



HAL
open science

Place des comparaisons indirectes dans l'évaluation des médicaments en vue de leur accès au remboursement en France

Marine Guichard

► **To cite this version:**

Marine Guichard. Place des comparaisons indirectes dans l'évaluation des médicaments en vue de leur accès au remboursement en France. Sciences pharmaceutiques. 2015. dumas-01206476

HAL Id: dumas-01206476

<https://dumas.ccsd.cnrs.fr/dumas-01206476>

Submitted on 29 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Bordeaux
U.F.R. DES SCIENCES PHARMACEUTIQUES

Année 2015

Thèse n° 76

MEMOIRE DU DIPLOME D'ÉTUDES SPECIALISEES DE PHARMACIE
HOSPITALIERE ET DES COLLECTIVITES

Conformément aux dispositions de l'arrêté du 04 octobre 1988 tient lieu de

Thèse pour l'obtention du

DIPLOME D'ÉTAT DE DOCTEUR EN PHARMACIE

Présentée et soutenue publiquement

Le 17 juillet 2015 à Bordeaux par

Marine GUICHARD

Né le 13 juillet 1987 à Barbezieux (16)

<p>Place des comparaisons indirectes dans l'évaluation des médicaments en vue de leur accès au remboursement en France.</p>

Présidente

Professeur Marie-Claude SAUX

Pharmacien Praticien Hospitalier, Professeur émérite des Universités.

Directrice de thèse

Docteur Isabelle BORGET

Pharmacien, Maître de Conférences des Universités – Praticien Hospitalier à l'Université Paris Sud et Gustave Roussy, Economiste de la santé.

Assesseurs

Docteur Anne D'ANDON

Médecin, Directrice du Service Evaluation des Médicaments de la Haute Autorité de Santé.

Docteur Amélie DAVELUY

Pharmacien, Praticien Hospitalier, Centre d'addictovigilance de Bordeaux.

Docteur Driss BERDAI

Médecin des hôpitaux, Praticien Hospitalier, Service de pharmacologie médicale, CHU de Bordeaux.

Remerciements

Au professeur M-C. SAUX, Présidente du jury

Pharmacien Praticien Hospitalier, Professeur émérite des Universités.

Pour m'avoir fait l'honneur de présider ce jury.

J'ai appris grâce à vous la pharmacie clinique et la pharmacocinétique, et je continue à apprendre chaque jour à vos côtés durant ce dernier semestre d'internat à l'OMEDIT.

Veillez trouver ici le témoignage de ma reconnaissance et de mon profond respect.

A Madame le Docteur Isabelle BORGET, Directrice de thèse

Pharmacien, Maître de Conférences des Universités – Praticien Hospitalier à l'Université Paris Sud et Gustave Roussy, Economiste de la santé.

Pour avoir accepté de diriger ma thèse.

Votre accompagnement tout au long de ce travail, votre disponibilité, et vos conseils m'ont été d'une aide précieuse.

Veillez trouver ici le témoignage de toute mon estime et de mes plus sincères remerciements.

A Madame le Docteur Anne D'ANDON

Médecin, Directrice du Service Evaluation des Médicaments de la Haute Autorité de Santé.

Pour avoir accepté de faire partie de mon jury de thèse et pour m'avoir donné la chance d'évoluer durant ces 6 mois à vos côtés au sein du Service Evaluation des Médicaments.

Vos qualités professionnelles et humaines sont à l'origine de ma vive admiration et de ma très respectueuse considération.

A Monsieur le Docteur Driss BERDAI

Médecin des hôpitaux, Praticien Hospitalier, Service de pharmacologie médicale, CHU de Bordeaux.

Pour avoir accepté de participer à mon jury de thèse et d'évaluer ce travail.

Soyez assuré de tout mon respect et de ma profonde gratitude.

A Madame le Docteur Amélie DAVELUY

Pharmacien, Praticien Hospitalier, Centre d'addictovigilance de Bordeaux.

Pour avoir accepté de participer à mon jury de thèse et d'évaluer ce travail. Ce fut un réel plaisir que de travailler à tes côtés.

Merci également à **M.CUCHERAT, B.FALISSARD, L.PIBOULEAU** et **C.RUMEAU-PICHON**, pour votre relecture attentive et votre disponibilité.

A toutes les personnes rencontrées pendant ces quatre années d'internat (le service de SSR de l'hôpital Xavier Arnoz, la pharmacie de l'hôpital Pellegrin, le CRPV de Bordeaux, la pharmacie de l'hôpital de Pau, le département Affaires réglementaires du laboratoire MSD France, la direction des études et des services stratégiques des laboratoires SERVIER, la HAS et enfin l'OMEDIT Aquitaine), merci d'avoir fait de ma formation une très belle expérience professionnelle et personnelle.

Aux co-internes rencontrés tout au long de mon internat :

A **Clément M.**, pour ton aide et ton soutien durant la préparation du concours de l'internat.

A **Richouchou** pour m'avoir donné le privilège d'être ton petit padawan et de si bien commencer l'internat!!! Egalement merci à **Caroline B.** et à **Cathy**.

A mes supers co-internes du SRR (**Antoine, Sandra, Christelle, Bastien**) ; Mariminette ne vous oublie pas !

A toute la team Bordelaise, sans qui cet internat n'aurait pas été aussi festif et plaisant (**Lucie, Caro, Fanny, Anne, Pauline, Bubu, Laetitia, Lise, Lisa, Albane...**). J'espère que la fin de l'internat ne nous éloignera pas.

A **Flavie** pour les moments de fou-rire et aussi de grandes remises en question ! Je te souhaite de réussir dans ta nouvelle voie, tu le mérites.

A **Marine H.** ma colocatrice paloise; merci pour ce doux été passé à Pau en ta compagnie.

A **Cyrille, Benoît et Wandrille** ; cet été passé au sein de la « Maison » fut mémorable.

A **Mathilde Grande** pour tous tes précieux conseils, ton soutien et tes corrections sur ma thèse.

A ceux qui m'ont si bien encadré :

Au **Dr Joanne Jenn**, merci pour ta gentillesse, ton dynamisme et ton implication dans la pharmacie clinique.

A **Hélène G.**, tu m'as fait découvrir la pharmacovigilance avec toute ton énergie et ton enthousiasme.

A toute la super équipe de la **pharmacie de Pau**, avec une attention spéciale pour ma super coloc **Brigitte** ; merci pour votre accueil et votre esprit de famille.

*A **Jean T., Corinne et Laetitia** ; merci de m'avoir initié au monde de l'industrie.*

*A **Marie et Débo**, pour ce super stage. J'espère que nous resterons en contact.*

*A toute **l'équipe du SEM de la HAS** ; vous m'avez beaucoup apporté.*

*A **Bertrice et à toute son équipe** dynamique de l'OMEDIT avec laquelle j'ai beaucoup de plaisir à apprendre et travailler chaque jour.*

A ma famille :

*A **mes parents**, pour votre soutien sans faille (même dans les moments de révisions où j'étais quelque peu insupportable !!!). Merci de m'avoir donné les moyens d'y arriver. Vous êtes des parents géniaux, je vous suis infiniment reconnaissante de tout ce que vous faites pour moi.*

*A **ma sœur**, qui sait à sa manière être présente quand il le faut (notamment pour courir me racheter des fluos, ou encore me refaire une fiche sur laquelle je viens de renverser un verre d'eau!...).*

*A mon super beau-frère **Laurent** ; promis, c'est fini les journées derrière mon bureau !*

*A mes deux grands-mères, **Gisèle et Monique**, pour toute la tendresse et le soutien que vous m'avez toujours apportés.*

*A **mon papi**, qui nous a quitté cette année. Toi qui t'inquiétais que je travaille si dur et que ces études durent tant, j'aurais aimé que tu sois là au moment où j'en vois la fin.*

*A **Florian**, pour ton soutien, ta patience, et ta bonne humeur avec laquelle je me réveille chaque matin.*

Liste des abréviations

ACB	Analyse coût bénéfice
ACE	Analyse coût-efficacité
ACR	Analyse coût-résultat
ACU	Analyse coût-utilité
AMM	Autorisation de mise sur le marché
ANSM	Agence nationale de sécurité du médicament
ASMR	Amélioration du service médical rendu
ATU	Autorisation temporaire d'utilisation
BPCO	Bronchopneumopathie chronique obstructive
CEESP	Commission d'évaluation économique et de santé publique
CEPS	Comité économique des produits de santé
CHMP	<i>Committee for medical products of human use</i>
CT	Commission de la transparence
ECR	Essai contrôlé randomisé
EMA	Agence européenne du médicament
EUnetHTA	<i>European network for health technology assessment</i>
HAS	Haute autorité de santé
HCSP	Haut conseil de la santé publique
HTAP	Hypertension artérielle pulmonaire
HTP-TEC	Hypertension pulmonaire thromboembolique chronique
IC	Intervalle de confiance
ICER	<i>Incremental cost effectiveness ratio</i>
IEC	Inhibiteur de l'enzyme de conversion
ISPOR	<i>International Society for pharmacoeconomics and outcomes research</i>
LEEM	Les entreprises du médicament
LFSS	Loi de financement de la Sécurité sociale
MTC	<i>Mixed treatment comparison</i> ou comparaison mixte
MTM	<i>Multiple treatments meta-analysis</i> ou méta-analyse en réseau
MVO	Maladie veino-occlusive
OMS	Organisation mondiale de la santé
ONDAM	Objectif national des dépenses d'assurance maladie
OR	<i>Odd ratio</i>

PIB	Produit intérieur brut
QALY	<i>Quality-adjusted life years</i>
RDCR	Ratio différentiel coût résultat
RR	Risque relatif
SE	<i>Standard error</i> ou erreur standard
SEESP	Service évaluation économique et santé publique
SEM	Service évaluation des médicaments
SEP-RR	Sclérose en plaques rémittente
SMR	Service médical rendu
TFO	Traitement de fond optimisé
UNCAM	Union nationale des caisses d'assurance maladie
VIH	Virus de l'immunodéficience humaine
CRI	Intervalle de crédibilité

Sommaire

Remerciements	2
Liste des abréviations	5
Sommaire	7
Liste des tableaux	11
Liste des figures	12
Liste des annexes	15
Introduction	16
PARTIE 1	18
Les comparaisons indirectes : méthodes et validité	18
1 Méthodes statistiques pour les comparaisons indirectes	19
1.1 Les méthodes inappropriées : les comparaisons indirectes naïves	20
1.1.1 Comparaisons naïves des estimations ponctuelles	20
1.1.2 Comparaison naïve des bras actifs	22
1.1.3 Approche par un test d'hétérogénéité	24
1.2 Les méthodes appropriées	24
1.2.1 La méthode de Bucher de comparaison indirectes « ajustées » ou « extrapolées » ..	24
1.2.2 Les méta-analyses en réseau (<i>Network meta-analysis</i> ou <i>Multiple-Treatments Meta-</i>	
Analysis MTM)	33
1.2.3 La méta-régression	44
2 Validité des données des comparaisons indirectes	47
2.1 Les différentes hypothèses de validité	47
2.1.1 Hypothèse d'homogénéité (<i>homogeneity</i>) pour les données de la méta-analyse	47
2.1.2 Hypothèse de similarité (<i>similarity</i>) pour les comparaisons indirectes (ou	
« échangeabilité », « transposabilité »)	51
2.1.3 Hypothèse de cohérence (<i>consistency</i>)	55
2.2 Le biais de publication	62
3 Guide de lecture critique	63
3.1 Pertinence globale pour le décideur	64
3.1.1 Le choix de la population est-il pertinent ?	64

3.1.2	Des traitements pertinents sont-ils manquants?	64
3.1.3	Les effets de traitements pris en compte sont-ils pertinents ?.....	65
3.1.4	Le contexte (les paramètres et les conditions) est-il applicable?.....	65
3.2	Analyse de la recherche et de la sélection des essais inclus	65
3.2.1	S'assurer de l'exhaustivité de la recherche.....	66
3.2.2	S'assurer de la qualité des études incluses	66
3.3	Vérification des hypothèses d'homogénéité et de similarité.....	67
3.3.1	Absence d'hétérogénéité statistique entre les études d'une même comparaison.....	67
3.3.2	Vérifier la présence de modificateurs d'effets de traitement pour l'hypothèse de similarité.....	68
3.4	Analyse de la méthode statistique mise en œuvre.....	68
3.4.1	Eviter les comparaisons indirectes naïves.....	68
3.4.2	Analyse de la méthode statistique mise en œuvre.....	69
3.4.3	Vérifier l'hypothèse de cohérence	69
3.4.4	Présentations des résultats	70
3.5	Interprétation des résultats.....	70
PARTIE 2	73
	Place des comparaisons indirectes dans l'évaluation médico-technique des médicaments en vue de leur accès au remboursement.....	73
1	La procédure d'accès au remboursement du médicament en France^{75,76}.....	74
1.1	L'Autorisation de Mise sur le Marché	75
1.2	L'inscription sur la liste des spécialités remboursables aux assurés sociaux et des spécialités agréées à l'usage des collectivités et divers services publics	75
1.2.1	La Commission de la transparence (CT) : missions et composition.....	76
1.2.2	Le Service médical rendu (SMR) et l'Amélioration du Service médical rendu (ASMR) 77	
1.2.3	La Primo-inscription	81
1.2.4	Le renouvellement d'inscription	82
1.2.5	Le recours aux comparaisons indirectes et les recommandations de la HAS	82
2	Place des comparaisons indirectes dans l'évaluation des médicaments par la Commission de la transparence : étude rétrospective 2009-2014.....	85
2.1	Objectif.....	85
2.2	Méthodes.....	85
2.3	Résultats.....	86

2.3.1	Nombre d'avis relevés	86
2.3.2	Typologie des avis contenant des comparaisons indirectes.....	88
2.3.3	Validité des comparaisons indirectes et critiques méthodologiques	94
2.3.4	Retentissement de la comparaison indirecte sur les conclusions de l'avis	102
2.4	Résumé et discussion	108
3	Conclusion	110
PARTIE 3		111
Place des comparaisons indirectes dans l'évaluation médico-économique des médicaments.....		111
1	L'évaluation médico-économique du médicament en France	112
1.1	Principes et objectifs de l'évaluation médico-économique.....	112
1.2	Le contexte de l'évaluation médico-économique en France	114
1.2.1	Le décret du 2 octobre 2012	114
1.2.2	La Commission évaluation économique et de santé publique (CEESP) de la HAS et les avis d'efficience.....	115
1.2.3	La place de l'évaluation médico-économique dans l'admission au remboursement	118
1.3	Les points clés d'une évaluation médico-économique	120
1.3.1	Le choix du type d'étude	121
1.3.2	Le choix de la population d'analyse	122
1.3.3	Le choix du comparateur et le recours aux comparaisons indirectes	122
1.3.4	L'horizon temporel.....	123
1.3.5	La modélisation	123
1.3.6	L'expression des résultats.....	124
1.3.7	Les analyses de sensibilités	126
2	Place des comparaisons indirectes dans l'évaluation médico-économique des médicaments : étude rétrospective.....	130
1.1	Objectifs.....	130
1.2	Méthodes.....	130
1.3	Résultats.....	131
1.3.1	Nombre d'avis relevés	131
1.3.2	Nombre d'avis concernés par le recours à une comparaison indirecte	131
1.3.3	Analyse des avis ne contenant pas de comparaison indirecte	132
1.3.4	Analyse des avis contenant des comparaisons indirectes	142

1.4	Résumé et discussion	156
3	Conclusion	158
	Conclusion.....	160
	Annexes.....	163
	Bibliographie	166

Liste des tableaux

<i>Tableau 1 : Résultats des estimations ponctuelles de l'essai A vs placebo et de l'essai B vs placebo⁴.</i>	21
<i>Tableau 2 : Résultats des estimations ponctuelles et intervalles de confiance de l'essai A vs placebo et de l'essai B vs placebo⁴.</i>	21
<i>Tableau 3 : Résultat de l'essai comparant A au placebo⁴.</i>	23
<i>Tableau 4 : Résultat de l'essai comparant B au placebo⁴.</i>	23
<i>Tableau 5 : Méta-analyse de rispéridone vs halopéridol dans le traitement de la schizophrénie¹¹.</i>	29
<i>Tableau 6 : Exemple de league table (Hazard ratios et intervalles de crédibilité à 95% pour la survie globale).</i>	43
<i>Tableau 7 : Méta-analyse de rispéridone vs halopéridol dans le traitement de la schizophrénie, test d'homogénéité¹¹.</i>	51
<i>Tableau 8 : Méta-analyse de rispéridone vs halopéridol dans le traitement de la schizophrénie, évaluation de l'incohérence¹¹.</i>	59
<i>Tableau 9 : Concordance entre la conclusion de la comparaison indirecte (CI) avec l'ASMR.</i>	104
<i>Tableau 10 : Résumé de l'analyse des avis d'efficience sans comparaison indirecte.</i>	134
<i>Tableau 11 : Résumé de l'analyse des avis d'efficience avec des comparaisons indirectes.</i>	147
<i>Tableau 12 : Résumé des critiques (en rouge) et points positifs (en vert) émis à l'encontre des méta-analyses en réseau présentées dans les avis d'efficience LEMTRADA et TECFIDERA.</i>	153

Liste des figures

Figure 1 : Principe de la comparaison indirecte.....	19
Figure 2 : Comparaison indirecte naïve non valide car les différences d'effet par rapport au placebo dans chaque étude ne sont pas reconnues ⁶	22
Figure 3 : Comparaison indirecte ajustée B vs A. La comparaison indirecte ajustée tient compte du risque de base r_0 ⁶	27
Figure 4 : Utilisation de la méta-analyse pour le calcul des comparaisons indirectes ajustées ⁴	28
Figure 5 : Deux exemples d'intégration par méta-analyse des résultats des comparaisons directes avec ceux issus des comparaisons indirectes ajustées ⁸	31
Figure 6 : Exemple de géométries de réseaux ¹⁸	35
Figure 7 : Réseaux des essais d'antihypertenseurs dans lesquels l'incidence du diabète est rapportée, d'après Elliott et Meyer, 2007 ¹⁹	36
Figure 8 : Approche Fréquentiste et Bayésienne d'une méta-analyse en réseau ⁶	37
Figure 9 : Méthode d'estimation Bayésienne.....	39
Figure 10 : Forest plot des résultats de la méta-analyse en réseau concernant l'efficacité et le taux d'abandon (tolérance) des antipsychotiques par rapport au placebo ²⁹	43
Figure 11 : "Rankograms" montrant la probabilité pour chaque traitement d'être à un rang déterminé dans la hiérarchie de traitements. (rank=rang) ¹⁸	44
Figure 12 : Comparaison indirecte avec et sans ajustement avec le modificateur d'effet en utilisant un modèle de méta-régression ⁶	46
Figure 13 : Hypothèses sous-jacentes aux comparaisons indirectes et comparaisons mixtes ⁴²	47
Figure 14 : Comparaison indirecte d'une étude A vs B et A vs C avec des proportions différentes de patients atteints de maladie modérée et sévère ³	53
Figure 15 : Comparaison directe et indirecte des traitements A et B ¹¹	55
Figure 16 : Exemple d'un réseau en boucle fermée qui présente une cohérence entre la comparaison directe et indirecte ⁶	56
Figure 17 : Exemple d'un réseau en boucle fermée qui présente une incohérence entre les comparaisons directes et indirectes ⁶	57
Figure 18 : Résultats des différentes méthodes de comparaisons de l'halopéridol et de la rispéridone ¹¹	59
Figure 19 : Illustration du processus de comparaisons indirectes ajustées en cas d'existence de plusieurs chemins de comparaison de base ⁴	61
Figure 20 : Exemple d'hétérogénéité au sein des méta-analyses qui n'entraîne pas de remise en cause de l'hypothèse de similarité ⁶	67
Figure 21 : Exemple d'hétérogénéité au sein des méta-analyses qui entraîne une remise en cause de l'hypothèse de similarité ⁶	68
Figure 22 : Le circuit du remboursement des médicaments en France.....	74
Figure 23 : Evolution du nombre d'avis contenant une ou plusieurs comparaisons indirectes sur la période de juillet 2009 à octobre 2014.....	87

Figure 24: Méthodologie des comparaisons indirectes dans les avis de juillet 2009 à octobre 2014 (n=61 comparaisons indirectes).	89
Figure 25: Méthodologie des comparaisons indirectes par année de juillet 2009 à octobre 2014.	90
Figure 26 : Aires thérapeutiques (classes ATC) concernées par le recours aux comparaisons indirectes dans les avis (n=60), période juillet 2009-octobre 2014.....	91
Figure 27 : Répartition des motifs d'examen dans les avis comportant une ou plusieurs comparaisons indirectes, juillet 2009-octobre 2014.....	93
Figure 28 : Répartition entre les différents types de demandes déposées en 2013 auprès de la CT.	93
Figure 29: Recevabilité des comparaisons indirectes (n=61) dans les avis de juillet 2009 à octobre 2014.	94
Figure 30 : Répartition des principales critiques formulées à l'égard des comparaisons indirectes présentées entre juillet 2009 et octobre 2014.....	95
Figure 31 : Exploitation des résultats des comparaisons indirectes (n=61) dans les avis de juillet 2009 à octobre 2014.....	102
Figure 32 : Réseau de comparaisons de la MTC pour le taux annualisé de poussées dans l'avis TECFIDERA du 07 mai 14.....	107
Figure 33 : Principe de l'évaluation médico-économique.....	112
Figure 34 : Le ratio différentiel coût-résultat (RDCR) dans l'évaluation médico-économique.....	113
Figure 35: Procédure de remboursement et de fixation du prix pour les médicaments ⁹⁰	119
Figure 36 : Les différents types d'évaluation médico-économiques.....	121
Figure 37 : Exemple d'un modèle de Markov.	124
Figure 38 : Représentation graphique de la frontière d'efficacité dans un repère coût-résultat.....	125
Figure 39: Présentation d'une analyse de sensibilité univariée sous la forme d'un graphique de Tornado ⁹⁵ .	127
Figure 40: Présentation d'une analyse de sensibilité probabiliste sous la forme d'un nuage de points et d'ellipse de confiance ⁹⁶	128
Figure 41 : Présentation de l'incertitude par une courbe d'acceptabilité ⁹⁷	129
Figure 42 : Utilisation des comparaisons indirectes dans les avis d'efficacité (n=14) d'octobre 2013 à décembre 2014.....	131
Figure 43 : Raisons évoquées pour ne pas utiliser de comparaisons directes (n=14) dans les avis d'efficacité d'octobre 2013 à décembre 2014.	133
Figure 44 : Problème posé dans le choix des comparateurs dans l'avis d'efficacité XOFIGO (avis du 11/03/2014).....	137
Figure 45 : Stratégies comparées considérées ou non dans l'avis d'efficacité TIVICAY (27/05/2014).	139
Figure 46 : Stratégies comparées considérées ou non dans l'avis d'efficacité VEXTIBIX (27/05/2014) et réserves occasionnées.	142
Figure 47 : Méthodologie des comparaisons indirectes dans les avis d'efficacité (n=14) d'octobre 2013 à décembre 2014.....	143
Figure 48 : Indications thérapeutiques concernées par le recours aux comparaisons indirectes dans les avis d'efficacité (n=5) (octobre 2013 – décembre 2014).....	144
Figure 49 : Nombre d'avis d'efficacité concernés ou non par des réserves sur les comparaisons indirectes présentées (n=4) d'octobre 2013 à novembre 2014.....	145

