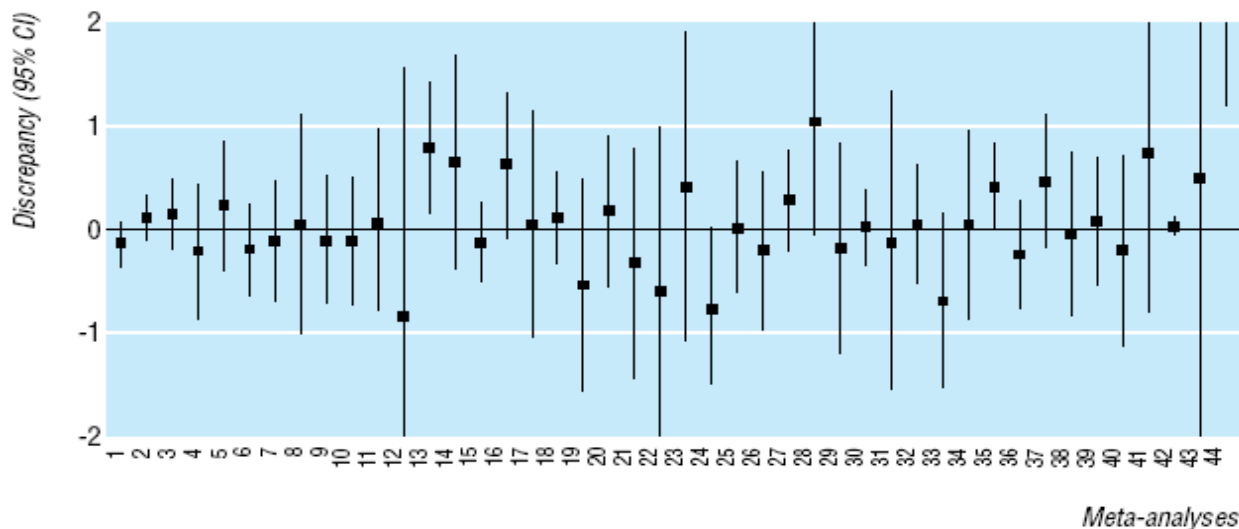


Figure 16 - Discordance (*discrepancy*) entre les résultats des comparaisons directes et indirectes pour les 44 méta-analyses étudiées par Song *et al.*, 2003 (14)
 Reproduced from Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326(7387):472. Copyright 2003, with permission from BMJ Publishing Group Ltd.

Dans cette étude, seulement 3 cas sur 44 révèlent un désaccord statistiquement significatif entre l'estimation des comparaisons directes et celles des comparaisons indirectes. Deux de ces cas de désaccord concernent les traitements par antihistaminiques H2 ou inhibiteurs de la pompe à protons. Le dernier cas concerne le paracétamol et son association avec la codéine dans le traitement de la douleur chirurgicale (12,73) et révèle une discordance totale entre les deux estimations (Tableau 10).

Tableau 10 - Explication des discordances observées entre comparaisons directes et indirectes dans le cas de l'association paracétamol et codéine pour le traitement de la douleur chirurgicale dans le travail de Glenny *et al.*, 2005 (12)

	Score d'intensité de la douleur	
	Différence des moyennes entre les groupes [IC 95 %]	
	Comparaison directe	Comparaison indirecte ajustée
Ensemble des essais	6,97 [3,56 ; 10,37] N=13	-1,16 [-6,95 ; 4,64] N=12 et 31
Essais de chirurgie dentaire seulement	7,07 [3,37 ; 10,78] N=11	-1,40 [-8,27 ; 5,46] N=7 et 15
Essais utilisant 600-650 mg de paracétamol et 60 mg de codéine	7,28 [3,69 ; 10,87] N=10	5,72 [-5,37 ; 16,81] N=2 et 12

Cependant ces deux résultats sont obtenus en combinant des essais plutôt hétérogènes en termes de chirurgie et de doses (variant de 300 mg à 1 000 mg pour le paracétamol et de 30 à 60 mg pour la codéine).

Le type de chirurgie ne semble pas pouvoir expliquer la discordance étant donné que la restriction de l'analyse aux essais de douleurs post chirurgie dentaire (les plus nombreux) ne la fait pas disparaître. Par contre, comparaisons directes et indirectes produisent des résultats non différents quand l'analyse porte sur des essais utilisant les traitements à la même dose (600-650 mg de paracétamol et 60 mg de codéine). L'enseignement de cet exemple est assez trivial. Il met simplement en exergue la nécessité, en méta-analyse (destinée ou non à réaliser des comparaisons indirectes), de faire des regroupements homogènes d'essais vis à vis de la question posée et d'effectuer des analyses de sensibilité. Ici, il est possible d'évoquer la relation entre la dose et l'effet pour expliquer les discordances observées, même si d'autres facteurs non identifiés peuvent concourir aussi à ce résultat.

6.2.2 Etude de Bucher *et al.*, 1997

Le premier travail de comparaison a été réalisé en 1997 par Bucher *et al.* (74). Cette étude porte sur la prévention des pneumonies à *Pneumocystis Carinii* chez des patients infectés par le VIH et la comparaison de l'association triméthopime-sulfaméthoxazole à l'association dapsonne-pyriméthamine. Huit essais effectuant la comparaison directe sont disponibles conduisant, en méta-analyse, à un odds ratio de 0,64. L'estimation à l'aide d'une comparaison indirecte ajustée utilisant les résultats de 14 essais donne un odds ratio de 0,37. Le test de comparaison de ces deux estimations donne $p=0,11$.

Dans cet exemple, les deux approches ne conduisent pas aux mêmes estimations même si elles restent compatibles compte tenu de leurs degrés de précision (p d'hétérogénéité à 0,11).

Les auteurs ont comparé les caractéristiques des études sans trouver de véritable explication à cette discordance.

6.2.3 Etude de Song *et al.*, 2000

Antérieurement à leur travail de 2003, Song *et al.* (75) avaient conduit une première comparaison des résultats fournis par les comparaisons directes et indirectes à propos de l'antibioprophylaxie dans la chirurgie du colon. Une forte discordance avait été observée. Les auteurs imputent cette discordance au hasard en raison du faible nombre d'essais et de leur petitesse. En effet, seulement trois essais multibras sont impliqués dans ces calculs, fournissant à la fois les données nécessaires aux comparaisons indirectes et aux comparaisons directes. La non indépendance des différents bras au sein d'un même essai n'a pas été prise en compte.

Du fait de son caractère mono thématique (ne correspondant qu'à une seule question thérapeutique) et du faible nombre d'essais impliqués, ce travail ne peut pas constituer une étude empirique de la validité des comparaisons indirectes.

6.3 La comparaison directe reste-t-elle l'étalon-or ?

L'essai de comparaison directe appelé aussi essai « face à face » est considéré actuellement comme l'approche de référence pour comparer l'efficacité ou la sécurité de deux traitements actifs, même si en pratique ce type d'étude fait régulièrement défaut (cf. section 2.2). Cette approche apporte de nombreux avantages. Elle apporte la réponse à la question de l'efficacité relative des deux traitements avec toutes les protections méthodologiques de l'essai contrôlé, randomisé, double aveugle. Les patients chez lesquels le résultat est obtenu sont parfaitement définis et connus (il s'agit de la population de l'étude).

Cependant, des limites apparaissent en pratique.

6.3.1 Limites des essais de comparaison directe

La première limite est le faible recours à cette approche (cf. section 2.2). Dans bien des domaines les essais de comparaisons directes entre traitements actifs ne sont pas disponibles, notamment lors de l'évaluation post AMM d'un nouveau traitement, laissant sans réponse la question de la hiérarchie d'efficacité et/ou de sécurité des traitements concurrents concernés.

Souvent, ces essais sont réalisés en aval du dossier d'enregistrement. Les exigences méthodologiques et de qualité de ces essais peuvent être moins fortes que celles des essais destinés à l'enregistrement (76).

Les essais de comparaisons directes entre traitements actifs se révèlent souvent de plus faible niveau de preuve que les essais contrôlés *versus* placebo, en particulier au niveau de l'insu. Le masquage des traitements actifs, souvent déjà commercialisés, garantissant le double insu, posent des difficultés techniques, galéniques, analytiques. A titre d'exemple, l'encapsulation peut modifier la cinétique des médicaments et entraîner des modifications d'efficacité biaisant les résultats de l'essai « face à face » (77,78).

La technique du double placebo soulève des difficultés spécifiques avec des enjeux réglementaires et commerciaux. Lorsque les traitements comparés ne sont pas des médicaments, l'obtention d'un

double aveugle est encore plus difficile. Ces essais ne sont généralement pas non plus réalisés en double aveugle, à quelques exceptions près (*versus* « sham » par exemple).

Une autre limite pratique provient de la réalisation tardive de ces essais, après autorisation de mise sur le marché et commercialisation, laissant l'évaluation sans aucun élément de comparaison directe pour définir la place d'un nouveau traitement dans la hiérarchie des ressources thérapeutiques du domaine.

Enfin, se pose constamment la question de la sensibilité d'un essai (*assay sensitivity*) contre traitement actif. C'est à dire l'aptitude d'un essai à détecter d'éventuelles différences. Cette aptitude dépend de l'utilisation optimale des traitements (dose, schéma posologique), des patients inclus, de la puissance statistique, de la performance des critères de jugement, etc. Au niveau réglementaire, il est fortement recommandé (ICH E10, §1.5) d'introduire un groupe placebo dans les essais comparant deux traitements actifs pour démontrer cette aptitude de l'essai à mettre en évidence les différences quand elles existent. On peut citer comme exemple plusieurs essais comparant un nouveau triptan au triptan de référence dans le traitement de la crise aiguë de migraine (79).

Les essais comparant deux traitements actifs sont actuellement, de plus en plus fréquemment, des essais de non infériorité qui pose des problèmes tout à fait spécifiques. Dans ce type d'étude, la conclusion de non infériorité s'effectue au prix d'une perte d'efficacité potentielle. De ce fait, ce type d'essais ne permet pas vraiment de répondre à la question de la hiérarchie des traitements.

Tableau 11 - Avantages et limites des essais de comparaisons directes entre traitements actifs

Avantages	Limites
<ul style="list-style-type: none"> • Mesure directe de la question posée • Haut degré potentiel de protection contre les biais • Absence d'hypothèse sur la cohérence des effets de A et B 	<ul style="list-style-type: none"> • Rarement effectués • Rarement en double aveugle • Souvent réalisés avec des standards de qualité moindre que ceux des essais contre placebo • Réalisés tardivement dans le développement des nouveaux traitements • Validation interne rare

Tableau 12 - Avantages et limites des comparaisons directes et indirectes relatives à l'obtention et à la qualité des données analysées

	Avantages et limites	
	Comparaison directe	Comparaison indirecte
Délai d'obtention des données	<ul style="list-style-type: none"> • Tardif 	<ul style="list-style-type: none"> • Précoce (dès la disponibilité des premiers essais)
Qualité méthodologique des essais	<ul style="list-style-type: none"> • Fréquent défaut d'insu • Qualité variable mais rien n'exclut une qualité méthodologique maximale 	<ul style="list-style-type: none"> • Utilise des résultats obtenus en double insu • Utilise des résultats produits avec des exigences maximales de qualité
Exhaustivité des comparaisons à tous les traitements concurrents du domaine	<ul style="list-style-type: none"> • Rare • Limitée dans certains domaines • Impossibilité d'exploiter les données dans leur ensemble 	<ul style="list-style-type: none"> • Constante avec les techniques reposant sur un réseau de comparaisons
Biais de publication	<ul style="list-style-type: none"> • Peu probable en raison de la taille des effectifs nécessaires 	<ul style="list-style-type: none"> • Evitable par l'inclusion de tous les essais réalisés • Apport du registre des essais cliniques, susceptible d'améliorer la pertinence des comparaisons indirectes

6.3.2 Etude de cas

Dans la maladie d'Alzheimer, la comparaison donépézil et galantamine est documentée de façon directe par seulement deux essais : l'un publié par Jones *et al.* (80) et l'autre par Wilcock *et al.* (81). Le Tableau 13 décrit ces deux essais et leurs résultats. Chacun de ces deux essais a été financé par un fabricant de l'une de ces deux molécules. L'essai de Jones *et al.* comparant donépézil à la galatamine est sponsorisé par le fabricant du donepezil tandis que l'essai de Wilcock *et al.* comparant galatamine au donepezil est financé par le fabricant de la galantamine.

Ces deux essais donnent des résultats contradictoires, concluant à chaque fois en faveur de la molécule du sponsor de l'essai.

Cet exemple n'a évidemment aucune vertu de généralisation mais il met en évidence que l'essai « face à face » n'est peut être pas la solution au problème de la comparaison de deux traitements. Ces discordances de résultats peuvent provenir notamment du caractère ouvert de ces essais.

Dans un article récent, Song *et al.* proposent 3 études de cas supplémentaires où les comparaisons indirectes se sont révélées plus conservatrices que les comparaisons directes, posant à nouveau la question de la validité de ces études (82).

Tableau 13 - Description et résultats des 2 essais de comparaisons directes donepezil (DON) *versus* galatamine (GAL)

	Jones <i>et al.</i> , 2004 (80)	Wilcock <i>et al.</i> , 2003 (81)
Laboratoires	Fabricant du donepezil (Eisai Inc., Teaneck, NJ, Etats Unis, Pfizer Global Pharmaceuticals, Pfizer Inc., NY, Etats Unis)	Fabricant de la galantamine (Janssen-Cilag, Johnson & Johnson Pharmaceutical Research, Shire Pharmaceuticals)
Patients	Mild to moderate Alzheimer's disease	Alzheimer's disease
Comparaison	donepezil (jusqu'à 10mg/j) or galantamine (jusqu'à 24mg/j)	galantamine (24 mg/j) or donepezil (10 mg/j)
Durée	12 weeks	53 weeks
Effectif	120	182
Critères de jugement principal/secondaires	Physicians and caregivers satisfaction with treatment/ease of use in daily practice / ADAS-cog, MMSE, DAD-ADL	BrADL, MMSE, ADAS-cog/11, NPI, Screen for Caregiver Burden
Aveugle	Open	Open, rater-blinded
Conclusion de l'essai	Physician and caregiver ease of use/satisfaction scores, and assessments of cognition and ADL, showed significant benefits for donepezil compared with galantamine	Significant advantages were found in the treatment response to galantamine (<i>versus</i> donepezil) on cognition as measured by response rates on the MMSE and ADAS-cog/11

6.4 Perspectives

Une méta-analyse de synthèse des études empiriques étudiant la validité des comparaisons indirectes est planifiée dans le cadre de la *Cochrane Collaboration* par Song *et al.*¹². Pour l'instant, elle n'est qu'au stade de protocole. Seule une étude empirique de ce type est actuellement disponible.

¹² Song F, Altman DG, Glenny A, Eastwood AJ, Deeks JJ. Adjusted indirect comparison for estimating relative effects of competing healthcare interventions (Protocol). *Cochrane Database of Systematic Reviews* 2007, Issue 2.

7. Positionnement des instances réglementaires

European Medicines Agency : aucune recommandation portant sur les comparaisons indirectes n'a été trouvée sur le site Internet de l'EMA (recherche du 1^{er} octobre 2007).

Federal Drug Administration : aucune recommandation portant sur les comparaisons indirectes n'a été trouvée sur le site Internet de la FDA (recherche du 1^{er} octobre 2007).

National Institute for Health and Clinical Excellence : les documents de procédures du NICE font mention des comparaisons indirectes. Dans le cadre de l'élaboration des « single technology appraisal (STA) », un document destiné aux industriels « specification for manufacturer/sponsor submission of evidence » (83) fait mention des comparaisons indirectes dans les termes suivants : « The Institute has a strong preference for evidence from 'head-to-head' randomised controlled trials (RCTs) that directly compare the technology and the appropriate comparator(s). Wherever such evidence is available, and includes relevant outcome evidence, this is preferred over evidence obtained from other study designs. Where no head-to-head RCTs are available, consideration will be given to indirect comparisons, subject to careful and fully described analysis and interpretation. »

Le même texte figure dans le guide « Guide to the methods of technology appraisal (84).

Haute Autorité de Santé : aucune recommandation portant sur les comparaisons indirectes n'a été trouvée sur le site Internet de la HAS (recherche du 1^{er} octobre 2007).