<i>Figure 50 : Principe de la « fausse » comparaison indirecte décrite dans l'avis d'efficience OPSUMIT du 24/06/2014.....</i>	<i>148</i>
<i>Figure 51 : Comparaison indirecte effectuée avec la méthode de Bucher dans l'avis d'efficience ADEMPAS du 14/10/14.....</i>	<i>149</i>
<i>Figure 52 : Analyse du dossier d'efficience KADCYLA (avis d'efficience du 11/03/2014).</i>	<i>151</i>

Liste des annexes

<i>Annexe 1: Rankogram en valeurs absolues (Figure A) et en valeurs cumulatives (Figure B) pour présenter les probabilités et les classements d'une méta-analyse en réseau</i>	<i>163</i>
<i>Annexe 2 : Guide de lecture critique utilisé pour le relevé des données des études rétrospectives.....</i>	<i>164</i>
<i>Annexe 3 : Résumé de l'analyse de référence HAS⁹²</i>	<i>165</i>

Introduction

Si, selon l'adage, « *la santé n'a pas de prix* », elle a en revanche un coût lorsqu'elle devient un objectif collectif. En 2012, la France est le troisième pays où les dépenses de santé rapportées au PIB du pays sont les plus élevées toutes sources de financement confondues, derrière les Pays-Bas et les Etats-Unis, avec 11,6 % du produit intérieur brut¹ dédié à la santé. En France, la plus grande partie des dépenses est financée par la branche maladie de la Sécurité Sociale. Ainsi, en 2013, les dépenses dans le champ de l'Objectif national des dépenses d'assurance maladie (ONDAM) ont été de 173,8 Md€, en progression de 2,2 % par rapport aux objectifs arrêtés pour 2012². Le coût de la santé ne cesse ainsi d'augmenter avec la transition épidémiologique³ (les maladies aiguës sont stables, alors que les maladies chroniques sont en développement), le vieillissement, et l'évolution des techniques médicales. La maîtrise des dépenses de santé est ainsi devenue une question cruciale pour les finances publiques.

En France, la Haute Autorité de Santé (HAS) joue un rôle prédominant dans l'admission des médicaments au remboursement. Celui-ci est décidé sur la base des avis de la Commission de la transparence (CT) et, plus récemment, la Commission d'Evaluation Economique et de Santé Publique (CEESP) intervient également dans le processus de fixation du prix. La CT doit pouvoir comparer les médicaments ayant obtenu l'Autorisation de Mise sur le Marché (AMM) pour pouvoir se prononcer pour chacun sur un niveau de Service Médical Rendu (SMR) et d'Amélioration du Service Médical Rendu (ASMR), éléments déterminants pour l'admission du médicament au remboursement et pour la fixation de son prix. La CEESP, de son côté, doit également disposer de données comparatives pour pouvoir évaluer l'efficacité du médicament par rapport à ses comparateurs pertinents. Or, la plupart des essais contrôlés randomisés et des méta-analyses actuellement publiés ne permettent pas de répondre à la question pourtant déterminante du choix de traitement le plus efficace parmi l'arsenal thérapeutique disponible dans une situation donnée. En effet, les essais contrôlés randomisés sont parfois conduits contre placebo malgré l'existence de traitements actifs et les méta-analyses « classiques » synthétisent les résultats de ces essais. Il arrive aussi que les essais contrôlés randomisés soient développés *versus* un comparateur actif, alors que la stratégie thérapeutique en compte plusieurs.

Dans ce contexte, des méthodes se sont développées qui permettent de comparer l'efficacité (et/ou la tolérance) de deux traitements par l'intermédiaire de l'efficacité (et/ou la tolérance) respective de ces deux traitements par rapport à un comparateur commun, le plus souvent un placebo. On appelle ces méthodes les « *comparaisons indirectes* ». Pour aller encore plus loin, des méthodes dites de « *méta-analyses en réseau* » combinent comparaisons directes et indirectes afin d'estimer toutes les comparaisons possibles entre les différents traitements évalués dans le réseau.

L'objectif de ce travail de thèse est de déterminer la place et l'impact des méthodes de comparaisons indirectes dans l'évaluation des médicaments par la HAS en vue de leur accès au remboursement en France en 2014. Nous tenterons de répondre aux questions suivantes : des comparaisons indirectes peuvent-elles compenser l'absence ou compléter la rareté des comparaisons directes lors de l'évaluation médico-technique du médicament par la CT, dans un second temps, lors de son évaluation médico-économique par la CEESP? Quels principes fondamentaux doivent-elles suivre pour être jugées valides et retenues ?

Notre travail sera organisé en trois parties. Nous réaliserons tout d'abord un état des lieux des méthodes de comparaisons indirectes disponibles, ainsi que les conditions de validité dont elles dépendent. Dans une deuxième partie, nous présenterons le processus d'accès au remboursement, puis les résultats d'une étude rétrospective analysant les comparaisons indirectes relevées dans les avis de la CT entre 2009 et 2014, ainsi que leur poids dans l'évaluation et la note rendue (ASMR). Enfin, la dernière partie de cette thèse sera consacrée à l'évaluation médico-économique des produits de santé innovants et coûteux, et aux résultats d'une étude rétrospective visant à évaluer l'importance des comparaisons indirectes dans les avis d'efficacité de la CEESP en 2013 et 2014.

PARTIE 1

Les comparaisons indirectes : méthodes et validité.

1 Méthodes statistiques pour les comparaisons indirectes

Comparer de manière indirecte deux traitements signifie évaluer les performances cliniques de l'un par rapport à l'autre sans qu'il n'existe d'étude spécifique qui les inclue tous les deux. Le cas typique d'une situation de comparaisons indirecte est le suivant : on dispose du ou des résultats de différentes études (essais contrôlés randomisés (ECR)) comparant A^a à B et d'une ou plusieurs études comparant A à C, le but étant d'arriver à estimer la différence d'effet entre B et C (sous-entendu risque relatif, odds ratio, différence de risque, hazard ratio ou différence de moyenne). La Figure 1 suivante permet de visualiser plus clairement la démarche.

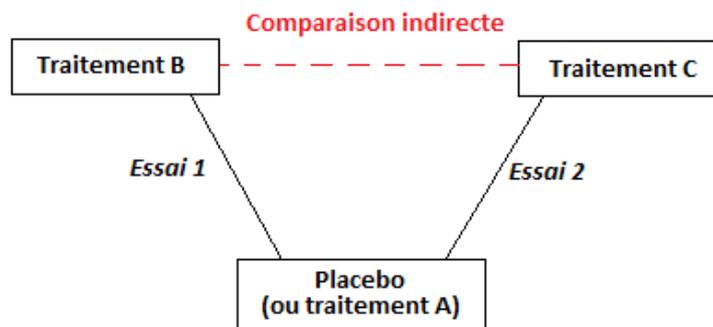


Figure 1 : Principe de la comparaison indirecte.

De manière générale, toute méthode permettant une estimation de l'effet de B par rapport à C à partir des estimations de l'effet de B et de C versus contrôle (ou traitement A) permet de réaliser des comparaisons indirectes. Cependant, certaines méthodes sont inappropriées ou non valides. Elles sont à distinguer des méthodes appropriées ou valides^{4,5}.

Nous allons nous attacher dans cette première partie à étudier les différentes méthodes de comparaisons indirectes, ainsi que les conditions de validité dont elles

^a L'approche reste vraie pour n'importe quel comparateur commun, « A » pouvant désigner le placebo ou un traitement actif. Dans la suite de la présente thèse nous utiliserons indifféremment les deux cas de figure.

dépendent. Cela permettra de comprendre les différents éléments à vérifier à la lecture d'une comparaison indirecte, éléments que nous listerons dans un guide de lecture critique.

1.1 Les méthodes inappropriées : les comparaisons indirectes naïves

Certaines méthodes de comparaisons indirectes parfois retrouvées dans la littérature sont intuitives, mais non valides car elles font l'objet de biais non acceptables. Ces procédés de comparaisons indirectes non ajustées ne doivent pas être utilisés pour effectuer des comparaisons indirectes sur l'efficacité et/ou la toxicité de médicaments. Il s'agit de comparaisons naïves d'estimations ponctuelles et celles de bras actifs.

1.1.1 Comparaisons naïves des estimations ponctuelles

1.1.1.1 Principe

Cette approche consiste à comparer l'estimation ponctuelle traduisant l'efficacité de A vs placebo à celle de B vs placebo (par exemple par le calcul d'un risque relatif). Elle présente de nombreuses limites.

1.1.1.2 Exemple

Il s'agit, par exemple, de comparer le risque relatif^b obtenu sur la mortalité lors de l'évaluation de A vs placebo à celui obtenu lors de l'évaluation de B vs placebo (Tableau 1). Le traitement entraînant la plus grande réduction relative de risque sera qualifié plus efficace que l'autre.

^b La différence d'effet est ici évaluée par le risque relatif, mais on aurait pu indifféremment choisir l'odds ratio, l'hazard ratio ou une différence de moyenne ou de risque. Ceci reste vrai pour la suite de la thèse.

	Risque relatif décès de toute cause	Réduction relative de la mortalité
Traitement A (versus placebo)	0,76	24%
Traitement B (versus placebo)	0,87	13%

Tableau 1 : Résultats des estimations ponctuelles de l'essai A vs placebo et de l'essai B vs placebo⁴.

Dans cet exemple, le traitement A entraîne une réduction relative de la mortalité de 24%, plus importante qu'avec le traitement B (13%). A sera donc qualifié de « plus efficace » que B. Cependant, le biais majeur de cette démarche est de ne pas tenir compte de la précision statistique des estimations.

	Risque relatif Décès de toute cause [IC95%]
Traitement A (versus placebo)	0,76 [0,53;0,99]
Traitement B (versus placebo)	0,87 [0,84;0,90]

Tableau 2 : Résultats des estimations ponctuelles et intervalles de confiance de l'essai A vs placebo et de l'essai B vs placebo⁴.

En effet, si l'on s'intéresse aux intervalles de confiance (Tableau 2), avec le traitement B nous sommes assurés d'une réduction de mortalité d'au moins 10% avec un degré de certitude de 95%, tandis qu'avec le traitement A nous n'avons l'assurance que d'une réduction de 1% de la mortalité (avec un degré de certitude de 95%).

Aussi, la prise en considération de la borne péjorative de l'intervalle de confiance montre que la comparaison des estimations ponctuelles est inappropriée et ne permet pas d'obtenir une méthode satisfaisante de comparaisons indirectes (en dehors de toutes considérations sur la méthodologie des essais réalisés, les populations incluses et les traitements délivrés).

1.1.2 Comparaison naïve des bras actifs

1.1.2.1 Principe

Dans ce type de comparaison, à partir des deux essais A *versus* placebo et B *vs* placebo, seules sont utilisées les effets du bras A du premier essai et celle du bras B du second essai (taux d'échec, médiane de survie etc.). L'effet de A *vs* B est alors estimé en comparant les résultats du bras A à celui du bras B.

Ainsi, seuls les bras traitements actifs sont considérés sans tenir compte de la randomisation, c'est-à-dire de l'appariement au sein de l'essai (Figure 2). De ce fait, il y a perte des différents caractères essentiels des deux essais:

- perte totale de leur caractère contrôlé puisqu'on ne tient plus compte du bras placebo : le résultat n'est plus à l'abri du biais de confusion^c,
- perte totale du bénéfice de la randomisation puisque celle-ci ne s'est pas faite sur la population globale de l'étude A et de l'étude B mais sur chacune séparément : la comparaison effectuée n'est plus à l'abri d'un biais de sélection^d.

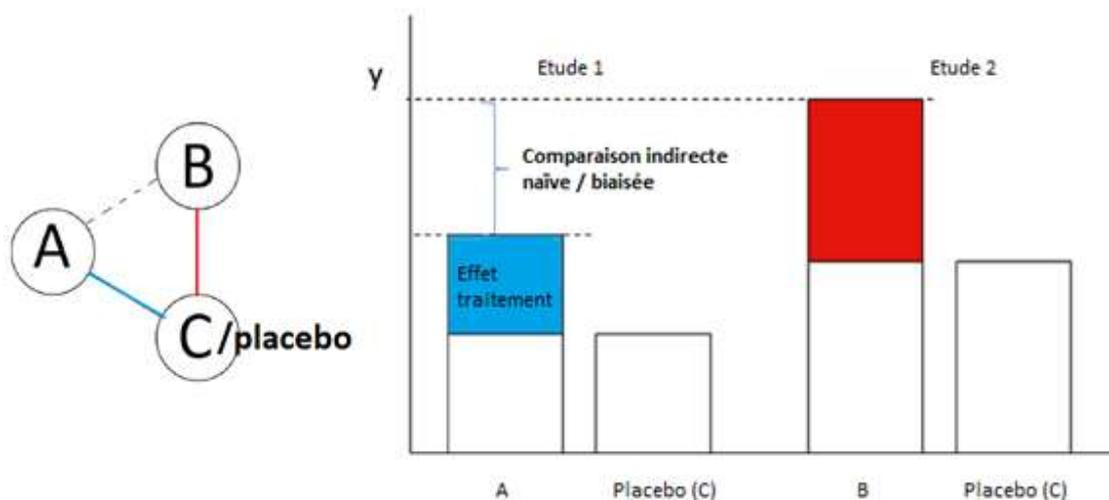


Figure 2 : Comparaison indirecte naïve non valide car les différences d'effet par rapport au placebo dans chaque étude ne sont pas reconnues⁶.

Notes : Y : effet traitement (ex : diminution des œdèmes). Dans cet exemple, la différence de l'effet du traitement de B par rapport à A est surestimée par la comparaison indirecte.

^c Biais de confusion : ensemble d'erreurs qui peuvent survenir dans l'interprétation des liens entre le médicament et son effet, notamment à cause de facteurs appelés « modificateurs d'effets », insuffisamment contrôlés par le protocole de recherche, et qui influencent l'effet du traitement.

^d Biais de sélection : erreur systématique faite lors de la sélection des sujets à étudier.

En plus des arguments théoriques que nous venons d'exposer contre la validité de cette approche, le caractère biaisé a été décrit de manière empirique. En effet, Glenny et al.⁷ ont montré que, lorsqu'elles sont comparées à des estimations faites à partir de comparaisons directes, les comparaisons indirectes non ajustées donnent lieu à un grand nombre de différences concernant l'importance et le sens de l'efficacité relative.

1.1.2.2 Exemple

Afin d'illustrer ce qui précède, prenons un exemple concret :

	Traitement A	Placebo	Risque relatif [IC95%] , valeur de p
Décès	45/1567 (2,87%)	103/1546 (6,66%)	0,43 [0,31;0,61] p <0,001 (S)

Tableau 3 : Résultat de l'essai comparant A au placebo⁴.
S : résultat significatif.

	Traitement B	Placebo	Risque relatif [IC95%] valeur de p
Décès	36/1251 (2,88%)	42/1298 (3,24%)	0,89 [0,57;1,38] p=0,6 (NS)

Tableau 4 : Résultat de l'essai comparant B au placebo⁴.
NS : non significatif.

A partir des résultats présentés dans ces deux tableaux, cette approche de comparaison naïve des bras actifs comparerait 2,87% à 2,88% et inciterait donc 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 vs B proposée est donc confondue par le risque de base.

Cette approche, complètement invalide sur le plan méthodologique, entraîne des résultats imprévisibles et ne prend pas en compte la randomisation ; elle ne doit donc pas être utilisée⁸. Ce fut pourtant la première méthode employée, c'est pourquoi elle est à

l'origine des réticences actuelles vis-à-vis des comparaisons indirectes. Le problème qu'elle pose n'est pas consubstantiel aux comparaisons indirectes. D'autres méthodes ont été développées par la suite, permettant de résoudre ce problème.

1.1.3 Approche par un test d'hétérogénéité

Une approche un peu plus aboutie que la comparaison naïve est parfois utilisée ; il s'agit d'effectuer un test d'hétérogénéité afin d'affirmer que l'effet de A est différent de celui de B. Le test d'hétérogénéité évalue si les résultats des deux essais peuvent être considérés comme similaires (l'hypothèse nulle est « les deux traitements ont une efficacité équivalente », c'est l'hypothèse d'homogénéité). Si le test d'hétérogénéité est significatif, cela oriente pour dire que l'effet de A est différent de l'effet de B. Ainsi, cette approche est appropriée pour savoir si deux traitements ont une « efficacité » différente, mais elle est limitée car elle ne permet pas de quantifier l'efficacité de l'un par rapport à l'autre. Les méthodes dites « appropriées » que nous allons voir permettent de pallier ce manque.

1.2 Les méthodes appropriées

1.2.1 La méthode de Bucher de comparaison indirectes « ajustées » ou « extrapolées »

L'estimation indirecte « ajustée » (« *adjusted indirect comparison* ») ou « extrapolée » présentée par Bucher et al. en 1997⁹ est la première méthode valide disponible permettant de réaliser des comparaisons indirectes. Afin de surmonter le problème potentiel des facteurs pronostics différents entre les participants des différentes études impliquées dans la comparaison, Bucher et son équipe suggèrent une méthode simple, dans laquelle la comparaison indirecte de A et B est ajustée selon les résultats de leur comparaison directe avec une intervention commune, C.

1.2.1.1 Principe

Cette méthode de comparaison indirecte deux à deux prend en compte l'estimation de l'effet de A vs placebo et celle de B vs placebo, c'est pourquoi elle est dite « ajustée ». De ce fait, la randomisation est préservée, et les propriétés méthodologiques de l'essai contrôlé randomisé sont conservées.

1.2.1.1.1 Calcul du risque relatif issu de la comparaison indirecte

Le calcul nécessaire est 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 : risque relatif issu de la comparaison du traitement A au placebo
- RR_b : risque relatif issu de la comparaison du traitement B au placebo
- $RR_{a/b}$: risque relatif issu de la comparaison indirecte du traitement A vs B

Le risque relatif $RR_{a/b}$ est obtenu grâce au calcul : $RR_{a/b} = \frac{RR_a}{RR_b}$.

Pour les différences (de risques, moyennes ; symbolisées par la lettre d), l'extrapolation se calcule ainsi :

$$-d_{A/B} = d_{B/placebo} - d_{A/placebo}$$

De manière usuelle, la méta-analyse se fait sur les variables exprimées en logarithme par rapport à leurs propriétés statistiques (loi normale). La conversion en termes de RR se fait dans un second temps. Aussi, lorsque l'efficacité relative est exprimée en logarithme du risque relatif (abrégé $\ln RR$), la formule devient alors :

$$\ln(RR_{a/b}) = \ln RR_a - \ln RR_b$$

La démonstration de cette équation est simple :

Dans un essai de comparaison directe, les patients ont un risque d'événement sans traitement de r_0 .

Dans le bras recevant le traitement A, le risque sous traitement r_a devient :

$$r_a = r_0 \times RR_a \quad (1)$$

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

$$r_b = r_0 \times RR_b \quad (2)$$

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 (1) et (2)

$$RR_{a/b} = \frac{RR_a}{RR_b}$$

Ce qui justifie le calcul de la comparaison indirecte ajustée.

Dans ces calculs, le risque de base r_0 disparaît complètement (sous réserve d'un risque comparable entre les deux populations). Cette extrapolation, basée sur les mesures des effets, est donc totalement insensible au risque de base des essais A et B (Figure 3).