Canadian Agency for Drugs and Technologies in Health : document de synthèse méthodologique publiée en mars 2009 (85).

Un article récent discute l'intérêt potentiel du recours aux techniques de comparaisons indirectes et de méta-analyse en réseau dans ce contexte (4).

8. Interprétation des résultats

L'interprétation des résultats de comparaisons indirectes, principalement ceux issus de réseaux de comparaisons présente deux types de pièges potentiels :

1. celui de porter des conclusions d'équivalence d'efficacité de deux traitements devant une absence de différence statistiquement significative ;
2. et celui d'établir, au sein d'une comparaison indirecte, une hiérarchie erronée par comparaisons naïves d'estimations ponctuelles.

Par exemple, à partir des résultats issus de la méta-analyse de l'effet des antihypertenseurs sur l'apparition de diabète (18), il n'est évidemment pas licite de conclure que les bêtabloqueurs sont identiques aux diurétiques. De même, la conclusion que les sartans (ARB) entraînent moins de nouveaux cas de diabète que les IEC (*ACE inhibitors*) ou que les antagonistes calciques (CCB) se base sur une comparaison naïve des estimations ponctuelles.

L'idéal pour éviter cette difficulté serait de prendre le traitement suspecté comme étant le mieux toléré comme référence et de montrer que les autres lui sont tous inférieurs de manière statistiquement significative.

9. Guide de lecture critique

Le guide de lecture critique proposé dans ce chapitre a pour objectif de vérifier si le travail de comparaisons indirectes analysé correspond à la situation optimale décrite dans la section suivante. Dans le cas contraire, les points non vérifiés représentent autant de réserves à émettre sur la validité des résultats disponibles. Il apparaît à la lumière d'une étude empirique récente que beaucoup de travaux de comparaisons indirectes présentent des problèmes méthodologiques importants, rendant impératif une lecture critique approfondie de ces travaux pour le moment (86).

9.1 Situation optimale

Pour être considéré comme acceptable, un résultat basé sur une comparaison indirecte devra vérifier les conditions suivantes :

- la comparaison indirecte a été effectuée avec une méthode statistique adaptée : comparaison indirecte ajustée, modèle linéaire mixte, méthodes Bayésiennes, méta-régression, etc. (cf. chapitre 1) ;
- une méta-analyse satisfaisante a été réalisée pour recueillir les essais contribuant aux comparaisons indirectes (recherche exhaustive des essais publiés et non publiés, sélection rigoureuse sur la qualité méthodologique des essais et une analyse poussée de l'hétérogénéité pour documenter l'hypothèse de stabilité des effets, cf. infra) ;
- les essais ayant participé aux comparaisons indirectes sont comparables en termes de traitements (dose, régime d'administration), de contextes cliniques (traitements concomitants y compris interventions autres que médicamenteuses), de pathologies (définition, critères diagnostiques, critères de gravité et évolution de la pathologie), de chronologie (même ère thérapeutique et même tendance séculaire). Les analyses en sous groupes des études sont rassurantes sur l'absence d'hétérogénéité de l'effet traitement en fonction de ces covariables¹³ ;
- si plusieurs chemins sont disponibles dans le réseau d'essais, les données permettent de tester l'incohérence de manière relativement satisfaisante et aucune incohérence n'a été détectée ;
- les résultats de comparaisons directes existants ne diffèrent pas du résultat de la comparaison mixte (« Mixed treatment comparison ») obtenu en combinant les comparaisons indirecte et directe qui sera mis en avant ;
- l'interprétation des résultats est correcte. En particulier, aucune conclusion d'équivalence n'est effectuée à partir d'un constat d'absence de différence.

9.2 Guide de lecture critique

Le guide de lecture critique d'un résultat obtenu par comparaison indirecte proposé ici comprend 6 grandes parties destinées à vérifier :

1. l'acceptabilité de l'approche utilisée ;
2. le processus de recherche et de sélection des données contribuant aux calculs de comparaison indirecte ;
3. l'homogénéité clinique des essais et la stabilité des effets ;
4. la cohérence des estimations ;
5. le degré de concordance du résultat avec celui des comparaisons directes existantes ;
6. l'interprétation correcte des résultats dans les conclusions proposées.

¹³ Absence de preuve que l'effet du traitement varie grandement en fonction de ces covariables.

9.2.1 Analyse de l'approche utilisée

Le premier point consiste à vérifier que la méthode employée est adaptée à la comparaison indirecte. Il s'agit d'exclure l'utilisation des méthodes inappropriées. Toutes les autres méthodes (cf. supra) sont acceptables.

i. Rejeter une approche de comparaison naïve des estimations ponctuelles issues d'essais contrôlés distincts

Il faut rejeter une approche où les estimations ponctuelles sont comparées les unes aux autres sans tenir compte de leurs intervalles de confiance (cf. section 4.1.1). Un exemple de résultats de ce type est donné dans l'étude de cas de la section 9.3.1.

ii. Rejeter une approche de comparaison naïve des bras actifs issus d'essais contrôlés distincts

Il faut rejeter une approche où les différents bras actifs des essais sont comparés, sans tenir compte des bras contrôles, rompant la randomisation (cf. section 4.1.2). Un exemple est donné dans l'étude de cas de la section 9.3.2.

9.2.2 Analyse de la recherche et de la sélection des essais inclus

Cette étape s'interroge sur le processus de constitution de la liste des essais participant aux méta-analyses nécessaires aux calculs de comparaison indirecte : s'agit-il d'essais sélectionnés pour donner les résultats de comparaisons indirectes souhaités ou s'agit-il de tous les essais à l'abri des biais réalisés dans le domaine ?

i. S'assurer que la recherche des essais a été exhaustive

Un processus de recherche exhaustive rigoureuse et reproductible des essais de type méta-analyse a-t-il été suivi ?

Les essais non publiés ont-ils été recherchés ? Les comparaisons indirectes sont sensibles, comme les méta-analyses classiques, au problème du biais de publication.

ii. Écarter une sélection opportune des essais

Il s'agit d'écarter la possibilité d'une sélection dirigée des essais en fonction du résultat escompté. Ici, l'analyse critique est identique à celle d'une méta-analyse. De fortes réserves devront être émises en cas d'absence de critères de sélection des essais ou d'utilisation de critères de sélection arbitraires.

A ce niveau, le travail de comparaison indirecte doit présenter les mêmes propriétés qu'une méta-analyse classique. L'analyse de la liste des essais exclus et des raisons d'exclusion est une aide utile pour l'analyse de la sélection des essais dont la qualité méthodologique garantit suffisamment l'absence de biais.

iii. S'assurer que toutes les comparaisons de base utilisables ont été considérées dans la sélection des essais

Un point en relation avec la recherche d'une incohérence (voir infra) est de s'assurer que toutes les comparaisons de base utilisables pour faire la comparaison indirecte recherchée ont bien été incluses. Pour la comparaison indirecte A vs B, les comparaisons de base utilisables peuvent être multiples : A vs PBO et B vs PBO mais aussi A vs D et B vs D, etc. Plus la diversité des comparaisons de base sera grande, plus le travail sera à même de tester avec puissance l'hypothèse d'incohérence.

9.2.3 Analyse de l'homogénéité clinique des essais et de la stabilité des effets

Cette étape vérifie que les essais recherchés et inclus cherchaient tous à répondre à la même question clinique au niveau du type de patients étudiés et des traitements comparés. Cette analyse vérifiera aussi les conditions d'acceptabilité de l'hypothèse de stabilité des effets au travers des essais fournissant les comparaisons de base.

iv. *Homogénéité des traitements étudiés dans les essais*

Les essais ont-ils étudié les mêmes modalités de traitements (en termes de dose, de régime d'administration) ? L'exemple de l'association de la codéine au paracétamol (cf. section 6.2.1) illustre parfaitement ce point.

v. *Absence de facteur d'interaction ou homogénéité des populations vis à vis des variables d'interaction*

Afin de vérifier la plausibilité de l'hypothèse de stabilité des effets indispensable à la validité des comparaisons indirectes (cf. section 4.2.2), il convient de vérifier que les analyses en sous groupes réalisées dans les essais ne révèlent pas de variable d'interaction, modifiant l'effet des traitements. Si de telles variables existent, la validité des comparaisons indirectes peut être remise en question s'il s'avère que les populations des essais de A *versus* contrôle et de B *versus* contrôle ne sont pas comparables vis à vis de ces covariables.

L'étude de cas de la section 9.3.7 donne un exemple de situation posant un problème d'interaction.

vi. *Absence d'hétérogénéité statistique entre les études d'une même comparaison*

La présence d'une hétérogénéité des résultats statistiques cliniquement significative entre les essais de A *versus* contrôle ou de B *versus* contrôle est susceptible de remettre en question l'hypothèse de stabilité des effets. En effet, si la méta-analyse trouve une hétérogénéité des résultats des essais effectuant la même comparaison, cela est évocateur d'une variabilité des effets d'un essai à l'autre, donc de l'existence de facteur(s) d'interaction.

En l'absence d'hétérogénéité, il n'a pas été possible de mettre en évidence de modification de l'effet d'un essai à l'autre. Si certains essais de A et de B se recoupent en termes de contextes et de types de patients, cela suggère que l'hypothèse de stabilité des effets peut être raisonnablement acceptée.

9.2.4 Analyse de l'hypothèse de cohérence

Le résultat d'une comparaison indirecte n'a de sens que si ce résultat est retrouvé à l'identique quel que soit l'enchaînement de comparaisons utilisé pour l'obtenir (cf. section 4.4), c'est à dire que si l'hypothèse de cohérence est vérifiée. Un résultat de comparaison indirecte sera d'autant plus fort que l'absence d'incohérence aura pu être éprouvée avec puissance. Des réserves seront à émettre chaque fois qu'une incohérence (cliniquement pertinente) aura été détectée ou qu'il n'aura pas été possible de la rechercher.

Ce point est à analyser pour chaque comparaison indirecte présente dans un travail car il concerne les résultats et non pas seulement la méthode.

vii. *L'incohérence a-t-elle été testée ?*

Était-il prévu de tester l'incohérence ? La méthode utilisée permet-elle de la tester ?

viii. *Le test de l'incohérence avait-il une efficacité minimale ?*

La possibilité de mettre en évidence un défaut de cohérence va dépendre du nombre d'essais effectuant les mêmes comparaisons et de la diversité des comparateurs pivots (chemins de comparaisons indirectes). S'il n'existe qu'un seul chemin (A vs PBO et B vs PBO) l'incohérence n'est pas testable ; seule est testable l'hétérogénéité des estimations des deux types de comparaisons.

9.2.5 Analyse de la méthode statistique mise en œuvre

Ce point est le plus technique du guide de lecture. Il a pour objectif de vérifier que la mise en œuvre de la méthode statistique utilisée est satisfaisante. L'existence de ce point dans le guide est justifiée par le fait que certaines des méthodes de comparaisons indirectes sont de mise en œuvre délicate, en particulier celle faisant appel aux méthodes d'estimation par méthode de Monte Carlo et chaînes de Markov (MCMC) et *Gibbs sampling*. En pratique le recours à une expertise statistique est souvent nécessaire pour juger de l'acceptabilité technique d'un résultat de ce type.

Pour les méthodes utilisant l'estimation par MCMC et Gibbs sampling :

Un diagnostic de convergence a-t-il été effectué ? Dans l'idéal, ce diagnostic fait appel à au moins une analyse graphique de l'historique de deux chaînes. L'utilisation d'outils de diagnostic comme CODA ou BOA est fortement souhaitable.

Les résultats obtenus sont-ils indépendants des « a priori » choisis ? Une étude de sensibilité des résultats démontrant que ces résultats sont indépendants du choix des « a priori » est indispensable.

Les « a priori » choisis sont-ils vraiment non informatifs ? En général, il est recommandé d'utiliser au moins pour les distributions normales une variance de 100 et pour les précisions une distribution gamma (0,0001, 0,0001). En particulier, il convient d'écartier les situations où l'obtention de la convergence nécessite l'utilisation d'un « a priori » informatif.

Un exemple d'utilisation satisfaisante de ces méthodes est rapporté dans l'étude de cas de la section 9.3.5.

9.2.6 Les résultats des comparaisons indirectes sont-ils renforcés par la disponibilité de comparaisons directes

ix. Y-a-t-il des comparaisons directes ?

Les essais de comparaisons directes ont-ils été recherchés ?

x. Y-a-t-il concordance des estimations des comparaisons directes et indirectes ?

Les estimations des comparaisons directes sont-elles compatibles avec celles des comparaisons indirectes ou mixtes ? Cette concordance est évaluée à l'aide d'un test d'hétérogénéité des 2 types d'estimations.

xi. Le résultat mis en avant inclut-il les comparaisons directes ?

En cas de disponibilité d'essais de comparaisons directes, le résultat mis en avant les intègre-t-il (« mixed treatment comparison ») ? En effet, dans ce cas, la meilleure estimation possible est celle qui englobe l'ensemble de l'information disponible.

9.2.7 Interprétation des résultats

xii. Aucune conclusion d'équivalence n'a été faite à partir de résultat non significatif

La conclusion d'absence de différence d'efficacité à partir de résultats de comparaison indirecte non significatifs est abusive. La problématique est strictement identique à celle rencontrée dans un essai thérapeutique non concluant.

xiii. Aucune hiérarchie n'est proposée à partir de comparaisons naïves des estimations ponctuelles

Il existe un risque de hiérarchiser les traitements concurrents par un processus de comparaisons naïves des estimations ponctuelles lors de comparaisons indirectes basées sur des réseaux (cf. section 8). Dans ce cas, la conclusion pose les problèmes évoqués dans la section 4.1.1, même s'il y a utilisation sous jacente d'une méthode appropriée de comparaison indirecte.

9.3 Études de cas

Ces études de cas ont pour objectif d'illustrer certains points du guide de lecture critique en explicitant les concepts par l'exemple.

9.3.1 Exemple de comparaison naïve des estimations ponctuelles

Un exemple de comparaison indirecte inappropriée par simple comparaison des estimations ponctuelles est donné par une méta-analyse s'intéressant à la prévention des thromboses veineuses profondes à la phase aiguë de l'accident vasculaire cérébral (AVC) (44).

La conclusion de l'article est « Indirect comparison of low and high doses of UFH and LMWH suggests that low-dose LMWH have the best benefit/risk ratio in patients with acute ischemic stroke by decreasing the risk of both DVT and pulmonary embolism, without a clear increase in ICH or ECH. »

Ce travail a comparé les faibles et fortes doses d'héparine (non fractionnée ou de bas poids moléculaires) dans la prévention des thromboses veineuses profondes à la phase initiale de l'AVC. La méta-analyse a été réalisée en sous groupes d'essais comme le montre le Tableau 14.

Tableau 14 - Analyse en sous groupes en fonction de la dose d'anticoagulant d'après Kamphuisen et Agnelli, 2007 (44)

	Traitement n/N	Contrôle n/N	OR [IC 95 %]
Fortes doses d'héparine	2/816	26/754	0,10 [0,03 ; 0,31]
Faibles doses d'héparine	60/482	175/432	0,21 [0,14 ; 0,30]

n : nombre d'événements ; N : nombre de patients

9.3.2 Exemple de comparaison naïve des bras actifs issus d'essais contrôlés

Une méta-analyse publiée en 2005 dans *Ophthalmology* (87) donne un exemple de comparaison indirecte inappropriée. L'objectif de cette méta-analyse était de comparer l'efficacité sur la pression intra oculaire (PIO) des différents traitements couramment utilisés dans le glaucome.

Les résultats présentent les moyennes des changements avant-après de pression intra oculaire observées dans les différents bras de ces essais, chaque traitement ayant été analysé sans tenir compte du bras contrôle. Ces résultats sont ensuite utilisés pour ordonner les traitements d'après ces valeurs de changement conduisant à la conclusion « This meta-analysis suggests that bimatoprost, travoprost, latanoprost, and timolol are the most effective intraocular pressure-reducing agents in POAG and OH patients » (87).

Cette approche est inappropriée car elle rompt la randomisation et revient à présenter des valeurs de nature observationnelle sans aucune prise en compte des facteurs de confusion à l'aide des groupes contrôles (cf. section 4.1.2). On s'attendrait à trouver, dans le tableau de résultats, des estimations d'efficacité par comparaison de traitements et non pas des estimations rattachées à un seul traitement (sans mention du comparateur).