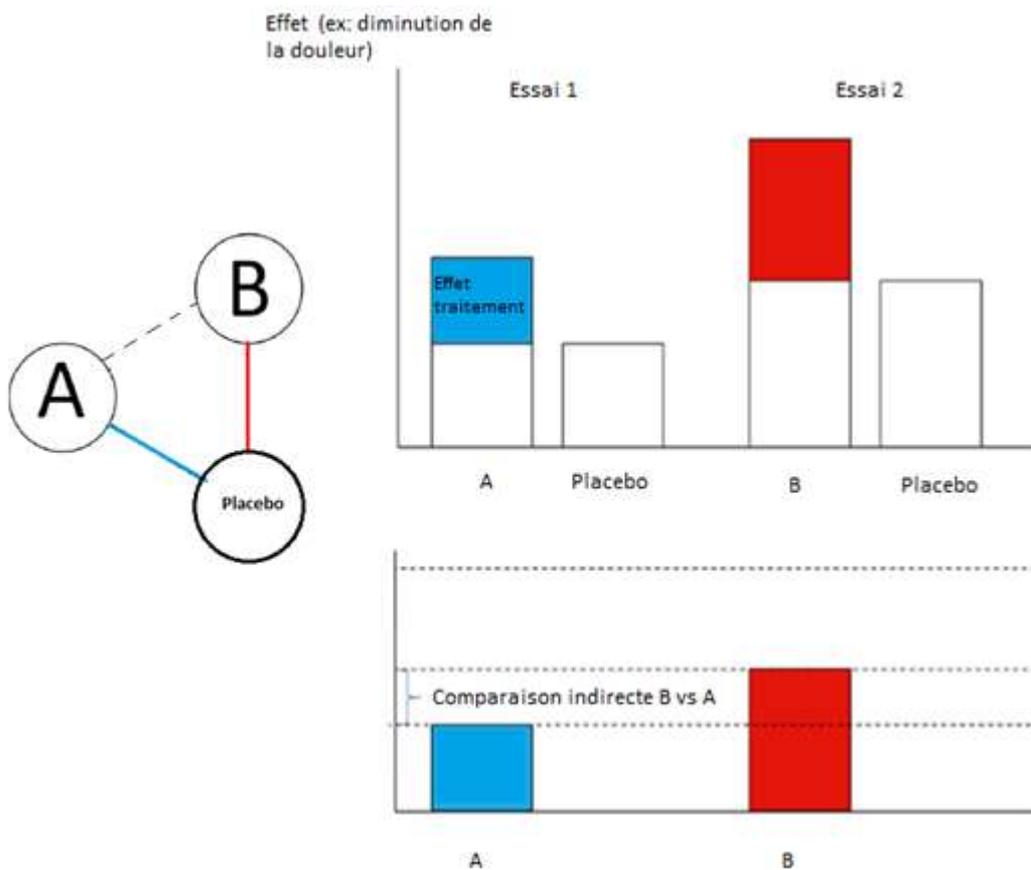


Figure 3 : Comparaison indirecte ajustée B vs A. La comparaison indirecte ajustée tient compte du risque de base r_0 ⁶.

1.2.1.1.2 Calcul de l'intervalle de confiance du risque relatif

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.

$$\text{var}(\ln RR_{ab}) = \text{var}(\ln RR_a) + \text{var}(\ln RR_b)$$

ou en utilisant SE l'erreur standard (écart dû à l'estimation du paramètre) :

$$SE(\ln RR_{ab}) = \sqrt{SE(\ln RR_a)^2 + SE(\ln RR_b)^2}$$

Sachant que $SE(\ln RR_a) = [\ln(\text{borne inférieure}) - \ln(\text{borne supérieure})] / 2 / 1,96$

Avec « borne inférieure » et « borne supérieure » : les bornes inférieures et supérieures de l'intervalle de confiance du risque relatif RR_a (même calcul pour RR_b).

Au total :

$IC_{95\%} RR_{ab} = [e^{\ln(RR_{ab}) \pm 1,96 \times SE(\ln RR_{ab})}]$

1.2.1.2 Utilisation de la méta-analyse

En pratique, on se retrouve souvent confronté à 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, mais sélectionner un seul essai de A et un seul de B introduirait un choix arbitraire inacceptable dans la démarche.

Aussi, la solution consiste à effectuer dans un premier temps la méta-analyse de tous les essais de A versus placebo et celle des essais de B versus placebo¹⁰. 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 4).

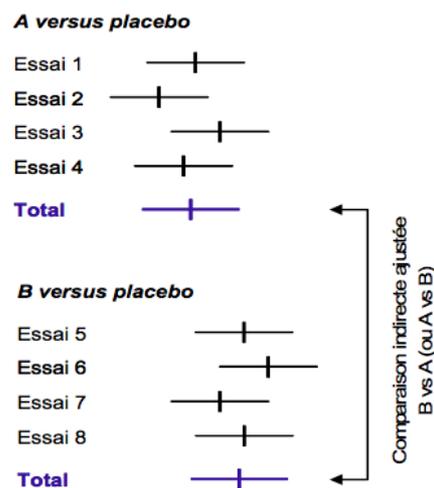


Figure 4 : Utilisation de la méta-analyse pour le calcul des comparaisons indirectes ajustées⁴.

De plus, dans la mesure où cette technique regroupe des études réalisées sur des populations et avec des protocoles quelques peu différents, le caractère généralisable des

résultats issus d'une méta-analyse est *a priori* supérieur à celui des résultats issus d'une seule étude. Cette synthèse des données contribue à augmenter la stabilité de l'effet des traitements et surtout permet de tester cette stabilité, comme nous le verrons dans le chapitre 2.

1.2.1.3 Exemple

Pour illustrer la méthode de comparaison indirecte ajustée, prenons l'exemple de la rispéridone versus halopéridol dans le traitement de la schizophrénie¹¹. Considérons une méta-analyse qui regroupe 3 essais contrôlés comparant la rispéridone au placebo et 9 essais contrôlés comparant l'halopéridol au placebo. L'efficacité relative est exprimée en odd ratio et logarithme de l'odd ratio. La comparaison indirecte consiste à comparer la rispéridone à l'halopéridol ($OR_{\text{resp/hal}}$) grâce aux résultats des comparaisons directes. Les résultats des essais sont regroupés dans le Tableau 5.

Comparaison	Nombre d'essais	Log OR (SE)	OR (IC 95%)
Essais contrôlés contre placebo			
Rispéridone vs placebo	3	-0,909 (0,218)	0,40 (0,26 ; 0,62)
Halopéridol vs placebo	9	-1,707 (0,318)	0,18 (0,10 ; 0,34)
Risperidone vs halopéridol			
Comparaison indirecte ajustée	3/9	0,798 (0,386)	2,22 (1,04 ; 4,72)

Tableau 5 : Méta-analyse de rispéridone vs halopéridol dans le traitement de la schizophrénie¹¹.

L'application des formules de comparaisons indirectes ajustées nous permettent de calculer

$\ln OR_{\text{resp/hal}}$ et $OR_{\text{resp/hal}}$:

- $\ln OR_{\text{resp/hal}} = \ln OR_{\text{resp}} - \ln OR_{\text{hal}}$

$$\ln OR_{\text{resp/hal}} = -0,909 - (-1,707)$$

$$\ln OR_{\text{resp/hal}} = 0,798$$

- $OR_{risp/hal} = \frac{OR_{risp}}{OR_{hal}}$

$$OR_{risp/hal} = \frac{0,40}{0,18} = 2,22 \text{ ou tout simplement } e^{(0.798)} = 2,22 .$$

De même avec l'intervalle de confiance :

- $SE(\ln OR_{risp/hal}) = \sqrt{0,218^2 + 0,318^2}$

$$SE(\ln OR_{risp/hal}) = 0,386$$

- La formule pour calculer les bornes inférieure (-)/supérieure (+) de l'IC pour le log OR est :

Log OR +/- [1,96 x SE (log OR)] (1,96 correspond au quantile d'ordre 2,5% de la loi normale)

$$\text{Soit } 0,798 \text{ +/- } (1,96 \times 0,386) = [0,04144 ; 1,55456]$$

Pour obtenir l'IC en termes d'OR, il faut passer en exponentielle :

$$e^{(0,04144)} = 1,04 \text{ et } e^{(1,55456)} = 4,72$$

soit IC (OR) = [1,04 ; 4,72].

1.2.1.4 L'approche mixte des comparaisons indirectes et directes (Mixed Treatment Comparison, MTC)

1.2.1.4.1 Principe

Les comparaisons indirectes peuvent conserver de l'intérêt même en cas de disponibilités d'essais de comparaisons directes. En effet, l'essai de comparaison directe est souvent unique. Assez fréquemment, il a été arrêté à l'analyse intermédiaire ce qui entraîne 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* » (MTC), 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^{12,13}.

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 en tenant compte de la totalité de l'information disponible. Cette condition de concordance ou « cohérence » des estimations issues des comparaisons directes et indirectes est développée dans le chapitre 2.1.3.1. Le regroupement des estimations issues des comparaisons indirectes ajustées avec celles des comparaisons directes s'effectue simplement par méta-analyse.

1.2.1.4.2 Exemple

Song *et al.*⁸ donnent deux exemples de l'application de cette technique issus des publications faites par Soo *et al.*¹⁴ et Trindade et Menon¹⁵ où, l'estimation des comparaisons directes étant en parfaite phase avec celles issues des comparaisons indirectes ; la meilleure estimation possible est celle qui combine les résultats des deux approches (Figure 5).

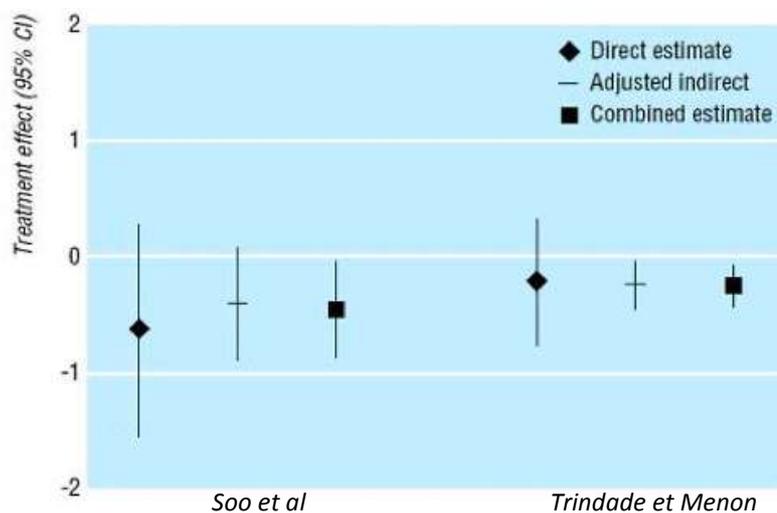


Figure 5 : Deux exemples d'intégration par méta-analyse des résultats des comparaisons directes avec ceux issus des comparaisons indirectes ajustées⁸.

1.2.1.5 Comparaison de plus de deux traitements : limite des comparaisons indirectes ajustées

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. Par exemple s'il existe 4 traitements T1, T2, T3 et T4 dans un domaine thérapeutique, les comparer entre eux revient à effectuer 6 comparaisons directes :

- 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 possibilités d'en trouver une du seul fait du hasard (les 6 comparaisons 2 à 2). Or, si chaque comparaison est effectuée 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 un risque sur quatre de conclure à l'existence d'un traitement plus efficace qu'un autre alors qu'en réalité ils n'y pas de différence entre les traitements.

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 cette limitation ont été développées : il s'agit des méthodes de *méta-analyses en réseaux* et des méthodes de *méta-régression*.

1.2.2 Les méta-analyses en réseau (*Network meta-analysis* ou *Multiple-Treatments Meta-Analysis* MTM)

Comme alternative à plusieurs méta-analyses séquentielles et plusieurs comparaisons indirectes («*pairwise comparison*»), un modèle statistique qui reflète les relations mathématiques entre les estimations relatives des résultats de plusieurs comparaisons directes et indirectes peut être utilisé afin d'estimer globalement et simultanément toutes les comparaisons deux à deux. Aussi, une approche plus sophistiquée que les comparaisons indirectes ajustées, dite « méta-analyse en réseau », a été introduite pour la première fois par Lumley en 2002¹⁶. La méta-analyse en réseau utilise le plus souvent des méthodes d'estimation bayésiennes^{13,12} et permet de combiner comparaisons directes et indirectes afin d'estimer toutes les comparaisons possibles entre les différents traitements évalués dans le réseau.

Les objectifs de la méta-analyse en réseau sont multiples : répondre aux questions de recherche en l'absence de comparaisons directes, améliorer la précision des estimations en combinant preuves directes et indirectes, classer les traitements les uns par rapports aux autres, et évaluer l'impact de certains éléments du réseau d'essais.

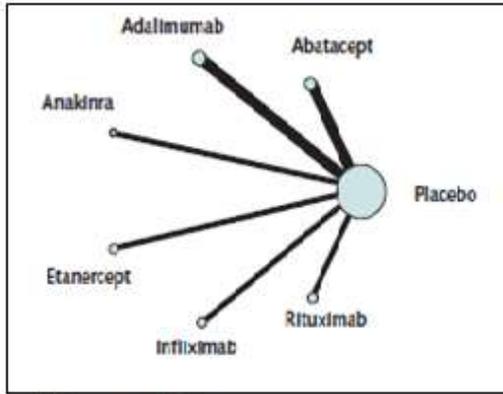
1.2.2.1 Principe de la méta-analyse en réseau et d'un réseau d'essais

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 le réseau, les nœuds représentent les traitements. Les flèches (ou les lignes) reliant les sommets représentent les comparaisons issues des essais randomisés disponibles impliquant les traitements situés de part et d'autres de la ligne. Le nom de l'étude et le risque relatif (ou odd ratio) avec l'intervalle d'incertitude correspondant sont parfois précisés le long de la ligne. Le sens de la flèche donne le sens de la comparaison et des risques relatifs. Parfois, la largeur des lignes est proportionnelle au nombre d'essais de comparaison de chaque paire de traitements, et la taille de chaque nœud est proportionnelle au nombre de participants répartis de façon aléatoire (taille de l'échantillon). S'il n'y a pas de ligne entre 2 nœuds, cela signifie qu'il n'existe aucune étude (c'est-à-dire aucune preuve directe) comparant les deux médicaments. Avec une méta-analyse en réseau,

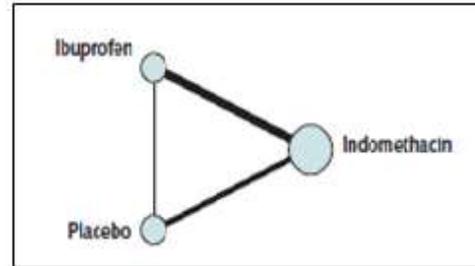
l'efficacité relative de ces deux traitements peut être estimée, même si aucune étude ne les compare directement entre eux.

Différentes représentations des réseaux de comparaisons existent en fonction des situations (Figure 6) :

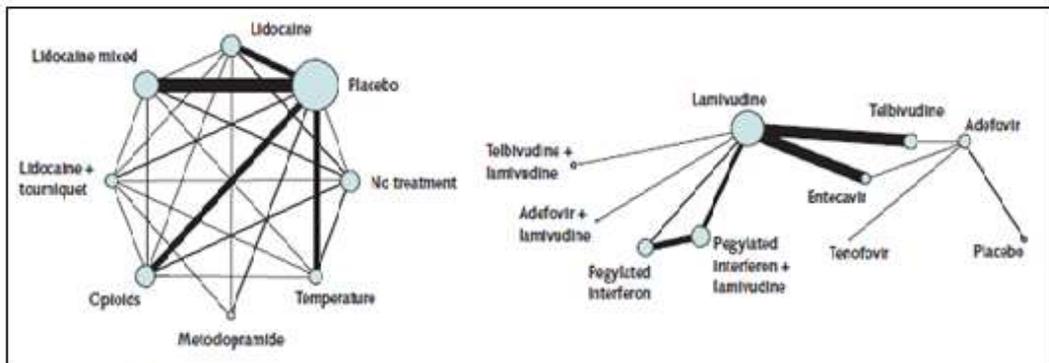
- a) l'étoile montre un réseau dans lequel toutes les interventions ont un seul comparateur commun, comme le placebo par exemple ;
- b) la boucle fermée implique trois interventions et peut fournir des données par des comparaisons directes et indirectes (« *mixed treatment comparison* ») ;
- c) le réseau connecté dans lequel toutes les interventions ont été comparées les unes aux autres dans plusieurs essais contrôlés randomisés ;
- d) le réseau complexe dispose de plusieurs boucles et des bras qui peuvent avoir des connections ;
- e) le réseau non connecté ou boucle ouverte dans lequel l'effet du traitement de C par rapport B peut être obtenu grâce à la comparaison indirecte des études A versus B et A versus C.



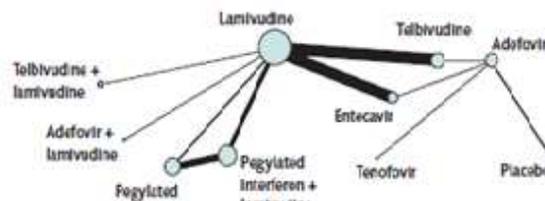
a) Réseau en étoile



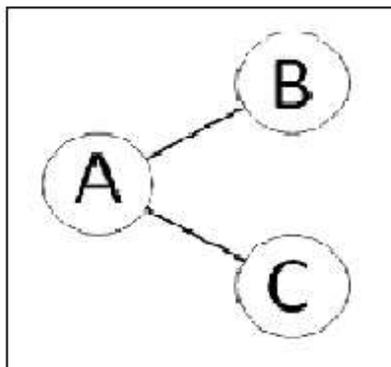
b) Boucle seule fermée



c) Réseau connecté



d) Réseau complexe



e) Réseau non connecté

Figure 6 : Exemple de géométries de réseaux¹⁸.

Les termes « comparaison indirecte », « mixed treatment comparison (MTC) », et « méta-analyse en réseau » sont parfois utilisés de façon interchangeable. Afin de clarifier les choses, nous proposons d'utiliser le terme « méta-analyse en réseau » lorsque la base de preuves se compose de plus de deux essais contrôlés reliant plus de deux interventions. Si le réseau est constitué d'au moins une boucle fermée, le terme « MTC » est approprié.

L'analyse d'une boucle ouverte (comme dans la Figure 2 et la Figure 3) peut-être appelée « comparaison indirecte (ajustée) ».

La Figure 7 donne un exemple d'un réseau issu de la méta-analyse des traitements antihypertenseurs publiée en 2007 par Elliot et Meyer¹⁹. Cette méta-analyse s'intéressait à l'incidence du diabète avec les traitements antihypertenseurs. Les comparaisons directes disponibles sont représentées par les flèches avec indication du nom de l'essai, de l'odds ratio (OR) et de son intervalle de confiance. Dans ces essais, les patients n'avaient pas de diabète au moment de la randomisation.



ACE (angiotensin-converting-enzyme-inhibitor) = IEC
 CCB (Calcium Channel Blocker) = antagonistes calciques
 ARB (Angiotensin II Receptor Antagonist) = sartans

Figure 7 : Réseaux des essais d'antihypertenseurs dans lesquels l'incidence du diabète est rapportée, d'après Elliott et Meyer, 2007¹⁹.

Il convient d'être prudent à la lecture de ce type de figure, car 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 ratios : à l'issue de cette analyse, les sartans (ARB) seraient considérés comme étant supérieurs aux IEC (ACE inhibitor), parce que l'OR

sartan versus diurétique (0,12) est plus intéressant que l'OR IEC versus diurétique (0,66). Or l'interprétation correcte de cette figure doit se limiter à : « Cette méta-analyse en réseau suggère que le placebo, les antagonistes calciques (CCB), les IEC et les sartans entraînent moins de cas incidents de diabète que les diurétiques (leurs OR étant inférieurs à 1). Aucune conclusion n'est possible pour les bêtabloquants car l'intervalle de confiance contient la valeur 1.»⁴.

1.2.2.2 Estimation globale d'un réseau d'essais

La méta-analyse en réseau peut être effectuée avec une méthode fréquentiste ou une méthode Bayésienne (Figure 8).

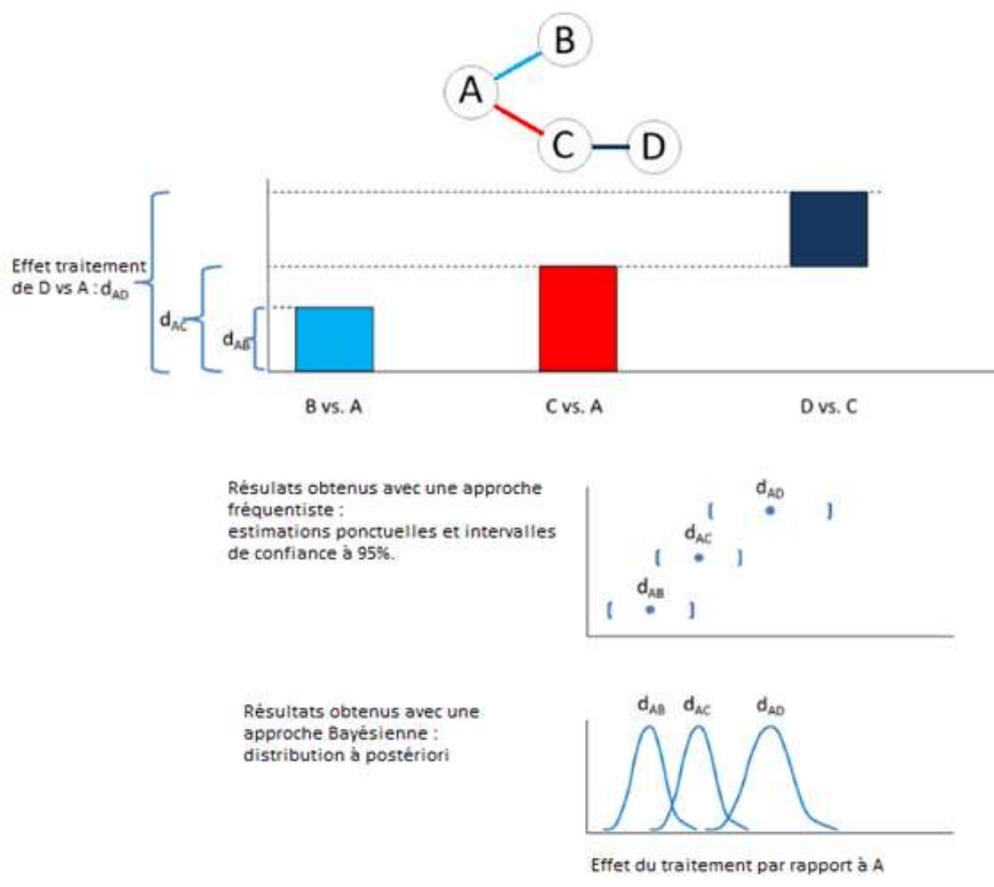


Figure 8 : Approche Fréquentiste et Bayésienne d'une méta-analyse en réseau⁶.

1.2.2.2.1 Méthode fréquentiste

Avec la méthode fréquentiste, le résultat de l'analyse des comparaisons de traitements est une estimation de l'effet relatif du traitement avec une valeur de P et un IC à

95%. La valeur de P indique si les résultats sont statistiquement « significatifs » ou « non significatifs ». L'utilité de la valeur de P, cependant, est limitée pour la prise de décision, car elle ne fournit pas d'information sur la probabilité qu'une hypothèse (par exemple, « un traitement est meilleur que l'autre ») soit vraie ou fausse. Il faut noter que les IC obtenus avec une approche fréquentiste ne peuvent pas être interprétés en termes de probabilités; en effet, un IC à 95% ne signifie pas qu'il y a de 95% de chance que la « vraie valeur » se situe entre les limites de l'intervalle mais que si l'expérience est répétée un grand nombre de fois alors 95% des intervalles de confiance calculés contiendront la vraie valeur du paramètre ²⁰. En outre, lorsque le décideur est confronté à un choix entre plus de deux traitements, l'interprétation d'une valeur de P associée à chaque comparaison par paires dans une méta-analyse en réseau devient encore plus difficile, car il ne fournit pas d'informations sur le classement des traitements. D'autre part, certaines méthodes fréquentistes impliquent des approximations et des hypothèses qui ne sont pas déclarées explicitement ou vérifiées lorsque les méthodes sont appliquées. Aussi, il faudra être vigilant quant à ces éléments lors de la lecture d'une comparaison indirecte.

Lumley ¹⁶ a décrit une méthode d'estimation à l'aide d'un modèle linéaire mixte; sa méthode a été appliquée dans relativement peu d'articles^{19,21,22}. Cette approche fréquentiste a été développée avec le logiciel R^e ; elle est également réalisable en utilisant le logiciel SAS^e.

1.2.2.2 Méthode Bayésienne

Contrairement à la statistique dite « classique » ou « fréquentiste » où le paramètre inconnu est considéré comme fixe, en statistique bayésienne, le paramètre inconnu, noté θ , est aléatoire et il lui est affecté une loi de probabilité qui est un moyen de traduire numériquement un degré de connaissance.

Dans une approche bayésienne, l'état d'incertitude sur le paramètre avant l'obtention de données cliniques est traduit en spécifiant sa loi de probabilité via sa distribution *a priori* notée $g(\theta)$. Cette loi représente l'ensemble des informations *a priori* disponibles sur le paramètre θ (ainsi que les imprécisions qui s'y rattachent). Sa

^e Le logiciel R et le logiciel SAS (Statistical Analysis System) sont des logiciels orientés vers les statistiques et la représentation graphique des données.

détermination est un aspect central et critique de l'approche bayésienne sur lequel nous reviendrons ultérieurement.

Supposons maintenant que l'on dispose des données d'un essai clinique, notées y , dont la probabilité d'occurrence dépend de θ . Cette dépendance est exprimée par la probabilité conditionnelle de y pour chaque valeur possible de θ , $f(y|\theta)$, aussi appelée vraisemblance de θ ou modèle d'échantillonnage. La statistique bayésienne se fonde sur le théorème dit de Bayes qui permet de calculer la loi de probabilité de θ en prenant en compte les données de l'essai, c'est-à-dire la densité *a posteriori* $g(\theta|y)$, selon la formule suivante (écrite ici pour un paramètre continu) : $g(\theta|y) \propto f(y|\theta) \times g(\theta)$. La forme continue du théorème de Bayes indique que la distribution *a posteriori* s'obtient en multipliant la distribution *a priori*, par la vraisemblance, et en effectuant une normalisation, du fait qu'il s'agit d'une densité de probabilité. La loi *a posteriori* $g(\theta|y)$ synthétise les informations sur le paramètre d'intérêt après avoir observé les données. Elle prend en compte les croyances initiales (*a priori*) et les données (vraisemblance). En partant d'une incertitude initiale plus ou moins grande sur le paramètre, l'inférence bayésienne permet donc d'exprimer, une fois les données recueillies, la nouvelle incertitude sur les valeurs possibles du paramètre, comme illustré dans la Figure 9.

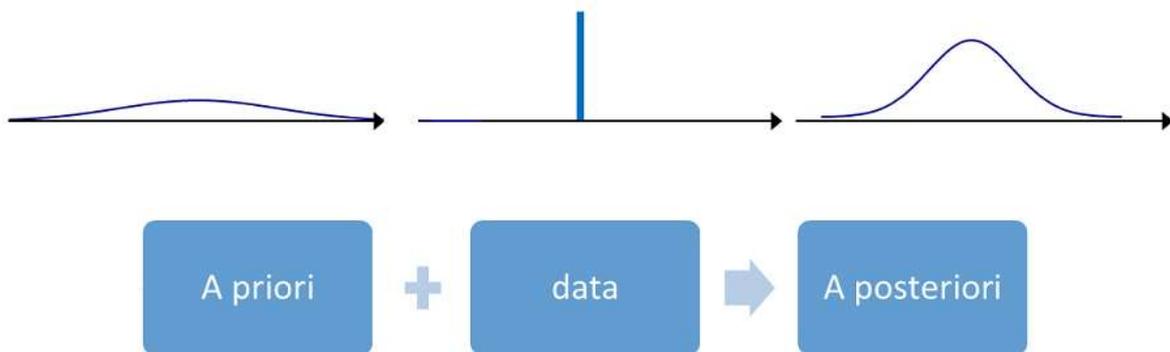


Figure 9 : Méthode d'estimation Bayésienne.

Le principe de base en inférence bayésienne est donc que la densité *a posteriori* $g(\theta|y)$ contient toute l'information sur le paramètre inconnu θ une fois les données observées. Des estimateurs ponctuels peuvent être fournis, comme par exemple, la médiane, la moyenne ou le mode de la distribution *a posteriori*. La loi *a posteriori* fournissant une

information en termes d'incertitude, il peut être également construit un intervalle probable pour le paramètre θ , appelé intervalle de crédibilité. Il ne s'agit pas d'un intervalle de confiance dont les bornes sont aléatoires et qui concerne un paramètre θ fixé. Pour un intervalle de crédibilité, les bornes sont fixes et θ est aléatoire. L'intervalle de crédibilité a donc une interprétation naturelle qui porte directement sur la valeur inconnue que l'on cherche à cerner.

Le choix de la loi *a priori* est crucial pour la suite de l'analyse : deux lois *a priori* différentes conduisent à deux systèmes d'inférence différents, même si l'accumulation de données finit par effacer cette différence. Tout au plus peut-on chercher à réduire l'influence de ce choix en retenant des lois dites non informatives, qui minimisent l'information apportée par la loi *a priori* face à celle apportée par les données, les résultats se rapprochant alors des méthodes fondées sur la seule vraisemblance. Un certain nombre de distributions *a priori* non informatives ont été proposées reposant sur des principes et critères divers. Citons comme exemples, l'*a priori* de Jeffreys²³ ou les *a priori* dits de référence décrits par Bernardo et Smith²⁴. Une autre approche est de considérer que la distribution *a priori* doit traduire les connaissances avant l'observation des données. Il s'agit alors de construire des distributions dites informatives, en se basant notamment sur les connaissances des experts du domaine étudié ou sur les résultats d'essais antérieurs. Le paradoxe de l'approche bayésienne est qu'il n'existe pratiquement jamais une unique loi *a priori*, mais plutôt un ensemble de lois compatibles avec les informations *a priori* disponibles. Ainsi, Spiegelhalter et al.²⁵ ont-ils préconisé le choix de plusieurs distributions, afin d'étudier la congruence des résultats obtenus à partir de différentes distributions *a priori*.

Le calcul de la densité *a posteriori* $g(\theta|y)$ peut se faire de façon analytique mais cela suppose que des conditions précises soient réunies en ce qui concerne les familles des lois *a priori* et *a posteriori* (modèles dits conjugués). Dans le cas de modèles à structure complexe comme ceux des méta-analyses en réseau, le calcul repose sur des méthodes d'approximation, basées sur la simulation selon la loi *a posteriori*⁽²⁷⁾. Les méthodes de simulation utilisées pour l'inférence statistique bayésienne sont celles dites de Monte Carlo par Chaînes de Markov (*Monte Carlo Markov Chains* en anglais ou MCMC). L'algorithme de Metropolis (et ses variantes) est l'un des algorithmes MCMC les plus utilisés. Le principe est

de constituer un large échantillon aléatoire de θ , ($\theta(1)$, $\theta(2)$, ... $\theta(n)$), issu de la loi *a posteriori*. L'expression mathématique de la loi *a posteriori* $g(\theta|y)$ reste inconnue mais on peut représenter la distribution *a posteriori* de θ par un histogramme et calculer empiriquement des grandeurs statistiques la décrivant, telles que moyenne, variance, percentiles, etc. Les algorithmes de Metropolis ont été implémentés dans différents logiciels facilitant leur utilisation : le logiciel BUGS (pour *Bayesian inference Using Gibbs Sampling*) sous ses deux versions, WinBUGS et OpenBUGS et plus récemment le logiciel JAGS (pour *Just another Gibbs sampler*) et le logiciel STAN (pour *Stanislaw Ulam*, pionnier des méthodes de Monte Carlo). Le projet BUGS a été initié en 1989 au Medical Research Council Biostatistics Unit à Cambridge ^(28,29). Les logiciels BUGS utilisent l'échantillonnage de Gibbs qui est une des variantes de l'algorithme de Metropolis. Le calcul de la densité *a posteriori* avec ces logiciels, nécessite de fournir:

- les données y ;
- le modèle : loi *a priori* $g(\theta)$ et vraisemblance $f(y|\theta)$;
- la valeur initiale de chaque chaîne ;
- les nombres de chaînes et d'itérations souhaitées par chaîne.

Après un certain nombre de tirages, la chaîne simulée converge vers la densité *a posteriori* du paramètre du modèle, $g(\theta|y)$. Il s'agit là d'une garantie théorique qui ne nous renseigne pas sur le nombre d'itérations à partir duquel la chaîne de Markov a effectivement convergé. Pour produire des estimations avec suffisamment de confiance, il faut donc être capable d'évaluer le nombre de simulations à retirer de l'échantillon final, appelé période de rodage. En pratique, la convergence est difficile à affirmer. Il s'agit plutôt d'un diagnostic d'absence de non-convergence. Différents outils, notamment graphiques, sont disponibles sous les logiciels BUGS qui visent à examiner les tendances au sein de la chaîne elle-même et le bon mélange entre les chaînes (c'est-à-dire l'influence des valeurs initiales). Nous nous contenterons ici de les citer : graphique de la trace du paramètre d'intérêt, diagnostic d'auto-corrélation, test de Gelman-Rubin.

Le résultat de la méta-analyse bayésienne est donc une distribution conjointe *a posteriori* de tous les effets de traitement relatifs pour toutes les interventions incluses dans le réseau. Les effets du traitement sont généralement affichés avec l'intervalle de crédibilité à 95% (CRI)^{26,27}. La méthode bayésienne permet également d'estimer la probabilité qu'un traitement soit supérieur à un ou plusieurs autres. Les résultats sont alors formulés de la

façon suivante : « Il y a x % de probabilité que le traitement C soit meilleur que le traitement B », ou « Il y a y % de probabilité que le traitement D soit plus efficace que les traitements A, B, C, D, et E en ce qui concerne ce résultat R ». La méthode permet ainsi de classer les médicaments les uns par rapport aux autres en termes d'efficacité²⁸. Cependant, il faut toujours garder à l'esprit les tailles d'effet car un bon rang n'implique pas nécessairement une taille d'effet cliniquement significative. De plus l'intervalle de crédibilité est important à considérer.

Un des avantages majeurs de l'approche bayésienne est ainsi de fournir des outils facilitant la prise de décision i) en produisant des conclusions directement interprétables par le décideur en santé, ii) en quantifiant l'incertitude sur le paramètre d'intérêt (ici, l'effet traitement) et iii) en ordonnant les stratégies de santé.

1.2.2.3 Présentation des résultats d'une méta-analyse en réseau

1.2.2.3.1 Le réseau de preuves

Comme nous l'avons vu précédemment (Figure 7), la représentation graphique sous forme d'un réseau de preuves est un format accessible et compréhensible pour décrire les données ayant contribué à la méta-analyse. Cependant, il est important de mettre en avant les estimations issues des comparaisons directes et les estimations de la méta-analyse en réseau (qui peuvent être basées sur des preuves indirectes seules ou combinées avec les preuves directes), tous avec leurs intervalles d'incertitude correspondants. Cela donne une image complète et transparente de la preuve disponible. La plupart des méta-analyses en réseau publiées utilisent un tableau ou un graphique pour communiquer cette information. Les tailles d'effet sont souvent présentées sous forme d'une « league table » (Tableau 6) ou dans un « forest plot » (Figure 10) par rapport à un comparateur commun²⁹.

	Médicament A	Placebo	Médicament B	Médicament C	Médicament D
Médicament A		0.71 (0.59,0.84)	0.94 (0.72,1.23)	0.97 (0.74,1.26)	1.01 (0.79,1.29)
Placebo	1.42 (1.19,1.68)		1.33 (1.08,1.65)	1.37 (1.12,1.68)	1.43 (1.19,1.71)
Médicament B	1.06 (0.81,1.40)	0.75 (0.61,0.93)		1.03 (0.77,1.38)	1.07 (0.81,1.42)
Médicament C	1.03 (0.79,1.34)	0.73 (0.60,0.89)	0.97 (0.72,1.30)		1.04 (0.80,1.36)
Médicament D	0.99 (0.78,1.27)	0.70 (0.58,0.84)	0.93 (0.71,1.23)	0.96 (0.74,1.25)	

Tableau 6 : Exemple de league table (Hazard ratios et intervalles de crédibilité à 95% pour la survie globale). Les HR sont donnés pour le traitement en ligne relativement au traitement en colonne; un résultat inférieur à 1 est en faveur du traitement en ligne. Exemple : Le médicament A est supérieur au placebo en termes de survie globale. Le médicament D est supérieur au médicament A.

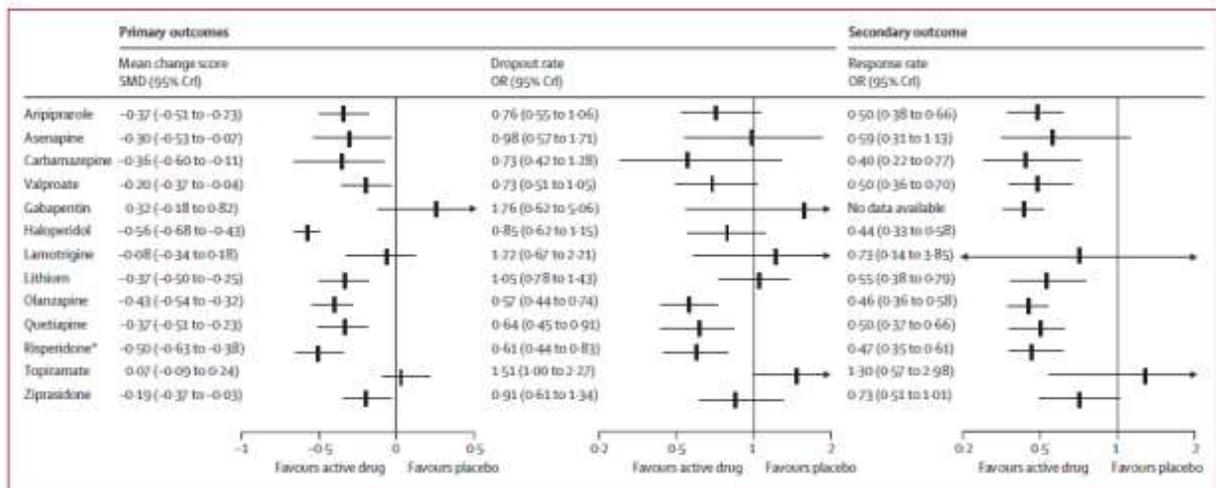


Figure 10 : Forest plot des résultats de la méta-analyse en réseau concernant l'efficacité et le taux d'abandon (tolérance) des antipsychotiques par rapport au placebo²⁹.

Les différences moyennes standardisées inférieures à 0 et les OR inférieurs à 1 sont en faveur du composé actif. * Les données de la rispéridone et la palipéridone ont été fusionnées. MTM = méta-analyse multi-traitements. OR = rapport des cotes. CrI = intervalle de crédibilité.

1.2.2.3.2 Le classement des probabilités pour chaque traitement d'être le meilleur

Comme nous l'avons vu, la méta-analyse en réseau permet l'estimation de la probabilité de chaque traitement d'être le meilleur pour chaque critère. Les classements des probabilités peuvent être résumés avec un graphique dans lequel on trouve sur l'axe horizontal le rang pour chaque traitement, et sur l'axe vertical une probabilité. Un tel graphique est appelé un « rankogram »¹⁸ (Figure 11). Il peut être représenté en valeurs

absolues ou en valeurs cumulatives (voir Annexe 1). La représentation est aussi possible grâce à un graphique en barres ou un tableau.

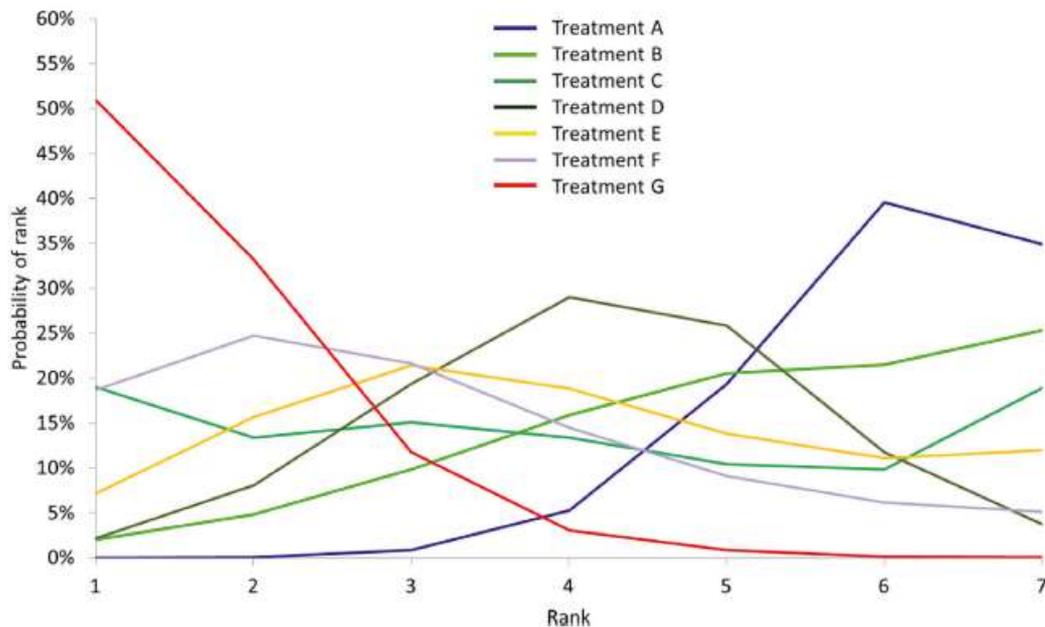


Figure 11 : "Rankograms" montrant la probabilité pour chaque traitement d'être à un rang déterminé dans la hiérarchie de traitements. (*rank=rang*)¹⁸.

1.2.2.3.3 Les analyses de sensibilité

Compte tenu de la complexité de certaines méta-analyses MTC, les auteurs peuvent évaluer la robustesse de leurs résultats en appliquant des analyses de sensibilité qui montrent la manière dont certains résultats varient si certains critères ou hypothèses changent³⁰. Par exemple, dans une MTC sur la prévention des exacerbations de bronchopneumopathie chronique obstructive (BCPO), les auteurs ont utilisé le taux d'incidence comme critère principal. Cependant, il y a un débat sur la pertinence de l'utilisation du taux d'incidence pour évaluer les essais sur la BCPO³¹ de sorte que les auteurs ont procédé à des analyses de sensibilité en utilisant le résultat binaire présence/absence d'exacerbation. Les résultats étaient suffisamment similaires pour conclure à la robustesse de l'analyse³².

1.2.3 La méta-régression

La méthode de méta-régression pour comparer des médicaments consiste également à colliger les informations présentes dans les essais cliniques randomisés et contrôlés

retrouvés dans la littérature et à en faire une synthèse quantitative à la manière des méta-analyses mais en utilisant la méthode de la régression linéaire multiple³³.

S'il existe un nombre suffisant d'études incluses dans la méta-analyse en réseau, il peut être possible d'effectuer une méta-régression dans laquelle l'effet relatif du traitement dans chaque étude est une fonction non seulement de la comparaison de cette étude, mais aussi d'un modificateur d'effet³⁴⁻³⁸ (parfois également appelé covariable). En effet, certaines caractéristiques des patients ou de l'étude (par exemple : sexe, âge, manière dont les résultats ont été mesurés ou définies, exigences du protocole, y compris les traitements concomitants autorisés, durée de suivi, époque pendant laquelle les études ont été menées, critères de gravité de la pathologie etc ...) peuvent influencer l'effet traitement³⁹. On appelle ces caractéristiques des « modificateurs de l'effet traitement » (« *effect modifiers* ») ; la méta-régression permet de prendre en compte ces paramètres.

En d'autres termes, avec un modèle de méta-régression, on estime l'effet du traitement, pondéré par les différences en termes de modificateur d'effet entre les études. Par exemple, s'il existe des différences dans la proportion de sujets atteints de maladie grave entre les études incluses dans la comparaison indirecte et que la gravité de la maladie affecte l'efficacité d'au moins une des interventions comparées, alors une méta-régression peut être utilisée pour ajuster la comparaison indirecte (Figure 12). En outre, le modèle peut prédire les résultats de comparaison indirecte pour différentes proportions de patients atteints de maladie grave dans une population.