9.3.3 Exemple de comparaison indirecte ajustée

La méta-analyse réalisée par Zhou *et al.* sur les statines en prévention cardiovasculaire a utilisé des comparaisons indirectes ajustées (47). Le Tableau 15 rapporte, pour chaque critère de jugement, les résultats des comparaisons de base (simvastatine vs pravastatine et atorvastatine vs pravastatine) et le résultat de la comparaison indirecte ajustée atorvastatine vs pravastatine.

Tableau 15 - Exemple de comparaisons indirectes ajustées dans la méta-analyse de Zhou *et al.*, 2006 (47)

	Risque relatif [IC 95 %]	P
Événements coronariens majeurs		
Simvastatine vs pravastatine	0,93 [0,84 ; 1,03]	0,18
Atorvastatine vs simvastatine	0,84 [0,66 ; 1,08]	0,18
Atorvastatine vs pravastatine	0,79 [0,61 ; 1,02]	0,06
Événements cérébrovasculaires majeurs		
Simvastatine vs pravastatine	0,87 [0,71 ; 1,07]	0,18
Atorvastatine vs simvastatine	0,90 [0,68 ; 1,20]	0,47
Atorvastatine vs pravastatine	0,78 [0,57 ; 1,07]	0,12

9.3.4 Exemple de comparaison mixte (*mixed treatment comparison*)

Un exemple d'utilisation des « mixed treatment comparisons » (MTC) est donné par la comparaison de trois chimiothérapies de seconde ligne dans le cancer ovarien : le topotécan, le paclitaxel et le DLP (liposomes pégylés contenant de la doxorubicine) (46).

Trois essais sont disponibles documentant les trois comparaisons 2 à 2 possibles avec ces 3 traitements :

	DLP	Topotécan	Paclitaxel
Essai A	X	X	
Essai B		X	X
Essai C	X		X

Les résultats de ces trois essais sur la mortalité totale montrent une incohérence étant donné que topotecan apparaît inférieur au paclitaxel tout en étant supérieur au DLP alors que paclitaxel apparaît inférieur au DLP :

		Survie globale
		Hazard ratio [IC 95 %]
Essai A	Topotécan vs paclitaxel	0,914 [0,681 ; 1,226]
Essai B	Topotécan vs DLP	1,216 [1 ; 1,478]
Essai C	Paclitaxel vs DLP	0,931 [0,702 ; 1,234]

La réalisation de MTC permet de réconcilier ces résultats en ne montrant plus aucune différence statistiquement significative entre ces trois traitements pouvant faire émettre l'hypothèse que l'incohérence initiale provenait du hasard :

		MTC - Survie globale
		Hazard ratio [intervalle de crédibilité 95 %]
	Topotécan vs paclitaxel	1,06 [0,85 ; 1,33]
	Topotécan vs DLP	1,14 [0,96 ; 1,36]
	Paclitaxel vs DLP	1,07 [0,86 ; 1,34]

9.3.5 Exemple d'utilisation satisfaisante d'une méthode d'estimation basée sur une méthode de Monte Carlo par chaînes de Markov (MCMC)

La méta-analyse de Babapulle *et al.* portant sur les stents enrobés donne un bon exemple d'utilisation satisfaisante d'une méthode de MCMC. Leur description de la méthode nous montre que les *a priori* utilisés étaient suffisamment non informatifs, et nous précise que le résultat est insensible aux « *a priori* » utilisés. Seul manque le diagnostic de convergence (54).

« Substantive prior knowledge can thereby be included into any Bayesian analysis by choice of initial (pre-data) distribution. However, because we incorporated all relevant past studies, we wanted our final (posterior) distribution to reflect the information in our dataset only and not to be influenced by our choice of initial (prior) distribution. Therefore, low-information prior distributions were used throughout, so that the data from the trials dominated the final inferences. In particular, we used normal (mean=0, variance=100) prior distributions for all means and gamma (0.0001, 0.0001) prior distributions for all precisions (which is defined as the reciprocal of the variance). Sensitivity analyses with different choices of low-information prior distributions showed robustness to this choice. In particular, a wide range of low-information values used for our gamma distributions did not change any of our posterior inferences. Therefore, our estimates of odds ratios and their associated 95% credible intervals (which are the Bayesian equivalent of standard confidence intervals) were not unduly affected by our choice of prior distribution. Inferences were calculated with the Gibbs sampler programmed in WinBUGS software (version 1.4, MRC Biostatistics Unit, Cambridge, UK) »

La méta-analyse sur les défibrillateurs implantables de Lam et Owen (88) donne aussi un exemple de méthode, aussi bien pour le choix des *a priori* :

« To ensure that overall effects were dominated by data from the trials and not influenced by choice of initial distribution we used low information (noninformative) prior distributions—that is,

we used vague normal (mean 0, variance 10,000) and uniform (0-2) prior distributions for means and standard deviations, respectively. We examined the impact of different choices of prior distribution in sensitivity analyses. »

que pour le diagnostic de convergence des chaînes :

« The Bayesian models were implemented using WinBUGS version 1.4.1 (Imperial College and Medical Research Council, 2004). After convergence was achieved from an initial 5,000 (burn-in) simulations, we constructed posterior distributions of the treatment effects from three chains of 50,000 simulations. MATLAB version 7.0 (MathWorks, Natick, MA, 2004) was used to carry out diagnostics and further data analyses. »

9.3.6 Étude de cas sur l'incohérence

La méta-analyse en réseau de l'effet des antihypertenseurs sur l'apparition de diabète illustre la gestion attendue de l'incohérence dans ce type de travail (18).

Le résumé précise « With an initial diuretic as the standard of comparison (eight groups), the degree of incoherence (a measure of how closely the entire network fits together) was small ($\omega=0.000017$, eight degrees of freedom). »

De plus, ce travail était à même de tester l'incohérence avec efficacité en raison de l'existence de nombreux chemins disponibles (y compris des chemins issus de comparaisons directes) pour chaque comparaison indirecte (

Figure 17).

Par exemple, l'estimation de l'efficacité des IEC par rapport aux diurétiques thiazidiques peut s'effectuer par au moins 4 chemins principaux :

- le chemin direct IEC vs diurétique (2 essais) ;
- le chemin IEC vs placebo (3 essais) et placebo vs diurétique (3 essais) ;
- le chemin IEC vs bêtabloquant et bêtabloquant vs diurétique ;
- le chemin IEC vs calcium bloqueur (CCB) (3 essais) et CCB vs diurétique (2 essais) ;
- et même les chemins IEC vs CCB, CCB vs ARB et ARB vs diurétique ou IEC vs CCB, CCB vs placebo et placebo vs diurétique, etc.

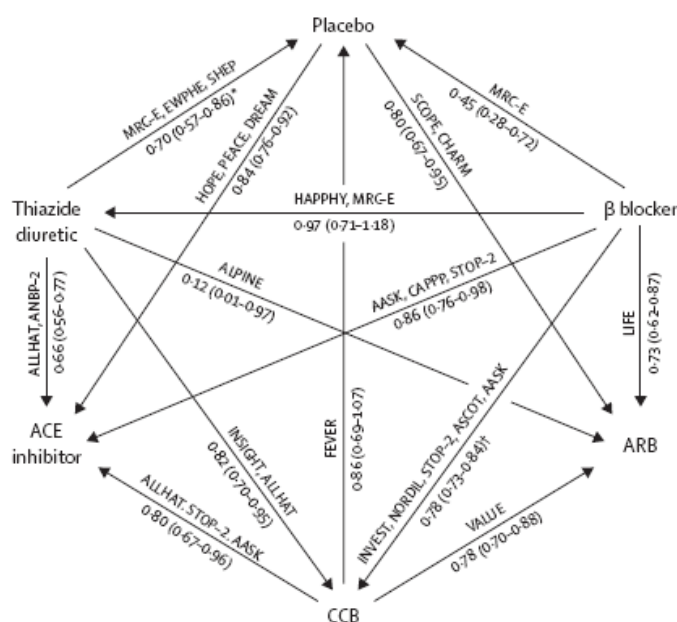


Figure 17 - Réseau de comparaisons de la méta-analyse de Elliott et Meyer, 2007

Reprinted from Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. Lancet 2007;369(9557):201-7. Copyright 2007, with permission from Elsevier

9.3.7 Interaction et hypothèse de stabilité des effets

Pour la fibrinolyse à la phase aiguë de l'infarctus du myocarde, le délai entre le début des symptômes et la fibrinolyse est une covariable qui interagit fortement avec la taille de la réduction de mortalité. Plus la fibrinolyse est tardive, moins la réduction de mortalité est importante. La méta-analyse des essais de fibrinolyse (89) *versus* placebo le démontre et permet de quantifier la relation entre délai et réduction de mortalité.