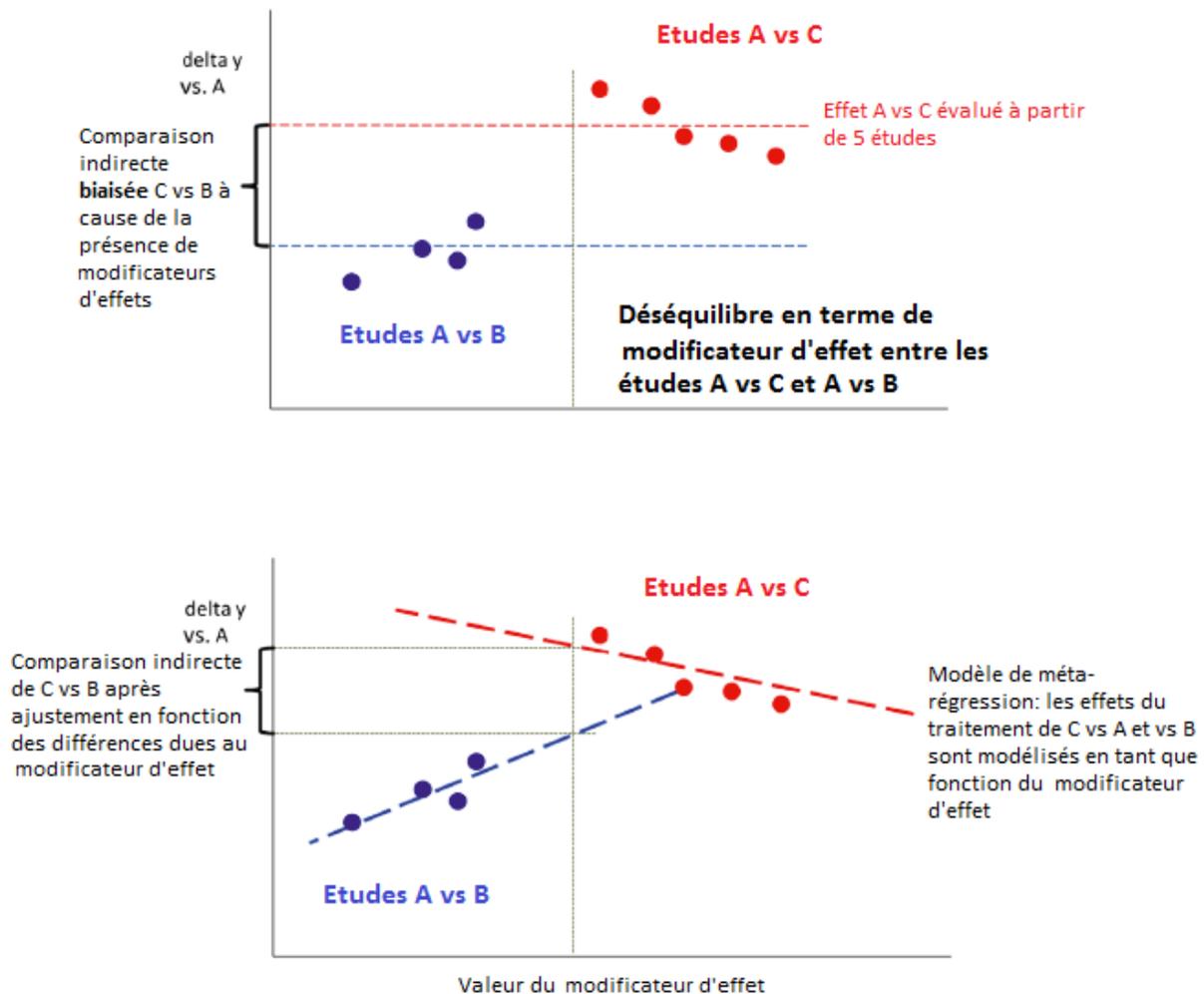


Figure 12 : Comparaison indirecte avec et sans ajustement avec le modificateur d'effet en utilisant un modèle de méta-régression⁶.

Remarque : avec cette approche, les effets du traitement de A vs C et A vs B dépendent de la valeur du modificateur d'effet (ex : âge des patients).

Pour exemple, dans une MTC évaluant différentes statines pour la protection cardiovasculaire⁴⁰, les auteurs ont utilisé une méta-régression pour définir s'il était approprié de combiner les résultats des populations traitées en prévention primaire, les résultats des populations traitées en prévention secondaire, les différentes statines, et les différentes doses de statines, en examinant si les effets de traitement des essais avec ces caractéristiques étaient différents.

Un défi avec la méta-régression est la puissance qui dépend du nombre d'études. Aussi, quand le nombre d'études disponibles dans le réseau est faible, la faisabilité d'une méta-régression peut se poser⁴¹.

En conclusion, l'objectif recherché avec la méta-régression est d'intégrer l'impact des modificateurs d'effet du traitement, puisque ceux-ci peuvent impacter la validité des comparaisons indirectes.

2 Validité des données des comparaisons indirectes

2.1 Les différentes hypothèses de validité

Une question importante au sujet des comparaisons indirectes et des comparaisons mixtes concerne leur validité, c'est-à-dire les paramètres qui peuvent remettre en questions leurs conclusions⁴² et les disqualifier. La méthode de comparaison indirecte est tellement sensible que les critères de validité sont exigeants. Les différentes hypothèses de validité sont illustrées dans la Figure 13 et détaillées dans la partie qui suit.

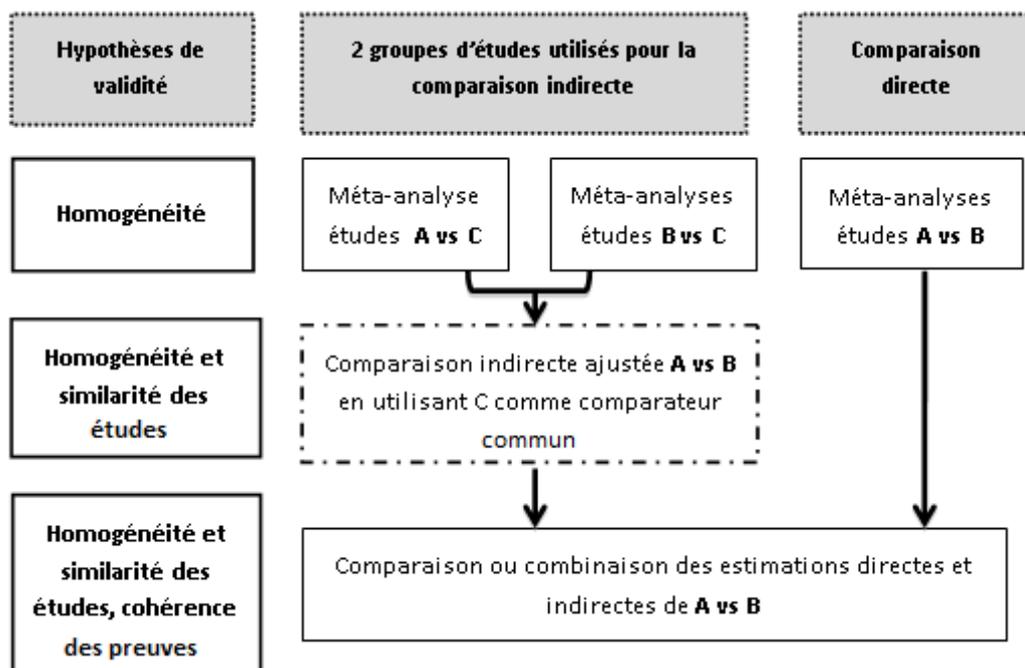


Figure 13 : Hypothèses sous-jacentes aux comparaisons indirectes et comparaisons mixtes⁴².

2.1.1 Hypothèse d'homogénéité (*homogeneity*) pour les données de la méta-analyse

Comme nous l'avons vu précédemment, lorsque plusieurs essais sont disponibles pour une comparaison donnée, les résultats de plusieurs essais peuvent être regroupés dans

les méta-analyses avant d'effectuer une comparaison indirecte. La plupart, sinon la totalité, des méta-analyses comprennent des études qui sont cliniquement et méthodologiquement diversifiées, et donc on s'attend à ce que d'une étude à l'autre l'hétérogénéité des effets du traitement soit présente.

2.1.1.1 Test d'hétérogénéité

Des études sont dites homogènes si elles sont similaires en ce qui concerne la population étudiée, pour le protocole d'investigation et pour la méthode d'analyse. Des études sont dites hétérogènes quand elles divergent entre elles pour ces critères. Dans les méta-analyses, il est important que les études rassemblées soient aussi homogènes que possible. Nous pouvons montrer la présence d'une hétérogénéité statistique entre différentes études par une analyse critique des méthodologies de recherche employées dans les études, par une représentation graphique des comparaisons des différentes études ou par le test d'hétérogénéité (déjà exposé brièvement au chapitre 1.1.3). Ce test vise à déterminer s'il existe des différences réelles entre les études (hétérogénéité), ou si la variation entre les résultats est compatible avec le hasard seul (homogénéité). Un test d'hétérogénéité examine l'hypothèse nulle que toutes les études évaluent le même effet. Cependant, le test est sensible au nombre d'essais inclus dans la méta-analyse. De plus, il permet de conclure en termes de significativité sur l'homogénéité, mais ne permet pas de quantifier l'hétérogénéité. Higgins *et al.*⁴³ ont développé en 2003 une mesure pour quantifier l'hétérogénéité et évaluer le degré d'incohérence dans les résultats. Cette quantité, nommée I^2 , décrit le pourcentage de variation totale dans toutes les études qui est dû à l'hétérogénéité plutôt qu'au hasard. I^2 est facile à calculer à partir des résultats de base obtenus à partir d'une méta-analyse typique :

$$I^2 = 100\% \times (Q - \text{ddl}) / Q$$
 où Q est la valeur du test de Cochran et ddl le degré de liberté.

Les valeurs négatives de I^2 sont rapportées à zéro, de sorte que la valeur de I^2 varie entre 0% et 100%. Une valeur de 0% indique une absence d'hétérogénéité observée, et des valeurs plus grandes témoignent d'une hétérogénéité croissante. Des seuils existent⁴⁴ pour l'interprétation de I^2 :

- 0% à 40% : hétérogénéité peu importante;

- 30% à 60% : hétérogénéité modérée ;
- 50% à 90% : hétérogénéité forte ;
- 75% à 100% : hétérogénéité considérable^f.

L'hétérogénéité peut témoigner d'une interaction entre une covariable et l'effet du traitement. Elle peut aussi provenir d'une forte variabilité aléatoire de l'effet sans qu'il soit possible de rattacher ces fluctuations à un ou des facteurs bien précis. Les analyses en sous-groupes permettent de rechercher s'il existe une interaction avec une ou plusieurs covariables. Cette recherche d'interaction a pour but de montrer que la taille de l'effet varie en fonction des valeurs prises par une ou plusieurs covariables, qu'on appellera donc modificateurs d'effet traitement.

L'hétérogénéité peut être considérée comme une nuisance que l'on cherchera à éliminer en adoptant une méthode adaptée (méthode à effet aléatoire). Ces techniques prennent en compte l'hétérogénéité sans chercher à l'expliquer. A l'opposé, l'hétérogénéité peut être considérée comme informative, témoignant de changement dans l'effet du traitement en fonction des circonstances de sa mesure (profil des patients ou utilisation du traitement).

2.1.1.2 Modèle à effets aléatoires

Si l'hétérogénéité observée ne s'explique pas par un modificateur d'effet, il est possible de recourir à un modèle à effets aléatoires. Ce modèle prend en compte une certaine variabilité aléatoire de l'effet traitement d'un essai à l'autre. L'estimation obtenue tiendra compte de cette variabilité, l'intervalle de confiance de l'effet combiné sera plus large qu'avec le modèle fixe, et la variance du vrai effet traitement sera estimée. Les intervalles de confiance sont plus larges car, en plus des fluctuations aléatoires, ils prennent en compte la variabilité du vrai effet traitement.

^f L'importance de la valeur observée de I^2 dépend de l'amplitude et du sens des effets et de la force de la preuve de l'hétérogénéité (par exemple, la valeur de P du test du chi-carré, ou un intervalle de confiance pour I^2).

L'utilisation d'un modèle à effets aléatoires s'accompagne d'un risque de méconnaître l'existence d'une interaction et d'arriver à une conclusion réductrice qui perd une partie de l'information apportée par les essais.

L'utilisation d'un modèle à effets aléatoires est généralement préconisé pour combiner des données⁴⁵. L'utilisation d'une méta-régression permet alors d'estimer les effets du traitement en fonction des différents niveaux de modificateurs d'effets.

2.1.1.3 Modèle à effets fixes

Un modèle à effets fixes suppose que le véritable effet du traitement soit commun à toutes les études comparant les mêmes traitements. Si une étude présente un effet qui diverge, cela ne peut être dû qu'au hasard. Cela implique qu'il n'y ait pas de modificateurs d'effets, ou qu'ils ont la même distribution dans toutes les études dans la méta-analyse. Compter uniquement sur un test statistique d'homogénéité pour soutenir l'utilisation du modèle à effets fixes à la place du modèle à effets aléatoires ne peut pas être considéré comme suffisant. Un argument en faveur de l'utilisation d'un modèle à effets fixes au lieu d'un modèle à effets aléatoires est d'être sûr de la similitude des études en fonction des modificateurs d'effets importants et la certitude, sur la base de l'expérience avec le domaine clinique considéré, que l'intervention est susceptible d'entraîner un effet fixes, indépendamment des populations étudiées.

2.1.1.4 Exemple

Reprenons l'exemple de comparaison indirecte ajustée (Tableau 7) ; les résultats de trois essais contrôlés contre placebo de la rispéridone et les résultats de neuf essais contrôlés de l'halopéridol contre placebo ont été rassemblés¹¹. L'hypothèse sous-jacente était que ces essais étaient suffisamment homogènes pour être quantitativement combinés (colonne I² dans le Tableau 7).

Comparaison	Nombre d'essais	Log OR (SE)	OR (IC 95%)	I ² %
Essais contrôlés contre placebo				
Risperidone vs placebo	3	-0,909 (0,218)	0,40 (0,26 ; 0,62)	37%
Halopéridol vs placebo	9	-1,707 (0,318)	0,18 (0,10 ; 0,34)	11%
Risperidone vs halopéridol				
Comparaison indirecte ajustée	3/9	0,798 (0,386)	2,22 (1,04 ; 4,72)	

Tableau 7 : Méta-analyse de rispéridone vs halopéridol dans le traitement de la schizophrénie, test d'homogénéité¹¹.

Les valeurs de I² sont inférieures à 40% et orientent donc bien vers une faible hétérogénéité.

2.1.2 Hypothèse de similarité (*similarity*) pour les comparaisons indirectes (ou « échangeabilité », « transposabilité »)

Comme nous l'avons vu précédemment, certaines caractéristiques des patients ou de l'étude, les « modificateurs de l'effet traitement » (« *effect modifiers* ») peuvent influencer l'effet traitement³⁹. Par exemple, si une intervention médicale ne fonctionne que pour les hommes et pas pour les femmes, un essai chez les hommes démontrera un effet positif du traitement par rapport au placebo, alors qu'un essai chez les femmes n'en démontrera pas. Le facteur sexe est alors un modificateur de l'effet du traitement.

S'il existe un déséquilibre dans les caractéristiques des patients ou de l'étude en termes de modificateur de l'effet relatif de traitement (par exemple le pourcentage de femmes et d'hommes dans notre cas) les comparaisons indirectes seront biaisées parce qu'elles ne respectent pas l'hypothèse de similitude (aussi appelée échangeabilité ou transposabilité). D'après cette hypothèse, dans les conditions des essais B, le traitement A aurait donné les mêmes résultats que ceux obtenus dans les essais A. Prenons l'exemple de la molécule Y versus la molécule X dans le traitement de la fibrillation auriculaire :

- dans l'essai y, la molécule Y est comparée à la warfarine. RR=0.80

- dans l'essai x, la molécule X est comparée à la warfarine. RR=0.95.

Si l'hypothèse d'échangeabilité est respectée, la molécule Y dans les conditions de l'essai x aurait donné un RR=0.80, et la molécule X dans les conditions de l'essai y aurait donné un RR=0.95.

2.1.2.1 Similitude clinique

Bien que la comparaison indirecte ajustée ou la méta-analyse en réseau conserve la randomisation, celle-ci ne tient pas sur l'ensemble des essais utilisés pour l'analyse car les patients ne sont pas répartis de façon aléatoire dans les différents essais. En conséquence, des différences systématiques dans la distribution des caractéristiques des patients dans les différents essais peuvent s'ensuivre. Pour qu'une comparaison indirecte ajustée soit valide, une hypothèse de similitude est nécessaire en termes de modificateurs de l'effet relatif de traitement. Cela signifie que les patients inclus doivent être suffisamment semblables dans les deux séries d'essais contrôlés par placebo, de sorte que l'effet relatif estimé par des études de A par rapport à C est généralisable aux patients dans les essais de B par rapport à C, et l'effet relatif estimé par des études de B par rapport à C est généralisable aux patients dans les essais de A par rapport à C (Figure 14).

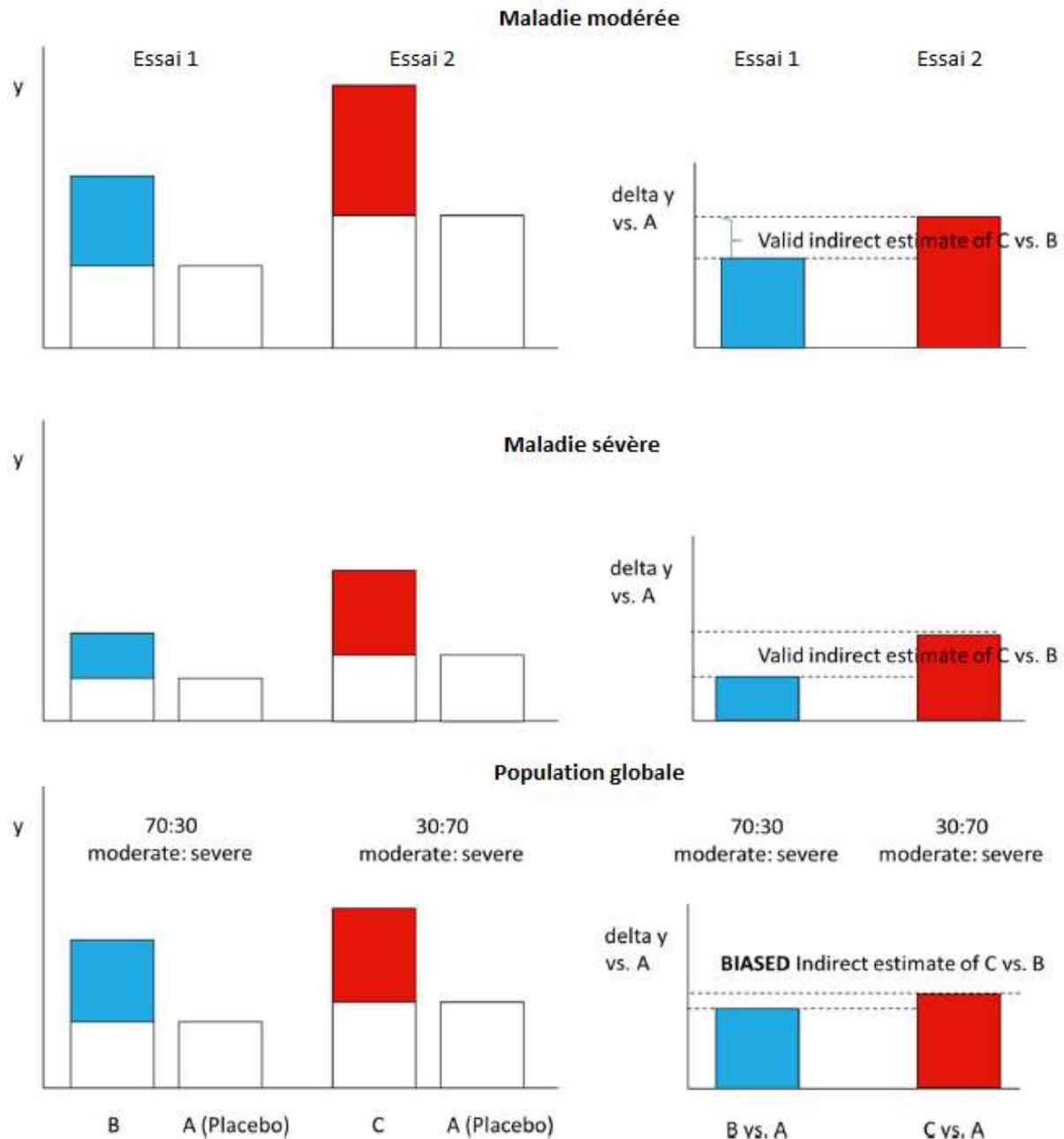


Figure 14 : Comparaison indirecte d'une étude A vs B et A vs C avec des proportions différentes de patients atteints de maladie modérée et sévère³.

La gravité de la maladie est un modificateur de l'effet (« effect modifier »). La comparaison indirecte pour le sous-groupe de maladie modérée est valide, ceci est également le cas pour le sous-groupe sévère. Par contre la comparaison indirecte de ces résultats pour l'ensemble de la population des deux études est biaisée, car la répartition du modificateur d'effet « gravité du traitement » est différente pour les études A vs B et A vs C.

« y » : effet traitement ; « valid indirect estimate » : estimation indirecte valide ; « biased indirect estimate » : estimation indirecte biaisée.

L'hypothèse de similarité est explorée à l'aide des analyses en sous-groupe et plus particulièrement à l'aide du test d'interaction. Ce test recherche si les différences observées

entre les tailles des effets dans les différents sous-groupes peuvent être expliquées par le seul hasard ou non. Lorsque le test est significatif il n'est plus raisonnable de conclure que les différences observées sont dues au seul fait du hasard. Il est alors possible de conclure que l'effet du traitement varie effectivement entre les sous-groupes. Le test d'interaction est similaire au test d'hétérogénéité en méta-analyse.