Si le fibrinolytique A a été évalué dans un essai *versus* placebo où le délai moyen depuis le début des symptômes était de 3h, son résultat ne pourra pas être combiné avec celui d'un essai de B *versus* placebo où le délai moyen était de 12h. Il n'est en effet plus possible d'accepter l'hypothèse de stabilité des effets dans cette configuration.

À effet des fibrinolytiques A et B identique et égal à l'effet moyen de tous les fibrinolytiques mesurés dans la méta-analyse, une réduction de mortalité d'environ 28 % sera observée dans l'essai de A, tandis qu'elle sera de 15 % dans l'essai de B, faisant apparaître une supériorité de A sur B lors de la comparaison indirecte.

10. Conclusion

De nombreuses méthodes permettent de réaliser des comparaisons indirectes de traitements. Ces méthodes permettent d'obtenir des données sur l'efficacité et la tolérance relative des différents traitements concurrents en l'absence totale ou partielle d'essais de comparaisons directes (« face à face ») de ces traitements.

Ces approches sont donc potentiellement de grand intérêt dans l'évaluation car elles peuvent aider à combler le manque de comparaisons directes fréquemment rencontré entre les traitements. Même lorsque les comparaisons directes sont disponibles, ces approches gardent leur intérêt en permettant de synthétiser les données disponibles d'efficacité et de tolérance relatives de l'ensemble des traitements concurrents d'un domaine.

L'analyse théorique de cette approche et des méthodes disponibles, associée au retour d'expérience en provenance des premières applications publiées, permettent d'esquisser les grandes lignes de l'évaluation critique de ce type de travaux et d'évaluer de manière standard leur fiabilité.

Ainsi, compte tenu des possibilités techniques disponibles, des résultats des études de validité et de leurs intérêts potentiels, les comparaisons indirectes peuvent maintenant être envisagées de façon assez systématique.

Le débat s'ouvre pour savoir dans quelle mesure elles peuvent contribuer à des évaluations de technologies de santé et des recommandations. Leur acceptabilité doit être envisagée au cas par cas en prenant en compte le domaine thérapeutique et les conséquences de l'absence d'essais de comparaison directe.

11. Points clés

Points clés	Voir chapitre
Pour pallier le manque d'essais comparatifs des traitements entre eux dans de très nombreux champs thérapeutiques, le recours aux comparaisons indirectes est envisageable quand se pose la question de la hiérarchie en termes d'efficacité et/ou de sécurité de traitements concurrents.	§ 1 § 2.1
Cette démarche formalisée peut permettre d'estimer, dès l'enregistrement, l'efficacité relative d'un nouveau traitement par rapport aux traitements existants sans devoir attendre, plusieurs années après sa commercialisation, les données de comparaisons plus conventionnelles issues d'essais « face à face ».	§ 2
Les résultats de comparaison directe ne doivent pas faire reléguer au second plan ceux des comparaisons indirectes ; ces deux estimations devront être soigneusement confrontées et si possible combinées à l'aide d'une approche mixte (<i>mixed treatment comparison</i>).	§ 4.3 § 6.3
La mise en œuvre de cette approche devrait s'effectuer préférentiellement par l'intermédiaire de grandes méta-analyses en réseau pour dessiner la hiérarchie complète d'un champ thérapeutique donné (envisageant s'il y a lieu d'autres ressources thérapeutiques que le médicament).	§ 4.4
Une seule grande étude sur la validité des comparaisons indirectes étant pour l'instant disponible, de nouvelles vérifications empiriques devront être entreprises en utilisant les méthodes modernes de comparaison indirecte. Des efforts de recherche devront être soutenus pour étudier des questions concernant la fiabilité des comparaisons directes et les performances des différentes méthodes de comparaison indirecte proposées.	§ 6.2
La diffusion des méthodes de comparaisons indirectes pourrait permettre de mieux définir les stratégies de hiérarchisation des traitements entre eux à partir de données factuelles.	§ 1

Annexe 1. Liste des participants

Ce rapport a été réalisé par le docteur Michel Cucherat (EA 3736, Faculté de médecine Laennec, Lyon) avec la participation du docteur Valérie Izard (HAS, Saint-Denis).

Le groupe de lecture était composé de cinq méthodologistes :

- Professeur Alain Bernard, Service de chirurgie thoracique, CHU Dijon
- Professeur Bruno Falissard, INSERM UMR-S669, Paris-Sud, Paris Descartes - Département de Santé Publique Evaluation et Information Médicale, Hôpital Paul Brousse, Villejuif
- Docteur Patrice Nony, Service de Pharmacologie Clinique - UMR CNRS 5558, Hôpital cardiologique, Lyon
- Professeur Eric Vicaut, Centre d'Evaluation du Dispositif Médical et Unité de Recherche Clinique Lariboisière St Louis, Paris
- Docteur Mahmoud Zureik, INSERM U700, Faculté de médecine Université Paris 7, Paris.

Annexe 2. Bibliographie

1. Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. *Value Health* 2008;11(5):956-64.
2. Gartlehner G, Moore CG. Direct versus indirect comparisons: a summary of the evidence. *Int J Technol Assess Health Care* 2008;24(2):170-7.
3. Malone DC. Using indirect comparisons in pharmaco-economic studies: time for implementation [editorial]. *Clin Ther* 2007;29(11):2454-5.
4. Sutton A, Ades AE, Cooper N, Abrams K. Use of indirect and mixed treatment comparisons for technology assessment. *Pharmacoeconomics* 2008;26(9):753-67.
5. Pérard L, Hot A, Cucherat M, Simon M, Desmurs H, Coppéré B, *et al.* Essais de non-infériorité dans la maladie thromboembolique veineuse. Une lecture critique est nécessaire! *Rev Med Interne* 2007;28(11):731-6.
6. Garattini S, Bertelé V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007;370(9602):1875-7.
7. Temple RJ. When are clinical trials of a given agent vs. placebo no longer appropriate or feasible? *Control Clin Trials* 1997;18(6):613-20.
8. Cour des comptes. La sécurité sociale. La dépense de médicaments 2007. <<http://www.ccomptes.fr/fr/CC/documents/RELFS/07-securite-sociale.pdf>> [consulté le 29-10-2008].
9. Brown BM, Hettmansperger TP. Kruskal-Wallis, multiple comparisons and Efron's paradox. *Aust N Z J Stat* 2002;44(4):427-38.
10. Thangavelu K, Brunner E. Wilcoxon-Mann-Whitney test for stratified samples and Efron's paradox dice. *J Stat Plan Inference* 2007;137:720-37.
11. Boos DD, Brownie C. Mixture models for continuous data in dose-response studies when some animals are unaffected by treatment. *Biometrics* 1991;47(4):1489-504.
12. Glenny AM, Altman DG, Song F, Sakarovich C, Deeks JJ, d'Amico R, *et al.* Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9(26).
13. Julious SA, Wang SJ. How biased are indirect comparisons, particularly when comparisons are made over time in controlled trials? *Drug Inf J* 2008;42(6):625-33.
14. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326(7387):472.
15. Soo S, Moayyedi P, Deeks J, Delaney B, Innes M, Forman D. Pharmacological interventions for non-ulcer dyspepsia (review). *Cochrane Database of Systematic Reviews* 2000; Issue 2.
16. Canadian Coordinating Office for Health Technology Assessment, Trindade E, Menon D. Selective serotonin reuptake inhibitors (SSRIs) for major depression. Part 1. Evaluation of the clinical literature. Ottawa: CCOHTA; 1997.
17. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;17(3):279-301.