Il faudra donc compter sur les analyses en sous-groupe réalisées dans les essais pour identifier les modificateurs d'effets s'ils existent et, le cas échéant, vérifier si la répartition de ceux-ci est identique dans tous les essais à l'aide des tableaux des caractéristiques de base. Mais le diagnostic peut être rendu difficile par la possibilité que des variables pertinentes n'aient pas été analysées. En effet, la difficulté vient du fait que les modificateurs d'effets sont parfois mal connus. Par conséquent, il y a toujours un risque résiduel de biais, même si tous les modificateurs d'effets observés sont équilibrés dans les comparaisons directes. 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 traitement concomitants similaires) ou chez des patients dont les caractéristiques sont proches.

2.1.2.2 Similitude méthodologique

Les effets relatifs estimés dans les essais peuvent aussi être associés à la qualité méthodologique des essais. Des preuves empiriques indiquent que les essais de mauvaise qualité peuvent signaler des effets plus importants de traitement que des essais de bonne qualité, en particulier lorsque les résultats sont mesurés subjectivement. Il peut être prouvé mathématiquement que les résultats de la comparaison indirecte ajustée seront non biaisés lorsque les deux séries d'essais sont pareillement biaisés⁴⁶.

2.1.2.3 Respect de l'hypothèse de similarité clinique et méthodologique

Si l'hypothèse de similarité n'est pas remplie, les estimations de comparaisons indirectes ajustées ne sont pas valides et sont alors trompeuses. Par exemple, pour la comparaison indirecte ajustée de la rispéridone et l'halopéridol¹¹ (Tableau 5 et Tableau 7), les caractéristiques des patients, la dose de la durée de traitement étaient similaires entre les essais. Cependant, l'amélioration clinique a été définie de manière différente entre les

deux séries d'essais contrôlés par placebo (ce n'était pas la même échelle utilisée) ; les résultats de la comparaison indirecte ajustée doivent donc être interprétés avec prudence.

Finalement, les deux types de similitude, clinique (caractéristiques similaires des patients, des interventions, des contextes d'étude, de la durée de suivi, des critères de jugement) et méthodologique (risque de biais dans les études), doivent être prises en compte dans la comparaison indirecte ajustée (Figure 15).

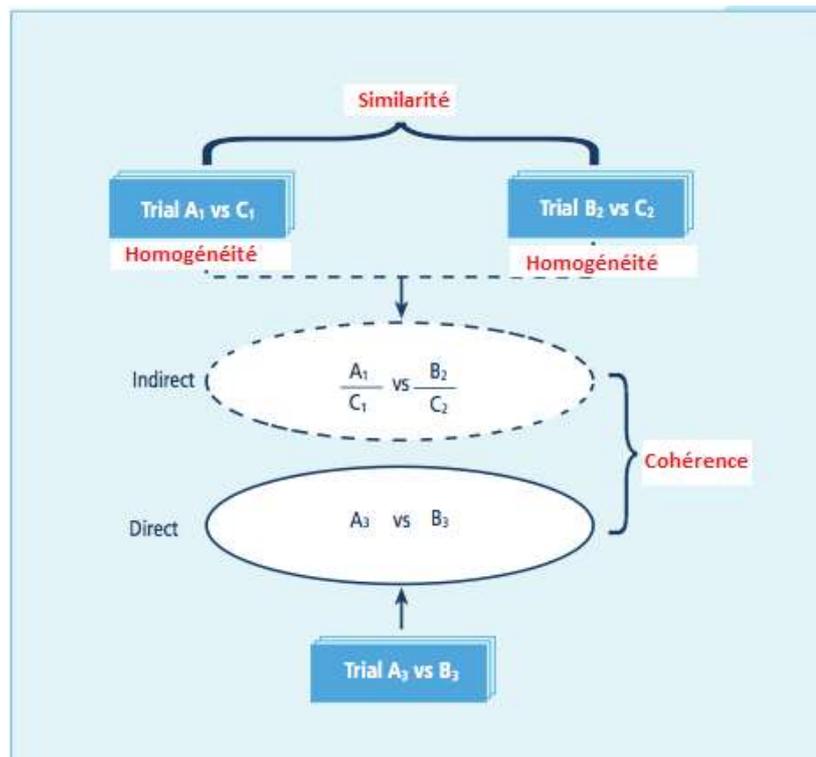


Figure 15 : Comparaison directe et indirecte des traitements A et B¹¹.

A ceci s'ajoute une autre hypothèse, l'hypothèse de cohérence.

2.1.3 Hypothèse de cohérence (consistency)

2.1.3.1 Hypothèse de cohérence pour combiner les résultats des comparaisons directes et indirectes

Sous l'hypothèse de cohérence (*consistency*), il n'y a pas de différence entre les résultats des comparaisons directes et ceux des comparaisons indirectes. En effet, lorsque les preuves directes et indirectes sont disponibles (comme c'est le cas avec un réseau constitué d'un boucle fermée), nous pouvons comparer l'estimation directe avec

l'estimation indirecte afin d'examiner si les deux estimations sont cohérentes (cette approche est la MTC, comme abordé précédemment).

S'il n'y a pas de différence dans la répartition des modificateurs d'effets traitement des différentes comparaisons directes qui forment la boucle, alors il n'y aura pas de différence entre l'estimation directe et indirecte pour chacune des comparaisons^{12,13} (Figure 16). Dans le cas où il n'y a pas d'incohérence il est intéressant de combiner les résultats des comparaisons directes et indirectes (« *Mixed treatment comparison* »). La combinaison des deux résultats permet une estimation plus précise de l'effet traitement⁴⁷.

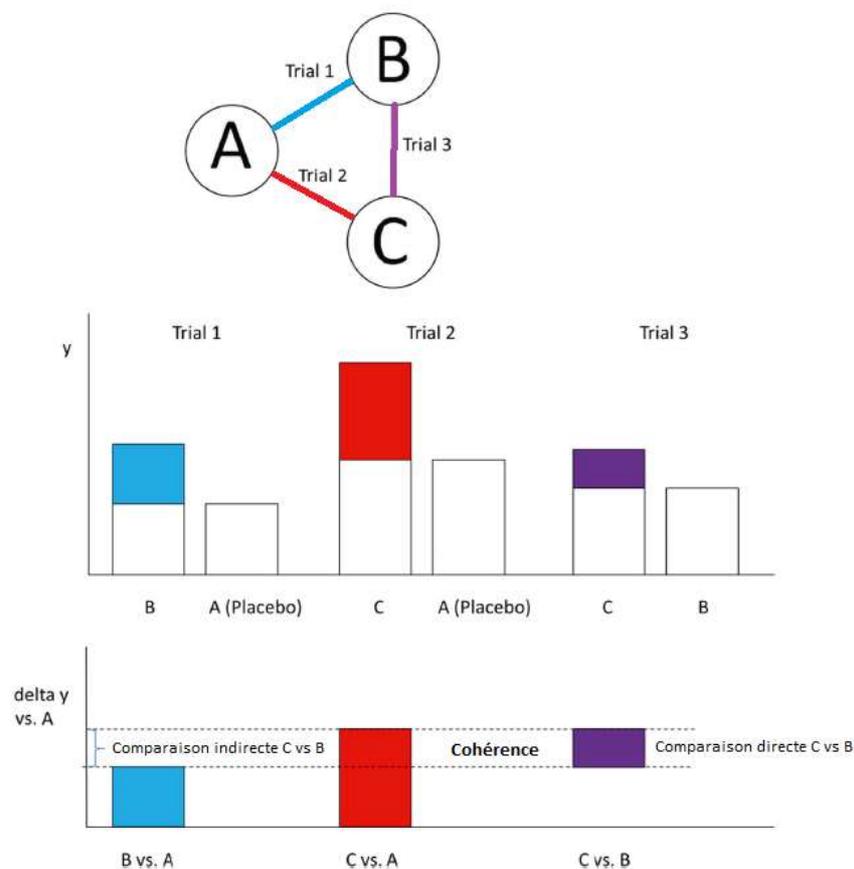


Figure 16 : Exemple d'un réseau en boucle fermée qui présente une cohérence entre la comparaison directe et indirecte⁶. « trial » = essai

Toutefois, s'il existe des différences dans la répartition des modificateurs d'effets des différentes comparaisons directes de la boucle, l'estimation directe est attendue être différente de l'estimation par comparaison indirecte correspondante, et la combinaison des deux résultats serait inappropriée (Figure 17). Par conséquent, il est important qu'en présence d'une boucle fermée, toutes les comparaisons directes soient comparées à la

comparaison indirecte correspondante pour la taille ou la distribution des modificateurs d'effets de traitement.

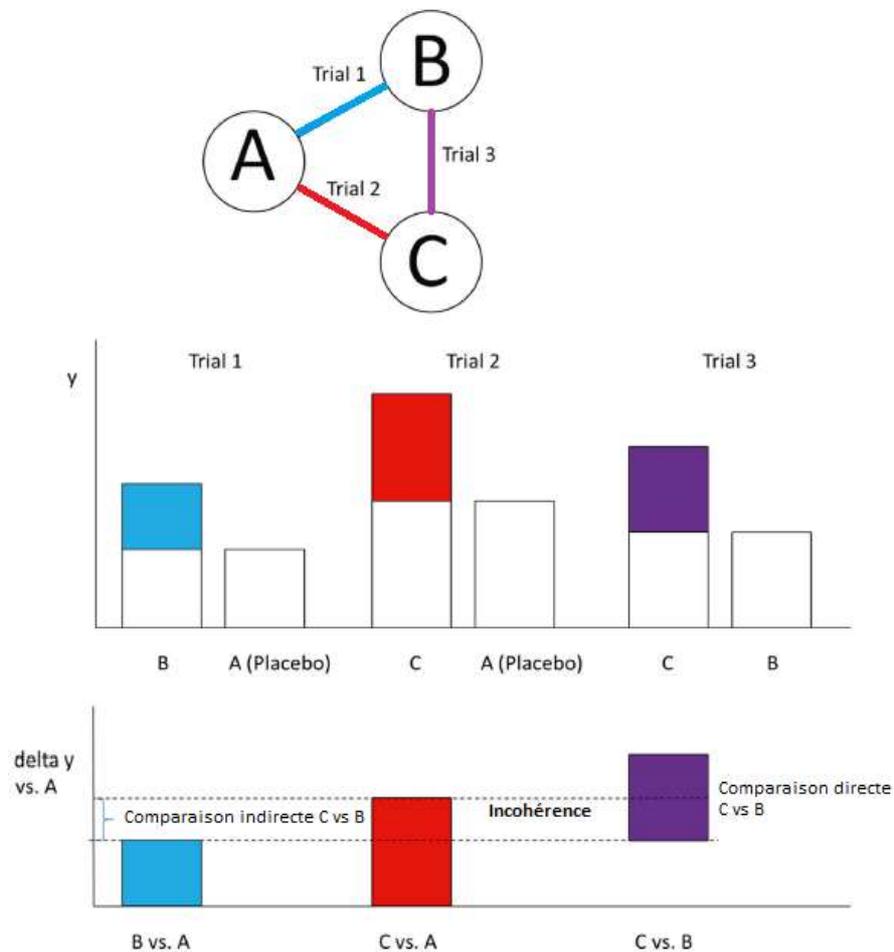


Figure 17 : Exemple d'un réseau en boucle fermée qui présente une incohérence entre les comparaisons directes et indirectes ⁶.
« trial » = essai

La publication d'Edward J.⁴⁸ donne une liste de facteurs pouvant entraîner une incohérence entre les résultats de comparaisons directes et ceux de comparaisons indirectes :

- Le hasard ;
- Une « véritable » diversité :
 - Des différences entre les participants inclus (par exemple les critères d'inclusion, le stade clinique, le risque de base) ;
 - Des différences dans les interventions (par exemple la dose, la durée d'administration, l'administration préalable de traitements [deuxième ligne de traitement]) ;

- Des différences de traitement de fond (qui ont pu changer à cause des évolutions des dernières années) ;
- Des différences dans la définition ou la mesure des résultats (s'il y a interaction avec l'effet du traitement) ;
- Des biais dans les comparaisons directes :
 - Le biais de publication ;
 - Sélections des résultats et des analyses ;
 - Taille de l'effet surestimé ;
 - Failles dans la gestion des perdus de vue, dans la gestion de la randomisation, de l'aveugle.
- Des biais dans les comparaisons indirectes :
 - Chacun des items ci-dessus peuvent affecter les résultats des comparaisons directes sur lesquels les comparaisons indirectes sont fondées.
 - Comparaison indirecte inappropriée.

Pour exemple, la comparaison indirecte ajustée de la rispéridone et l'halopéridol donne des résultats incohérents par rapport aux résultats de la comparaison directe (Tableau 8 et Figure 18). Bien que nous puissions quantitativement combiner les deux estimations, l'estimation de rapport commun (odds 1,23 ; IC 95%: [0.44 ; 3.45]) peut être invalide et trompeuse. Il est important de rechercher les causes possibles de divergence entre les preuves directes et indirectes. Des méthodes statistiques existent pour contrôler ce type d'incohérence, généralement appelé test d'incohérence^{49,50}.

Comparaison	Nombre d'essais	Log OR (SE)	OR (IC 95%)	I ² %
Essais contrôlés contre placebo				
Risperidone vs placebo	3	-0.909 (0.218)	0,40 (0,26 ; 0,62)	37%
Halopéridol vs placebo	9	-1,707 (0,318)	0,18 (0,10 ; 0,34)	11%
Risperidone vs halopéridol				
Comparaison directe	10	-0,262 (0,142)	0,77 (0,58 ; 1,02)	14%
Comparaison indirecte ajustée	3/9	0,798 (0,386)	2,22 (1,04 ; 4,72)	
Combinaison des comparaisons directes et indirectes	10+3/9	0,207 (0,527)	1,23 (0,44 ; 3,45)	85%

Tableau 8 : Méta-analyse de rispéridone vs halopéridol dans le traitement de la schizophrénie, évaluation de l'incohérence¹¹.

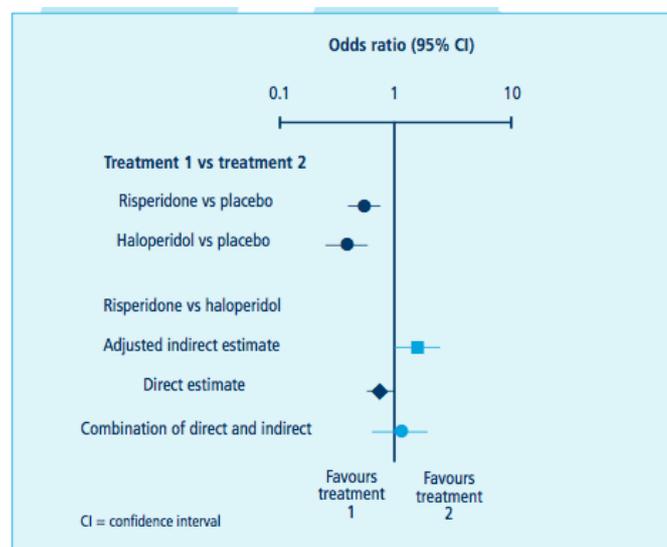


Figure 18 : Résultats des différentes méthodes de comparaisons de l'halopéridol et de la rispéridone¹¹.

Song *et al.*⁸ ont comparé les résultats obtenus par comparaison directe avec les résultats obtenus par comparaison indirecte ajustée. Parmi les quarante-quatre situations envisagées (comparaison de deux traitements pour une pathologie donnée), il n'y a eu incohérence entre les deux approches que dans trois cas seulement.

Une autre évaluation⁵¹, plus récente, examinant 112 exemples de comparaisons de preuves indirectes versus preuves directes, a constaté que les résultats étaient statistiquement incompatibles dans 14% des cas. Cette même évaluation a constaté que les comparaisons incluant un plus petit nombre d'essais et des mesures de résultats subjectives ont un plus grand risque d'incohérence. Par exemple, une méta-analyse portant sur l'efficacité analgésique du paracétamol (acétaminophène) en association avec de la codéine dans la douleur chirurgicale indique par comparaison directe que cette intervention est plus efficace que le paracétamol seul (différence moyenne dans le changement d'intensité de la douleur : 6,97 ; IC 95% [3,56 à 10,37]). Par contre, la comparaison indirecte ajustée n'a pas montré de différence significative entre le paracétamol associé à la codéine et le paracétamol seul (-1.16 ; IC à 95% [-6,95 à 4,64])⁵². Dans cet exemple, les éléments de preuve directe et indirecte ont été statistiquement et significativement incohérents (P = 0,02). Cette analyse répond à un critère subjectif (à savoir l'intensité de la douleur), et est donc plus susceptible de démontrer statistiquement une incohérence. L'incohérence peut s'expliquer ainsi : dans la comparaison directe les essais ont inclus plus de patients avec une intensité de la douleur plus basse au départ, et qui ont été plus sensibles à l'ajout de la codéine.

Lorsque la comparaison directe diffère de la comparaison indirecte ajustée, nous devrions normalement donner plus de crédibilité à la preuve d'essais comparatifs directs. Cependant, les données d'essais comparatifs directs ne sont pas toujours valables. Certaines études de cas ont indiqué que la comparaison indirecte ajustée peut fournir des résultats moins biaisés que des comparaisons directes sous certaines circonstances⁴⁶. En effet, les essais directs peuvent produire des estimations trompeuses notamment lorsque des conflits d'intérêt influent sur le choix des comparateurs utilisés ou sur les résultats rapportés. Par conséquent, les comparaisons indirectes peuvent parfois fournir des estimations avec un meilleur niveau de confiance.

2.1.3.2 Evaluation de l'incohérence lorsqu'il existe plusieurs chemins

Dans certains cas, il s'avère que des comparaisons indirectes ajustées pourraient être réalisées à l'aide de plusieurs « chemins » et non pas seulement par un seul comparateur commun. On entend par le terme « chemin », le comparateur commun des comparaisons de base sur lesquelles s'appuie l'extrapolation effectuée par la comparaison indirecte ajustée.

Dans le cas simple, la comparaison indirecte de A versus B s'effectue à l'aide du chemin $A \rightarrow \text{placebo} \rightarrow B$ puisque les deux comparaisons de bases utilisées le sont versus placebo. Mais 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 $A \rightarrow C \rightarrow B$. 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) (Figure 19).

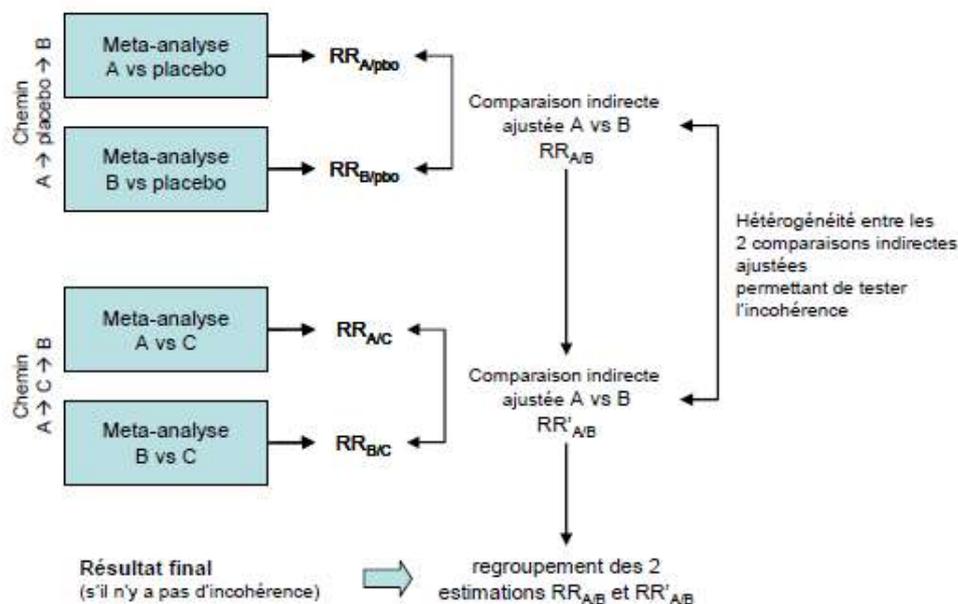


Figure 19 : Illustration du processus de comparaisons indirectes ajustées en cas d'existence de plusieurs chemins de comparaison de base⁴.

L'existence de ces deux résultats va permettre de tester s'il y a incohérence (*discrepancy*) entre ces deux chemins, c'est-à-dire si le résultat de la comparaison indirecte varie en fonction du chemin. 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, quel que 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.

Comme nous l'avons vu précédemment, la question de l'hétérogénéité est cruciale. En fait, les différents essais contrôlés randomisés inclus dans l'analyse sont réalisés à divers moments, au sein de populations distinctes, avec des dosages parfois différents, et cette hétérogénéité peut conduire à de possibles biais. Ces biais ne sont pas les seuls possibles, comme nous allons le voir dans le chapitre qui suit.

2.2 Le biais de publication

Le biais de publication définit la tendance de certains auteurs et éditeurs à publier prioritairement et rapidement des résultats positifs, alors que les résultats négatifs ou non conclusifs ne font pas, ou seulement tardivement, l'objet d'une publication³³.

A titre d'exemple, un biais de publication a été mis en évidence concernant la fluoxétine. Au moment de l'enregistrement de ce traitement par la FDA, seul un nombre limité d'études avaient été publiées. Plus tard, des études négatives qui existaient pourtant au moment de l'enregistrement ont fait l'objet de publications. Une méta-analyse a montré que l'effet de la fluoxétine était divisé par deux lorsque la totalité des études étaient prises en compte dans l'analyse⁵³⁻⁵⁵.