-
18. Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *Lancet* 2007;369(9557):201-7.
19. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004;23(20):3105-24.
20. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331(7521):897-900.
21. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21(16):2313-24.
22. Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995;14(24):2685-99.
23. Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996;15(24):2733-49.
24. Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, *et al.* Health outcomes associated with various antihypertensive therapies used as first-line agents. A network meta-analysis. *JAMA* 2003;289(19):2534-44.
25. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med* 2004;23(11):1663-82.
26. Teramukai S, Matsuyama Y, Mizuno S, Sakamoto J. Individual patient-level and study-level meta-analysis for investigating modifiers of treatment effect. *Jpn J Clin Oncol* 2004;34(12):717-21.
27. Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21(11):1559-73.
28. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999;18(20):2693-708.
29. Van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21(4):589-624.
30. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *Lancet* 2005;366(9493):1267-78.
31. Eckert L, Lançon C. Duloxetine compared with fluoxetine and venlafaxine: use of meta-regression analysis for indirect comparisons. *BMC Psychiatry* 2006;6(30).
32. Nixon RM, Bansback N, Brennan A. Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med* 2007;26(6):1237-54.
33. Packer M, Antonopoulos GV, Berlin JA, Chittams J, Konstam MA, Udelson JE. Comparative effects of carvedilol and metoprolol on left ventricular ejection fraction in heart failure: results of a meta-analysis. *Am Heart J* 2001;141(6):899-907.
34. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21(11):1575-600.
35. Chou R, Carson S, Chan BKS. Gabapentin versus tricyclic antidepressants for diabetic neuropathy and post-herpetic neuralgia: discrepancies between direct and indirect meta-analyses of randomized controlled trials. *J Gen Intern Med* 2009;24(2):178-88.

-
36. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JPT, Churchill R, *et al.* Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;373(9665):746-58.
37. Edwards SJ, Clarke MJ, Wordsworth S, Welton NJ. Carbapenems versus other beta-lactams in the treatment of hospitalised patients with infection: a mixed treatment comparison. *Curr Med Res Opin* 2009;25(1):251-61.
38. Karthikeyan G, Eikelboom JW, Yusuf S, Hirsh J. An indirect comparison of the efficacy and safety of factor Xa inhibitors with thrombin inhibitors in preventing venous thromboembolism after hip or knee surgery [abstract]. *Circulation* 2008;118(Suppl 2):S1051.
39. Lee YH, Woo JH, Rho YH, Choi SJ, Ji JD, Song GG. Meta-analysis of the combination of TNF inhibitors plus MTX compared to MTX monotherapy, and the adjusted indirect comparison of TNF inhibitors in patients suffering from active rheumatoid arthritis. *Rheumatol Int* 2008;28(6):553-9.
40. Mauri D, Polyzos NP, Salanti G, Pavlidis N, Ioannidis JPA. Multiple-treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. *J Natl Cancer Inst* 2008;100(24):1780-91.
41. Biondi-Zoccai GGL, Lotrionte M, Abbate A, Valgimigli M, Testa L, Burzotta F, *et al.* Direct and indirect comparison meta-analysis demonstrates the superiority of sirolimus- versus paclitaxel-eluting stents across 5854 patients. *Int J Cardiol* 2007;114(1):104-5.
42. Boonen S, Lips P, Bouillon R, Bischoff-Ferrari HA, Vanderschueren D, Haentjens P. Need for additional calcium to reduce the risk of hip fracture with vitamin D supplementation: evidence from a comparative metaanalysis of randomized controlled trials. *J Clin Endocrinol Metab* 2007;92(4):1415-23.
43. Purkayastha S, Athanasiou T, Tekkis PP, Constantinides V, Teare J, Darzi AW. Magnetic resonance colonography vs computed tomography colonography for the diagnosis of colorectal cancer: an indirect comparison. *Colorectal Dis* 2007;9(2):100-11.
44. Kamphuisen PW, Agnelli G. What is the optimal pharmacological prophylaxis for the prevention of deep-vein thrombosis and pulmonary embolism in patients with acute ischemic stroke? *Thromb Res* 2007;119(3):265-74.
45. Lee YH, Ji JD, Song GG. Adjusted indirect comparison of celecoxib versus rofecoxib on cardiovascular risk. *Rheumatol Int* 2006;27:477-82.
46. Griffin S, Bojke L, Main C, Palmer S. Incorporating direct and indirect evidence using bayesian methods: an applied case study in ovarian cancer. *Value Health* 2006;9(2):123-31.
47. Zhou Z, Rahme E, Pilote L. Are statins created equal? Evidence from randomized trials of pravastatin, simvastatin, and atorvastatin for cardiovascular disease prevention. *Am Heart J* 2006;151(2):273-81.
48. Stettler C, Allemann S, Egger M, Windecker S, Meier B, Diem P. Efficacy of drug eluting stents in patients with and without diabetes mellitus: indirect comparison of controlled trials. *Heart* 2006;92(5):650-7.
49. Berner MM, Kriston L, Harms A. Efficacy of PDE-5-inhibitors for erectile dysfunction. A comparative meta-analysis of fixed-dose regimen randomized controlled trials administering the International Index of Erectile Function in broad-spectrum populations. *Int J Impot Res* 2006;18(3):229-35.
50. Batterham MJ. Investigating heterogeneity in studies of resting energy expenditure in persons with HIV/AIDS: a meta-analysis. *Am J Clin Nutr* 2005;81(3):702-13.

-
51. Biondi-Zoccai GGL, Agostoni P, Abbate A, Testa L, Burzotta F, Lotrionte M, *et al.* Adjusted indirect comparison of intracoronary drug-eluting stents: evidence from a metaanalysis of randomized bare-metal-stent-controlled trials. *Int J Cardiol* 2005;100(1):119-23.
52. Berry C, Norrie J, McMurray JJV. Ximelagatran compared with warfarin for the prevention of systemic embolism and stroke. An imputed placebo analysis. *Cardiovasc Drugs Ther* 2005;19(2):149-51.
53. Moore RA, Derry S, McQuay HJ. Indirect comparison of interventions using published randomised trials: systematic review of PDE-5 inhibitors for erectile dysfunction. *BMC Urol* 2005;5(18).
54. Babapulle MN, Joseph L, Bélisle P, Brophy JM, Eisenberg MJ. A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents. *Lancet* 2004;364(9434):583-91.
55. Yazdanpanah Y, Sissoko D, Egger M, Mouton Y, Zwahlen M, Chêne G. Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials. *BMJ* 2004;328(7434):249.
56. Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, *et al.* Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation. *Health Technol Assess* 2004;8(38).
57. Hind D, Calvert N, McWilliams R, Davidson A, Paisley S, Beverley C, *et al.* Ultrasonic locating devices for central venous cannulation: meta-analysis. *BMJ* 2003;327(7411):361.
58. Lim E, Ali Z, Ali A, Routledge T, Edmonds L, Altman DG, *et al.* Indirect comparison meta-analysis of aspirin therapy after coronary surgery. *BMJ* 2003;327(7427):1309.
59. Coomarasamy A, Knox EM, Gee H, Song F, Khan KS. Effectiveness of nifedipine *versus* atosiban for tocolysis in preterm labour: a meta-analysis with an indirect comparison of randomised trials. *BJOG* 2003;110(12):1045-9.
60. Bhandari M, Guyatt GH, Swiontkowski MF, Schemitsch EH. Treatment of open fractures of the shaft of the tibia. A systematic overview and meta-analysis. *J Bone Joint Surg [Br]* 2001;83B(1):62-8.
61. Agency for Healthcare Research and Quality. Manifestations and management of chronic insomnia in adults. Evidence report/Technology assessment n° 125. Rockville: AHRQ; 2005.
62. BlueCross BlueShield Association. Metal-on-metal total hip resurfacing. *Technol Eval Cent Asses Program Exec Summ* 2007;22(3):1-44.
63. BlueCross BlueShield Association. Off-label uses of bevacizumab: renal cell carcinoma and other miscellaneous non-colorectal cancer indications. *Technol Eval Cent Asses Program Exec Summ* 2006;21(9):1-26.
64. Costa J, Espírito-Santo C, Borges A, Ferreira JJ, Coelho M, Moore P, *et al.* Botulinum toxin type A therapy for cervical dystonia (Review). *Cochrane Database of Systematic Reviews* 2005;Issue 1.
65. Macfadyen CA, Acuin JM, Gamble C. Topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations (Review). *Cochrane Database of Systematic Reviews* 2005;Issue 4.
66. National Institute for Health and Clinical Excellence. Docetaxel for the adjuvant treatment of early node-positive breast cancer. NICE technology appraisal guidance 109. London: NICE; 2006.

-
67. National Institute for Health and Clinical Excellence. Pemetrexed for the treatment of non-small-cell lung cancer. NICE technology appraisal guidance 124. London: NICE; 2007.
68. National Institute for Health and Clinical Excellence. Adalimumab for the treatment of psoriatic arthritis. NICE technology appraisal guidance 125. London: NICE; 2007.
69. National Institute for Health and Clinical Excellence. Rituximab for the treatment of rheumatoid arthritis. NICE technology appraisal guidance 126. London: NICE; 2007.
70. Agency for Healthcare Research and Quality. Comparative effectiveness and safety of oral diabetes medications for adults with type 2 diabetes. Rockville: AHRQ; 2007.
71. National Institute for Health and Clinical Excellence. Infliximab for the treatment of adults with psoriasis. NICE technology appraisal guidance 134. London: NICE; 2008.
72. National Institute for Health and Clinical Excellence. Varenicline for smoking cessation. NICE technology appraisal guidance 123. London: NICE; 2007.
73. Zhang WY, Li Wan Po A. Analgesic efficacy of paracetamol and its combination with codeine and caffeine in surgical pain: a meta-analysis. *J Clin Pharm Ther* 1996;21(4):261-82.
74. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol* 1997;50(6):683-91.
75. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials* 2000;21(5):488-97.
76. Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326(7400):1167-70.
77. Salonen R. Drug comparisons: why are they so difficult? *Cephalalgia* 2000;20(Suppl 2):25-32.
78. Fuseau E, Petricoul O, Sabin A, Pereira A, O'Quinn S, Thein S, *et al*. Effect of encapsulation on absorption of sumatriptan tablets: data from healthy volunteers and patients during a migraine. *Clin Ther* 2001;23(2):242-51.
79. Pascual J, Mateos V, Roig C, Sanchez-del-Rio M, Jiménez D. Marketed oral triptans in the acute treatment of migraine: a systematic review on efficacy and tolerability. *Headache* 2007;47(8):1152-68.
80. Jones RW, Soininen H, Hager K, Aarsland D, Passmore P, Murthy A, *et al*. A multinational, randomised, 12-week study comparing the effects of donepezil and galantamine in patients with mild to moderate Alzheimer's disease. *Int J Geriatr Psychiatry* 2004;19(1):58-67.
81. Wilcock G, Howe I, Coles H, Lilienfeld S, Truyen L, Zhu Y, *et al*. A long-term comparison of galantamine and donepezil in the treatment of Alzheimer's disease. *Drugs Aging* 2003;20(10):777-89.
82. Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J Clin Epidemiol* 2008;61(5):455-63.
83. National Institute for Health and Clinical Excellence. Specification for manufacturer/sponsor submission of evidence. Single technology appraisal. London: NICE; 2006.
84. National Institute for Clinical Excellence. Guide to the methods of technology appraisal. London: NICE; 2004.

85. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. Indirect evidence: indirect treatment comparisons in meta-analysis. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.

86. Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009;338:b1147.

87. Van der Valk R, Webers CAB, Schouten JSAG, Zeegers MP, Hendrikse F, Prins MH. Intraocular pressure-lowering effects of all commonly used glaucoma drugs. A meta-analysis of randomized clinical trials. *Ophthalmology* 2005;112(7):1177-85.

88. Lam SKH, Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: Bayesian network meta-analysis of randomised controlled trials. *BMJ* 2007;335(7626):925.

89. Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. *Lancet* 1994;343(8893):311-22.

Annexe 3. Recherche de recommandations, rapports d'évaluation technologique, procédures méthodologiques sur Internet : liste des sites consultés

Agence française de sécurité sanitaire des produits de santé
Bibliothèque médicale Lemanissier
Bibliothèque Interuniversitaire de Médecine - BIUM
Catalogue et index des sites médicaux francophones - CISMeF
Comité d'Evaluation et de Diffusion des Innovations Technologiques - CEDIT
Expertise collective INSERM
Fédération Nationale des Centres de Lutte Contre le Cancer - FNCLCC
Société Française de Médecine Générale - SFMG
Adelaide Health Technology Assessment - AHTA
Agence d'Evaluation des Technologies et des Modes d'Intervention en Santé - AETMIS
Agency for Healthcare Research and Quality - AHRQ
Alberta Heritage Foundation for Medical Research - AHFMR
Australian Safety and Efficacy Register of New Interventional Procedures - Surgical
Blue Cross Blue Shield Association - BCBS - Technology Evaluation Center
Canadian Agency for Drugs and Technologies in Health - CADTH
Cancer Care Ontario
Centers for Disease Control and Prevention - CDC
Centre fédéral d'expertise des soins de santé - KCE
Centre for Clinical Effectiveness - CCE
Centre for Reviews and Dissemination databases
Clinical Knowledge Summaries
Canadian Medical Association Infobase
Cochrane Library
European Medicines Agency - EMEA
Food and Drug Administration - FDA
Guideline Advisory Committee - GAC
Guidelines and Protocols Advisory Committee- GPAC
Guidelines International Network - GIN
National Library for Health Guidelines Finder
Health Services Technology Assessment Text - HSTAT
Horizon Scanning
Institute for Clinical Evaluative Sciences - ICES
Institute for Clinical Systems Improvement - ICSI
Institute for Health Economics Alberta - IHE
Intute Health & Life Sciences - INTUTE
Medical Services Advisory Committee - MSAC
Minnesota Department of Health - Health Technology Advisory Committee - HTAC
National Comprehensive Cancer Network
National Coordinating Centre for Health Technology Assessment - NCCHTA
National Guideline Clearinghouse - NGC
National Health Services Scotland
National Institute for Health and Clinical Excellence - NICE
National Institutes of Health - NIH
New Zealand Health Technology Assessment - NZHTA
Ontario Health Technology Advisory Committee - OHTAC
Scottish Intercollegiate Guidelines Network - SIGN
Singapore Ministry of Health
U.S. Preventive Services Task Force - USPSTF
Veterans Affairs Technology Assessment Program
West Midlands Health Technology Assessment Collaboration - WMHTA

Index

A

a priori, 51
a priori non informatif, 28
adjusted indirect comparison, 18
ampleur d'effet, 35
approche mixte, 22
approche naïve, 16
assay sensitivity, 44

C

chaînes de Markov, 50
Comparaison des différentes méthodes, 35
comparaison directe, 9
comparaison face à face, 9
comparaison indirecte, 9
comparaison indirecte ajustée, 52
comparaison naïve, 51, 52
comparaisons directes, 51
concordance comparaison directe -
comparaison indirecte, 51
convergence, 50
Cour des Comptes, 12
critères de jugement continus, 35

D

diagnostic de convergence, 50
discordance, 41
discrepancy, 41
distribution a posteriori, 28
donepezil, 45
double aveugle, 43
double placebo, 44

E

effect size, 35
EMA, 46
essai de non infériorité, 12
essai versus traitement actif, 43
essais avec control actif, 9
essais de non infériorités, 13
estimation indirecte ajustée, 18

F

face à face, 43
FDA, 46
fibrinolyse, 54
fibrinolytique, 55

G

galantamine, 45
Gibbs sampling, 50

H

head to head comparison, 9
hypothèse de stabilité des effets, 19, 49, 50,
54

I

ICH E9, 13
infarctus du myocarde, 11, 54

L

lois de financement de la sécurité sociale, 12

M

maladie d'Alzheimer, 11, 45
MCMC, 50, 53
migraine, 11
modèle de Cox, 13
Monte Carlo, 50
Monte Carlo par chaînes de Markov, 53

N

NICE, 46
non transitivité des tests statistiques, 13

O

odds ratio, 36

R

risque relatif, 36
risque relatif., 9

S

sécurité sociale, 12
stabilité des effets, 19, 49, 50, 54

T

test de Wilcoxon, 13
test du logrank, 13
théorème de Bayes, 28
traitements concurrents, 11

V

validité, 41

W

weighted mean difference, 35



HAUTE AUTORITÉ DE SANTÉ

Toutes les publications de la HAS sont téléchargeables sur www.has-sante.fr