Dans un autre exemple de MTC de 12 traitements pour la dépression majeure, les auteurs ont choisi d'exclure les ECR contre placebo et de prendre en compte uniquement les essais cliniques directs impliquant des traitements actifs⁵⁶. Cependant, le biais de publication dans la littérature des antidépresseurs est bien connu^{57,58} et en excluant ainsi les essais contrôlés contre placebo, l'analyse perd la possibilité de bénéficier de l'exhaustivité des évidences⁽⁵⁶⁾. L'exclusion de certaines interventions éligibles, dans notre cas le placebo, peut non seulement diminuer la puissance statistique, mais cela peut également affecter la valeur des estimations ponctuelles des effets des comparaisons par paires, et cela peut ainsi impacter le classement relatif des traitements. C'est ainsi que donner une trop grande importance à l'interprétation du classement plutôt qu'aux comparaisons relatives peut être trompeur pour les évaluateurs, les cliniciens et les patients. Quand une méta-analyse en

réseau des antidépresseurs de seconde génération a été effectuée par la suite et a inclus les essais contrôlés contre placebo, en se fondant uniquement sur les essais utilisant la même échelle de dépression, les auteurs sont parvenus à une interprétation différente de la MTC précédente^{60,61}.

L'influence d'un biais de publication devra donc être systématiquement envisagée dans des études de sensibilité lors des comparaisons indirectes.

Aujourd'hui, les industriels sont contraints, au niveau européen, de mettre en place une base de données libre d'accès référencant les essais cliniques réalisés ou en cours avec les publications résultantes de ces études⁶. Cela devrait diminuer l'influence du biais de publication à l'avenir. Par ailleurs, ce biais n'est pas aussi problématique pour les anciens produits. En effet, lorsque le produit est bien introduit sur le marché, les études négatives finissent par être publiées. Elles ont un impact moindre, les cliniciens ayant chacun pu se faire une opinion de terrain. De plus, les anciens produits ont généralement été utilisés comme référence lors de nouveaux essais, ce qui est susceptible de contre balancer les premiers résultats.

3 Guide de lecture critique

Nous venons d'étudier les différentes méthodes statistiques de comparaisons indirectes ainsi que leurs critères de validité. Malgré le grand potentiel de ces méthodes pour aider à les décideurs et les cliniciens, il y a un besoin critique de formation, notamment sur la méta-analyse en réseau, et de moyens objectifs pour évaluer la qualité des réseaux de méta-analyses afin d'éclairer la prise de décision. C'est pourquoi, de nombreux questionnaires (« *user guide* ») ont été développés dans la littérature afin d'accompagner les utilisateurs de comparaisons indirectes, sans que cela ne requiert de connaissances approfondies en modélisation et en statistiques. L'objectif est d'aider le lecteur à évaluer la qualité et la pertinence de la méta-analyse en réseau. Le guide de lecture critique que nous proposons, qui synthétise différents questionnaires d'évaluation choisis dans la littérature^{4,6,48,62}, comporte différents items :

⁶ Règlement (UE) n°536/2014 du Parlement européen et du Conseil du 16 avril 2014 relatifs aux essais cliniques de médicaments à usage humain et abrogeant la directive 2001/20/CE.

1. Pertinence globale pour le décideur
2. Analyse de la recherche et de la sélection des essais inclus
3. Vérification des hypothèses d'homogénéité et de similarité
4. Analyse de la méthode statistique mise en œuvre
5. Interprétation des résultats

3.1 Pertinence globale pour le décideur

On utilise le terme « pertinence » pour qualifier dans quelle mesure la comparaison indirecte a de l'intérêt pour l'évaluateur en fonction de sa problématique, au-delà de toutes considérations méthodologiques.

3.1.1 Le choix de la population est-il pertinent ?

Cette question porte sur les populations des ECR qui constituent la base de la méta-analyse. La question est de savoir si elles correspondent à la population d'intérêt pour le décideur. Il s'agit de prendre en compte : la maladie d'intérêt, ses différents stades, la gravité, les comorbidités, les antécédents de traitement, l'âge, le sexe, et peut-être d'autres caractéristiques démographiques. Typiquement, les ECR inclus dans l'analyse sont identifiés au moyen d'une recherche documentaire systématique avec des études pertinentes en termes de populations prédéfinies par les critères de sélection de l'étude. Si ces critères sont signalés, c'est un bon point de départ pour juger de la pertinence en termes de population. Ainsi, des tableaux de données avec des critères d'inclusion/exclusions et les caractéristiques des patients pour chaque étude fournissent des informations précieuses pour juger de la pertinence. A noter que des critères d'éligibilité larges peuvent permettre d'améliorer la généralisation des résultats, mais à l'inverse cela augmente dangereusement l'hétérogénéité si les participants sont trop différents.

3.1.2 Des traitements pertinents sont-ils manquants?

Cette question revient à savoir si l'intervention(s) inclus dans la méta-analyse en réseau correspond à celle(s) d'intérêt pour le décideur et si tous les comparateurs pertinents ont été considérés. Les traitements qui seront comparés dans la méta-analyse en réseau peuvent être limités à tous les médicaments d'une classe, mais peuvent aussi inclure des

médicaments de différentes classes concurrentes et, dans certains cas, d'autres interventions médicales. Quelle que soit la portée des interventions, une justification claire du choix doit être décrite. Il est généralement préférable d'inclure toutes interventions, d'autant plus que les données sur des interventions non optimales ou abandonnées peuvent encore offrir des preuves indirectes pour d'autres comparaisons⁶³.

3.1.3 Les effets de traitements pris en compte sont-ils pertinents ?

L'accent est mis sur les résultats qui sont directement significatifs pour le patient ou le système de santé («*clinical endpoint*») tels que les événements cardiovasculaires (par exemple, infarctus du myocarde et accident vasculaire cérébral), alors que les critères de substitution («*intermediate endpoint*») tels que la modification des paramètres biologiques (par exemple, le taux de cholestérol), ont moins de poids. La durée de leur évaluation a aussi son intérêt. Par exemple, une méta-analyse en réseau qui inclut des ECR avec un suivi plus long, peut être plus pertinente pour éclairer les choix de traitement pour une maladie chronique, plutôt qu'une méta-analyse en réseau limitée à des études avec un suivi de courte durée (si un suivi est relié à l'effet du traitement).

3.1.4 Le contexte (les paramètres et les conditions) est-il applicable?

L'année où les études incluses dans la méta-analyse ont été effectuées peut avoir son importance puisque la stratégie de traitement peut avoir changé au fil du temps. Parmi les questions à se poser, on sera particulièrement attentif à celles-ci : la définition de la pathologie a-t-elle évolué au cours du temps ? Les populations de patients sont-elles susceptibles d'avoir évolué sur le plan du phénotype ou de la gravité entre les moments où ont été réalisées les différentes études ? Les modalités générales de prise en charge des patients ont-elles évolué au cours du temps ? Etc.

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

Cette étape s'intéresse au processus de constitution de la liste des essais inclus dans les méta-analyses nécessaires aux calculs de comparaisons indirectes. Différents points sont à vérifier.

3.2.1 S'assurer de l'exhaustivité de la recherche

Comme nous l'avons abordé précédemment, les comparaisons indirectes sont sensibles, comme les méta-analyses classiques, au problème du biais de publication⁶⁴. Aussi, il convient de se poser les questions suivantes : un processus de recherche exhaustive et reproductible des essais de type méta-analyse a-t-il été suivi ? S'agit-il d'essais sélectionnés pour donner les résultats de comparaisons indirectes souhaités ou s'agit-il de tous les essais réalisés dans le domaine¹⁷ ? Pour répondre à ces questions, différents points devraient être vérifiés :

- a) Les critères d'inclusions et de non inclusion des essais doivent être clairement énoncés.
- b) Les bases de données consultées doivent être précisées.
- c) La stratégie de recherche doit cibler toutes les interventions d'intérêt.

Si les critères ci-dessus sont remplis, cela reflète les efforts faits pour inclure tous les ECR publiés pertinents et disponibles. Un organigramme qui illustre la façon dont les essais ont été sélectionnés peut être utile. Une liste des principales caractéristiques des patients et des études doit être fournie sous forme de tableau. Concernant ce sujet, les rédacteurs de PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) donnent des recommandations précises sur la façon de présenter les résultats de revues systématiques⁶⁵.

Une étape supplémentaire pour identifier les études manquantes consiste à faire une recherche dans des bases de données de déclarations d'essais cliniques, comme <https://clinicaltrials.gov>. Il est aussi possible de demander directement au laboratoire pharmaceutique commercialisant le médicament l'ensemble des essais cliniques concernant le produit, comme le fait souvent Andrea Cipriani, chercheur en psychiatrie à l'Université d'Oxford au Royaume-Unis, pour réaliser ses méta-analyses en réseau⁶⁶.

3.2.2 S'assurer de la qualité des études incluses

Pour répondre à cette question, le rapport devrait fournir les renseignements concernant les caractéristiques clés de chaque ECR, comme la méthode de randomisation, les perdus de vue etc. Le plus souvent, le rapport fournit une vue d'ensemble des biais de

chaque essai clinique à l'aide d'une liste de contrôles spécifiques de validité, tel que la grille de la Cochrane Collaboration⁶⁷ ou le questionnaire de Jadad⁶⁸.

3.3 Vérification des hypothèses d'homogénéité et de similarité

3.3.1 Absence d'hétérogénéité statistique entre les études d'une même comparaison

Les essais étudiés doivent présenter les mêmes modalités de traitements (doses, régime d'administration) car la présence d'une hétérogénéité des résultats statistiques cliniquement significative entre les essais de B versus A ou de C versus A est susceptible de remettre en question l'hypothèse de similarité (Figure 20 et Figure 21).

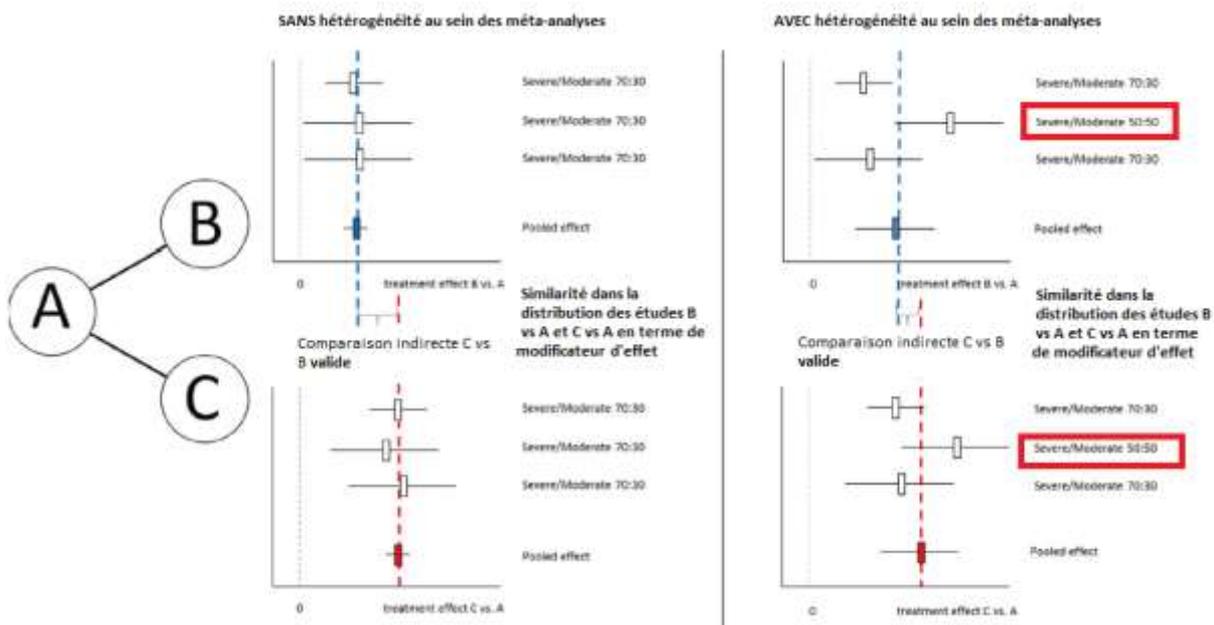


Figure 20 : Exemple d'hétérogénéité au sein des méta-analyses qui n'entraîne pas de remise en cause de l'hypothèse de similarité⁶.

Bien qu'il y ait une hétérogénéité entre les différentes études incluses dans la méta-analyse (schéma de droite), la comparaison indirecte reste valide car il n'y a pas de déséquilibre dans la répartition du modificateur d'effet « gravité de la maladie » entre les comparaisons B vs A et C vs A. L'hypothèse de similarité est respectée.

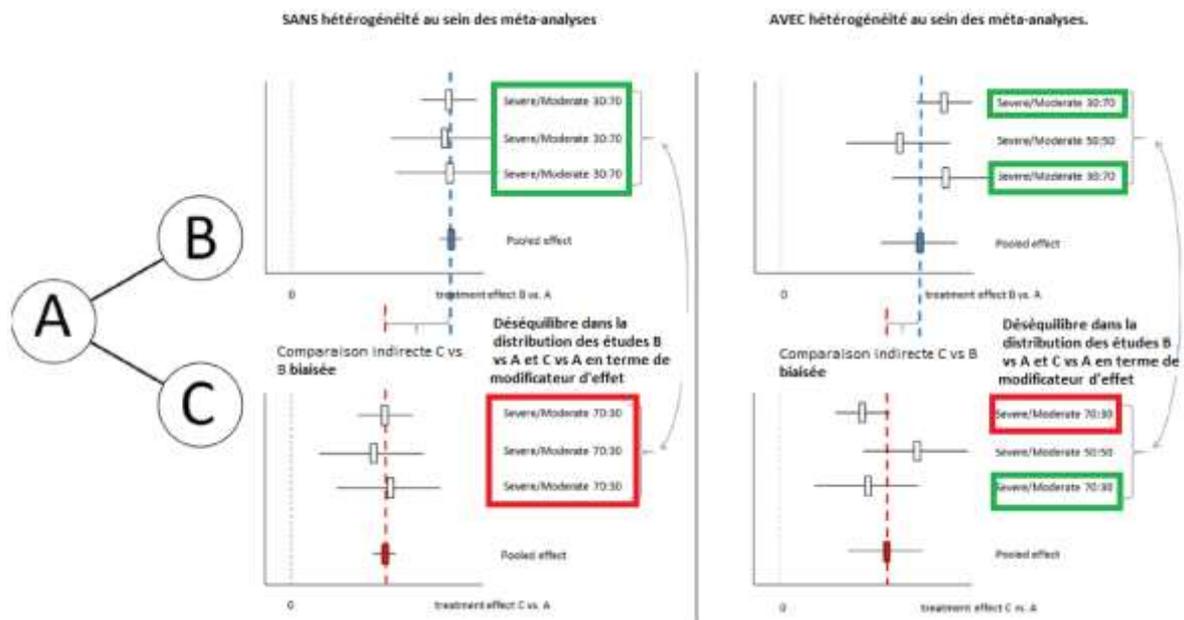


Figure 21 : Exemple d'hétérogénéité au sein des méta-analyses qui entraîne une remise en cause de l'hypothèse de similarité⁶.

La comparaison indirecte est biaisée car il y a un déséquilibre dans la répartition du modificateur d'effet « gravité de la maladie » entre les comparaisons B vs A et C vs A, indépendamment de la condition d'homogénéité entre les études incluses dans les méta-analyses. L'hypothèse de similarité n'est donc pas respectée.

3.3.2 Vérifier la présence de modificateurs d'effets de traitement pour l'hypothèse de similarité

Il convient de vérifier que les analyses en sous-groupe réalisées dans les essais ne révèlent pas de modificateur d'effet. Cela est essentiel pour juger s'il existe des différences entre les essais qui pourraient agir comme modificateurs d'effets, provoquant ainsi un biais dans l'analyse. Par exemple, des différences dans l'âge des patients ou les antécédents de traitement peuvent constituer des modificateurs d'effets. En outre, les zones géographiques où les études ont été réalisées peuvent inclure des différences supplémentaires entre les populations de patients.

3.4 Analyse de la méthode statistique mise en œuvre

3.4.1 Éviter les comparaisons indirectes naïves

Le premier point consiste à vérifier que la méthode employée est adaptée à la comparaison indirecte. Il s'agit d'exclure l'utilisation de méthodes inappropriées. Ainsi, il

faut rejeter une approche de comparaison naïve des estimations ponctuelles ainsi qu'une approche où les différents bras actifs des essais sont comparés sans tenir compte des bras contrôles, rompant la randomisation (comparaison naïve des bras actifs). Les autres méthodes sont valides.

3.4.2 Analyse de la méthode statistique mise en œuvre

Ce point représente la partie la plus technique. Le but est de vérifier que la mise en œuvre de la méthode statistique utilisée est satisfaisante. L'article doit fournir un aperçu global des méthodes statistiques utilisées, y compris la justification du choix des critères d'évaluation, des estimations des effets relatifs, et du choix d'un modèle à effets fixes ou aléatoires. Si les analyses ont été effectuées dans un cadre bayésien, le choix de la distribution *a priori* pour les paramètres du modèle devrait être défini. Une description des différentes analyses de sensibilités devrait également être fournie (il s'agit de refaire les calculs en faisant varier certains paramètres pour observer leurs impacts sur les résultats).

En pratique, le recours à une expertise statistique est souvent nécessaire pour juger de l'acceptabilité technique du résultat de certaines méthodes dont la mise en place est délicate.

3.4.3 Vérifier 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, c'est-à-dire 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 trouvée 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 car il concerne les résultats et non pas la méthode.

Aussi, la question qu'il faudra se poser est la suivante : Les estimations des comparaisons directes (quand elles existent) sont-elles compatibles avec celles des comparaisons indirectes ou mixtes ? Cette concordance est évaluée à l'aide d'un test d'hétérogénéité des deux types d'estimations.

En cas de disponibilité d'essais de comparaisons directes et en absence d'incohérence, la meilleure estimation possible est celle qui englobe l'ensemble de l'information disponible, c'est-à-dire le résultat de la comparaison directe et celui de la comparaison indirecte (« *mixed treatment comparison* »).

3.4.4 Présentations des résultats

Pour aider à comprendre les résultats d'une méta-analyse en réseau, une représentation graphique du réseau de preuves avec un aperçu des ECR inclus (et des comparaisons indirectes disponibles) est nécessaire. Comprendre la géométrie du réseau (nœuds et liens) permettra au lecteur d'examiner la situation dans son ensemble et de voir ce qui est comparé puisque des analyses comparatives de traitement multiples sont possibles lorsqu'il existe une connexion dans le réseau⁶⁹.

De plus, un tableau récapitulatif dans lequel les études sont présentées en lignes, les interventions en colonnes, et les résultats observés avec chaque intervention dans chaque étude dans les cellules, peut se révéler instructif. Plus précisément, la présentation des résultats des études individuelles permet aux examinateurs de comparer ceux-ci avec les résultats de la méta-analyse en réseau. Cela facilitera également la réplication de l'analyse, si on le souhaite. Enfin, cela permet de juger si l'hypothèse de la cohérence entre les preuves directes et indirectes est respectée. Tout aussi important, pour chaque effet du traitement qui est estimé, les mesures de l'incertitude doivent être présentées (c'est à dire des intervalles de crédibilité à 95% (ICr)).

3.5 Interprétation des résultats

Il est important d'évaluer la qualité de la MTC afin que le lecteur puisse déterminer si les conclusions fournies sont solides en termes de niveau de preuve. Les aspects suivants sont des indices de preuve de haute qualité dans une MTC :

- Les études individuelles sont à faible risque de biais ;
- Le biais de publication est peu probable ;
- La taille de l'échantillon est grande et les intervalles de confiance sont en conséquence étroits ;
- La preuve comprend des comparaisons directes ;

- Les résultats sont cohérents entre les comparaisons directes individuelles et les résultats des comparaisons indirectes.

Si tous ces éléments sont présents et que les différences en terme d'effet sont importantes, alors une grande confiance pourra être accordée aux résultats. Notre jugement sera plus nuancé en fonction des éléments manquant pour fonder la preuve.

Par ailleurs, comme nous l'avons vu précédemment, la validité du résultat d'une méta-analyse en réseau dépend de la convergence des chaînes de Markov et le calcul nécessite de déterminer un nombre non négligeable de paramètres pouvant influencer le résultat. Par conséquent, il est important que soient rapportées dans la publication, comme recommandé⁷⁰⁻⁷², les informations permettant de vérifier les calculs, à savoir : les modèles utilisés, les distributions *a priori* spécifiées (avec leur justification), les paramètres de calcul tels que le nombre d'itérations, la période de rodage, les valeurs initiales utilisées.

De plus, comme les méthodes MTC deviennent populaires, les cliniciens peuvent trouver plusieurs méta-analyses en réseau abordant la même question et, comme dans le cas des antidépresseurs de seconde génération que nous avons déjà mentionné, ces méta-analyses peuvent parfois aboutir à différentes conclusions^{56,73}. Des différences dans les critères d'inclusions et les échelles pour mesurer l'effet peuvent expliquer ces différences. Pour illustrer ce point, il n'y a pas moins de quinze MTC publiées et d'évaluations des technologies de santé sur l'efficacité relative des agents biologiques dans le traitement de la polyarthrite rhumatoïde. En consultant les rapports publiés, on peut observer des différences importantes quant aux nombres et aux séries d'essais inclus, à l'évaluation de l'hétérogénéité et du risque de biais, et aux critères d'inclusion ou d'exclusion⁷⁴. Compte tenu de l'importance de chacun de ces points et des écarts retrouvés entre les résultats des différentes MTC, une incertitude existe quant au classement des différents agents biologiques en termes d'efficacité.

En outre, l'identification du traitement le meilleur ou le plus approprié ne peut être effectué sur la base de la seule efficacité. Pour informer les prises de décision en matière de soins de santé concernant les recommandations de traitements et les politiques de remboursement, les conclusions d'une méta-analyse en réseau sur l'efficacité relative des traitements doivent être interprétées à la lumière des autres preuves disponibles et des autres caractéristiques des interventions concurrentes, telles que la sécurité et les modalités d'emploi.

Finalement, la méthode de Bucher est facile à mettre en œuvre et ses résultats sont facilement vérifiables. Cependant cette méthode n'est pas utilisable dans de nombreux cas puisqu'elle ne permet pas les comparaisons multiples sans un risque d'inflation du risque de première espèce et donc de conclure à tort à une différence entre les effets traitements. La méta-analyse en réseau permet-elle les comparaisons multiples, mais sa mise en œuvre est plus délicate et les résultats sont plus difficilement vérifiables.

Au vu des tous ces éléments et constats, nous allons maintenant nous intéresser à la place actuelle qu'occupent les comparaisons indirectes dans l'évaluation médico-technique des médicaments en vue de leur accès au remboursement.

PARTIE 2

Place des comparaisons indirectes dans l'évaluation médico-technique des médicaments en vue de leur accès au remboursement.

1 La procédure d'accès au remboursement du médicament en France^{75,76}

Après les premières phases de développement d'un candidat médicament qui ont permis d'identifier le mécanisme d'action et de définir la dose offrant le meilleur rapport bénéfice/risque, des études de phases III sont réalisées sur de plus larges populations de patients, dans le but de vérifier son efficacité et sa tolérance. Pour être mis sur le marché, le médicament devra bénéficier d'un rapport bénéfice/risque favorable, ce qui lui permettra d'obtenir une Autorisation de Mise sur le Marché (AMM). Pour être pris en charge par l'assurance maladie, il devra obtenir son inscription sur les listes des spécialités remboursables aux assurés sociaux et/ou aux collectivités, impliquant la fixation de son taux de remboursement et de son prix. Ces différentes étapes sont illustrées sur la figure qui suit (Figure 22).

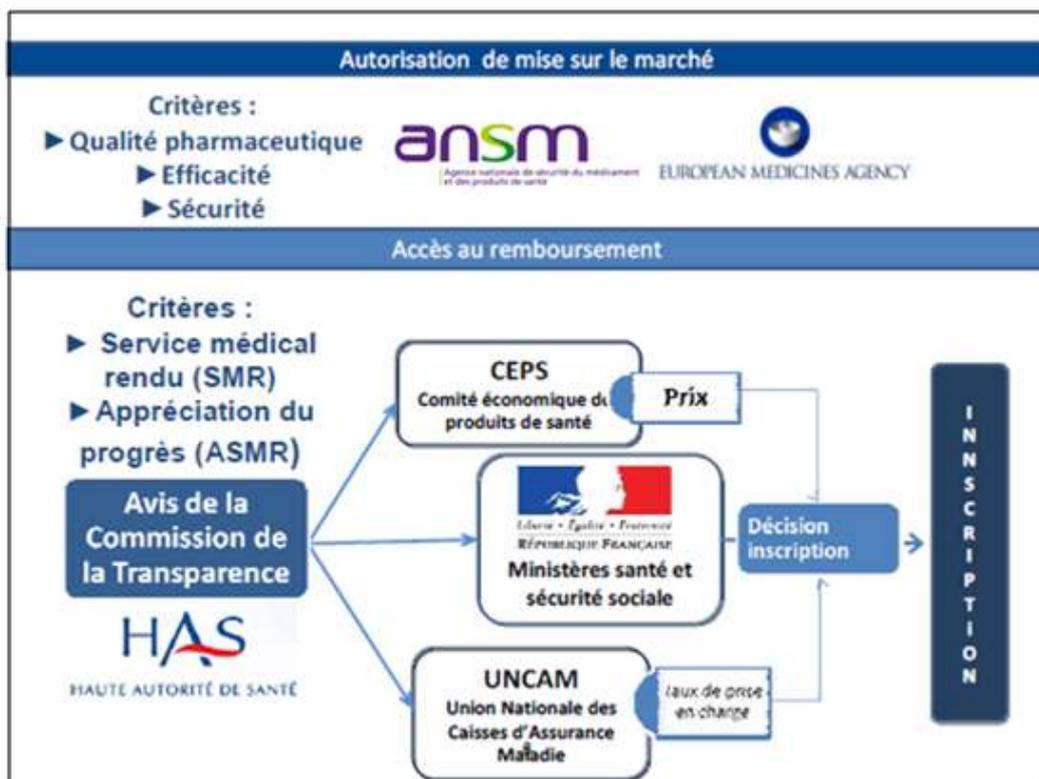


Figure 22 : Le circuit du remboursement des médicaments en France.

Nous allons détailler brièvement l'étape de l'obtention de l'AMM, puis celle de l'accès au remboursement de manière plus approfondie.

1.1 L'Autorisation de Mise sur le Marché

Pour qu'un nouveau médicament soit mis sur le marché dans une indication donnée, il faut que son dossier de demande d'enregistrement apporte la preuve que son efficacité est supérieure à ses risques connus dans même indication (rapport bénéfice/risque favorable)⁷⁵. On remarquera qu'il n'y a, à ce stade, pas de comparaison à l'existant ; il n'est donc pas nécessaire que le nouveau médicament soit plus efficace que les traitements de référence.

Des commissions rassemblant des experts, ont pour rôle de rendre un avis scientifique aux décideurs, à savoir la Commission Européenne ou les Directeurs des agences nationales de santé. Il s'agit soit du « Committee for Medical Products of Human Use » (CHMP) de l'Agence Européenne du Médicament (EMA) située à Londres (Angleterre), si l'évaluation est faite au niveau européen, soit de la Commission d'évaluation initiale du rapport bénéfice risque des produits de santé de l'Agence Nationale de Sécurité du Médicament (ANSM, ex-AFSSAPS) située à Saint-Denis (France), dans le cas d'une évaluation au niveau nationale.

Une fois que le médicament a obtenu son AMM, il est évalué par la Haute Autorité de Santé (HAS) en vue de son remboursement si le laboratoire souhaite que sa spécialité soit prise en charge.

1.2 L'inscription sur la liste des spécialités remboursables aux assurés sociaux et des spécialités agréées à l'usage des collectivités et divers services publics

En France, hormis les médicaments d'automédication disponibles en pharmacie d'officine en « non remboursables », il ne suffit pas qu'un nouveau médicament ait obtenu une AMM pour être remboursés. La réglementation française impose que les médicaments ne soient pris en charge que s'ils sont inscrits sur une liste positive^h établie après avis d'une commission scientifique indépendante, la Commission de la transparence (CT). La

^h Il y a deux listes : l'une pour les médicaments inscrits au remboursement dans les pharmacies de ville (liste Sécurité sociale) et l'autre pour les médicaments pris en charge à l'hôpital (liste d'agrément aux collectivités).

réglementation stipule aussi que cet avis, lorsqu'il est favorable à l'inscription, donne des éléments qui contribuent à la fixation du prix du médicament. Ainsi, il revient à la CT de la Haute Autorité de Santé (HAS), de donner aux ministres de la Santé et de la Sécurité Sociale un avis sur le bien-fondé de la prise en charge du médicament par la solidarité nationale, d'éclairer le Comité économique des produits de santé (CEPS) pour fixer le prix, et de recommander à l'Union Nationale des Caisses d'Assurance Maladie (UNCAM) le taux de prise en charge.

1.2.1 La Commission de la transparence (CT) : missions et composition

1.2.1.1 Composition de la CT et du Service Evaluation des Médicaments (SEM)

La CT est une commission consultative qui comprend 26 membres, tous compétents dans le domaine de l'évaluation des médicaments. Elle est ainsi composée de médecins de diverses spécialités cliniques et médecine générale, pharmaciens, biostatisticiens, épidémiologistes, médecins de santé publique, pharmacologues. Les membres de la Commission sont nommés par décision du Collège de la HAS pour une période de 3 ans renouvelable deux fois. Ils se réunissent toutes les deux semaines, le mercredi.

La CT s'appuie, pour rendre ses avis, sur le Service Evaluation des Médicaments (SEM). Il est composé d'une quinzaine de chefs de projet, médecins ou pharmaciens, formés à la méthodologie des essais cliniques, assistés d'un secrétariat. Le SEM assure l'instruction interne des demandes, et prépare un projet d'avis à partir de l'analyse critique du dossier déposé par le laboratoire et des données disponibles, en vue de la présentation en séance de la Commission. Il veille à la qualité des documents préparés, à la conduite et à la régularité des procédures ainsi qu'au respect des délais.

1.2.1.2 Missions de la CT

Les missions de la Commission sont précisées par les articles R 163-2 à 21 du Code de la sécurité sociale ⁷⁷. La Commission a notamment pour missions de :

- rendre un avis aux ministres chargés de la santé et de la sécurité sociale sur les demandes d'inscription et de renouvellement d'inscription des médicaments ;

- réévaluer, à son initiative ou à la saisine d'un tiers, à tout moment, le service médical rendu des médicaments inscrits sur les listes ;
- établir, pour diffusion, des documents d'information sur les produits de santé (par exemple : synthèses d'avis, fiches de bon usage des médicaments, etc.) ;
- donner un avis sur les documents à propos desquels elle est saisie ;
- établir des fiches d'information thérapeutique annexées aux arrêtés d'inscription des médicaments particulièrement coûteux et d'indications précises.

L'évaluation est réalisée sur la base d'une analyse critique des données scientifiques disponibles selon les bases de la médecine fondée sur les preuves (*Evidence Based Medicine*) et l'avis d'experts, dans l'indication et aux posologies de l'AMM. Les documents d'information ainsi que les avis de la CT sont publics, ils sont publiés sur le site Internet de la Haute Autorité de Santé (www.has-sante.fr).

Ainsi, en France la CT est une instance indépendante clé pour l'admission au remboursement des médicaments par l'assurance maladie et pour l'accès des patients à la meilleure qualité des soins possibles.

1.2.2 Le Service médical rendu (SMR) et l'Amélioration du Service médical rendu (ASMR)

Le Code de la sécurité socialeⁱ précise les critères sur lesquels la Commission doit s'appuyer pour rendre ses avis. En particulier, il définit les critères du service médical rendu (SMR)^j et de l'amélioration de ce service médical rendu (ASMR). La CT doit également estimer sur la population cible du médicament. Ces composantes de l'avis de la Commission, qui en constitueront les principales conclusions, répondent aux 5 questions posées par le décideur, qui sont :

- ce médicament doit-il être pris en charge par la solidarité nationale, à l'hôpital ou en ville ? Oui/Non
- à quel taux doit-on fixer la participation de l'assuré ?

ⁱ Décret 2004-1398 du 23 décembre 2004 (JO du 26/12/2004)/Articles R. 163-1 à 21 du Code de la sécurité sociale.

^j Article R. 163-3 du Code de la sécurité sociale.

- quelle est sa valeur ajoutée médicale, c'est-à-dire le progrès thérapeutique (ou diagnostique) apporté par le médicament par rapport à l'existant, dans le cadre d'une prise en charge clinique optimale au moment de l'évaluation ?
- quel est l'effectif de la population pour laquelle le remboursement est justifié ou le médicament représente une valeur ajoutée ?
- quel est l'impact, voire l'intérêt, de ce médicament pour la santé publique ?

Ces conclusions sont destinées aux acteurs de la décision (CEPS, UNCAM, ministre chargé de la Santé et de la Sécurité sociale) et aux professionnels impliqués dans le bon usage du médicament (laboratoire exploitant, professionnels de santé, en particulier le médecin qui le prescrit et le pharmacien qui le dispense, le malade, et le citoyen qui le consomme.

1.2.2.1 Le SMR

Dans son évaluation d'un nouveau médicament, la CT ne se satisfait pas de la seule démonstration d'une supériorité statistique par rapport au placebo. Pour la CT cette supériorité statistique n'est qu'un prérequis et elle accorde davantage de valeur à la quantité d'effet manifestée par le produit, et plus généralement, à son intérêt clinique, sa place dans la stratégie thérapeutique et son impact attendu en santé publique. Aussi, le SMR prend en compte :

- l'efficacité et les effets indésirables du médicament ;
- sa place dans la stratégie thérapeutique, notamment au regard des autres thérapeutiques disponibles ;
- la gravité de l'affection à laquelle il est destiné ;
- le caractère préventif, curatif ou symptomatique du traitement médicamenteux ;
- l'intérêt du médicament pour la santé publique.

Le SMR comprend 4 niveaux (important, modéré, faible, insuffisant pour une prise en charge par la solidarité nationale) qui conditionnent l'accès au remboursement et le taux de prise en charge des médicaments. Le SMR doit être apprécié pour chacune des indications cliniques autorisées du médicament. Un SMR insuffisant conduit à un avis défavorable de la Commission à la prise en charge du médicament par la solidarité nationale. Un SMR suffisant (important, modéré ou faible) constitue une recommandation à l'inscription sur l'une ou les

deux listes des médicaments remboursables, et pour le taux de ce remboursement pour la liste Sécurité sociale (65%, 30%, 15%). Si le ministre de la santé suit l'avis de la HAS (il en est presque toujours ainsi), il revient ensuite au directeur de l'UNCAM de fixer le taux de prise en charge selon ces dispositions.

Un certain nombre de facteurs sont identifiables comme pouvant conduire à l'attribution d'un SMR insuffisant⁷⁸:

- une quantité d'effet très minime sans pertinence clinique au regard d'effets indésirables notables ;
- un niveau d'efficacité dont la démonstration manque de fiabilité ;
- une efficacité seulement démontrée dans une population autre que celle qui correspond à l'autorisation de mise sur le marché ou démontrée dans une population dont la transposabilité à la population effectivement ciblée n'est pas certaine ;
- une absence de place dans la stratégie thérapeutique des affections visées par ses indications ;
- une indication correspondant à une pathologie ou à un symptôme à traiter bénin ou spontanément curable ;
- l'existence d'alternatives thérapeutiques (médicamenteuses ou non) ayant fait la preuve d'une efficacité plus fiable, plus importante ou dont les effets indésirables sont moins graves, le nouveau médicament étant alors susceptible d'induire une situation de perte de chance pour les patients ;
- une association de médicaments à doses fixes peu compatible avec une prise en charge adaptée à des situations pathologiques évolutives ou à des besoins thérapeutiques variables dans le temps.

A contrario, contribuent à l'attribution d'un SMR suffisant :

- une démonstration méthodologiquement fiable d'une efficacité dont la quantité d'effet est notable pour le patient ;
- une démonstration d'efficacité réalisée sur un critère clinique de morbi-mortalité (exceptionnellement sur un critère intermédiaire) ;

- une place reconnue du médicament dans la stratégie thérapeutique au regard de la maladie concernée ;
- une maladie grave, mais en rappelant que la gravité de la maladie ne permet pas de justifier à elle seule d'un SMR suffisant (les performances du médicament constituant l'élément essentiel de l'appréciation) ;
- des alternatives peu nombreuses ou absentes (besoin thérapeutique insuffisamment couvert ou non couvert) ;
- un médicament sans démonstration de supériorité par rapport à l'existant, mais dont la tolérance est supposée meilleure ou le mécanisme d'action, différent, laisse présager une alternative intéressante chez les patients non répondeurs ou intolérants aux traitements disponibles.
- une performance thérapeutique équivalente ou non inférieure par rapport à un comparateur pertinent ayant démontré son efficacité.

Pour les médicaments ayant obtenu un niveau de SMR suffisant, l'avis de la Commission doit aussi comporter une appréciation de l'Amélioration du Service Médical Rendu, par comparaison aux alternatives médicamenteuses disponibles.

1.2.2.2 L'ASMR

La CT doit préciser le niveau d'amélioration du service médical rendu (AMSR), c'est-à-dire, le progrès thérapeutique apporté par l'utilisation du nouveau médicament, comparativement aux médicaments déjà existants ou toute autre méthode thérapeutique disponible dans le domaine considéré. Elle mesure la valeur ajoutée médicale du médicament évalué et le progrès qu'il apporte dans la stratégie thérapeutique du moment. Cette appréciation constitue donc un instantané dans un environnement évolutif. Le niveau d'ASMR est l'un des éléments pris en compte par le CEPS pour fixer le prix du médicament.

L'obtention d'une ASMR peut traduire la reconnaissance soit d'un progrès quantitatif (A deux fois plus efficace que B), soit d'un progrès qualitatif (A permet d'atteindre des patients non répondeurs ou insuffisamment répondeur à B), ou traduire une meilleure tolérance. Apprécier le niveau d'ASMR d'un médicament présuppose que celui-ci soit comparé à un comparateur « pertinent ». Ce comparateur peut être un médicament, un

dispositif médical, un acte ou toute autre thérapie (ou méthode diagnostique) non médicamenteuse. Il se situe à la même étape de la stratégie thérapeutique que le nouveau produit. Si un comparateur pertinent existe, le fait que le nouveau médicament ait seulement été comparé au placebo n'est pas acceptable. Les résultats de la comparaison directe avec le (ou les) comparateur(s), qui prennent en considération un critère de jugement clinique approprié et sa pertinence pour le malade, la quantité d'effet et la qualité de la démonstration, vont permettre de définir un niveau d'ASMR parmi 5 options :

- la non-infériorité démontre l'absence de progrès : il n'y a pas d'ASMR (niveau V).
- en cas de supériorité démontrée, l'importance de la différence permet de quantifier l'ASMR en 4 niveaux :
 - mineur (IV)
 - modéré (III)
 - important (II)
 - majeur (I) : pour des médicaments qui ont démontré une très notable efficacité sur la mortalité dans la prise en charge d'une maladie grave.

} En fonction de l'intensité du surcroît d'intérêt clinique en termes d'efficacité ou de tolérance.

Des modalités nouvelles d'administration, de galénique, etc. peuvent être considérées comme un progrès si une conséquence clinique significative est démontrée.

Le niveau d'ASMR conditionne le prix du médicament. L'accord cadre⁷⁹ conclu en 2012 entre l'Etat et les Entreprises du médicament (LEEM), prévoit notamment qu'un médicament ayant une ASMR I à III peut bénéficier d'une garantie de prix européen. En revanche, les médicaments avec une ASMR V (absence de progrès) ne peuvent faire l'objet de remboursement que s'ils sont moins coûteux que leurs comparateurs déjà disponibles.

1.2.3 La Primo-inscription

Lors de la primo-inscription, la Commission doit évaluer le bien-fondé du remboursement du médicament, par l'Assurance maladie, en ville et/ou à l'hôpital, au vu de son SMR. De plus, l'ASMR évalue si le médicament est susceptible d'apporter un surcroît en efficacité ou en tolérance par rapport aux traitements déjà disponibles au même stade de la stratégie thérapeutique. Cette évaluation est nécessaire car l'AMM ne garantit pas l'existence d'un progrès thérapeutique par rapport aux médicaments déjà sur le marché. Ainsi, parmi les produits enregistrés, la CT doit faire des choix et ne retenir que ceux qui lui

paraissent rendre un service médical suffisant. Les avis de la CT publiés sur le site de la HAS détaillent pour chaque médicament le résultat de l'évaluation du SMR et de l'ASMR. Ces avis, consultatifs, sont destinés au CEPS, organisme interministériel qui fixe le prix des médicaments, au Directeur général de l'Union Nationale des Caisses d'Assurances Maladie (UNCAM) qui fixe le taux de remboursement et au Ministre qui, *in fine*, décide de la prise en charge ou non.

1.2.4 Le renouvellement d'inscription

Après sa primo-inscription, le médicament entre par la suite dans un circuit de réévaluation régulière (quinquennale pour les médicaments inscrits sur la liste Sécurité sociale) ou exceptionnelle (quel que soit la liste, par exemple à l'initiative des pouvoirs publics ou de la Commission elle-même). La pharmacovigilance reste assurée par l'ANSM. Aussi, tous les cinq ans après la première inscription, le laboratoire titulaire de l'AMM dépose un nouveau dossier de demande de renouvellement d'inscription sur la liste des spécialités remboursables aux assurés sociaux, à la fois au ministère et à la CT. A cette occasion, le titulaire de l'AMM doit fournir toutes les données dont il dispose pour confirmer certains éléments (par exemple des critères d'efficacité, de tolérance, etc...) de son médicament à travers les résultats des études post-inscription demandées par le CEPS et/ou la HAS lors de la primo-inscription. Ces résultats peuvent conduire la CT à modifier le SMR ou l'ASMR initialement appréciés. La demande de réinscription quinquennale ne concerne que les médicaments inscrits sur la liste des spécialités remboursables aux assurés sociaux. L'inscription sur la liste des spécialités pharmaceutiques agréées à l'usage des collectivités et divers services publics est actuellement unique et définitive. Cependant, hors de ce cadre réglementaire quinquennal de réinscription, la CT peut réaliser une réévaluation sur saisine du ministre, à la demande du titulaire de l'AMM ou de sa propre initiative dans des situations où le contexte médical a évolué, notamment avec l'apparition de nouveaux médicaments plus efficaces pouvant modifier la stratégie thérapeutique.

1.2.5 Le recours aux comparaisons indirectes et les recommandations de la HAS

Comme nous venons de le voir, pour apprécier l'ASMR d'un médicament, celui-ci doit être comparé aux médicaments déjà existants se situant à la même étape de la stratégie

thérapeutique. Cette évaluation du « nouveau » par rapport aux références plus anciennes nécessite que le nouveau médicament ait été comparé à au moins un des comparateurs dans des essais cliniques. Or, dans la mesure où l'exigence pour l'obtention de l'AMM n'est pas comparative, le plus souvent, les plans de développement des laboratoires pharmaceutiques comportent davantage d'essais *versus* placebo que d'essais *versus* traitement de référence. En France, ce manque est 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 9⁸⁰ : « *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.* ». En 2011, la loi Bertrand a renforcé le niveau d'exigence d'évaluation des médicaments, avec notamment la possibilité de demander la soumission d'essais cliniques effectués contre comparateurs actifs et contre placebo ; tout refus de la part de l'industriel devant être justifié⁸¹. Cependant, il n'est pas toujours possible d'obtenir des comparaisons d'un médicament versus tous ses comparateurs actifs, ce qui serait très coûteux pour un laboratoire, et pas forcément utile, notamment dans les aires thérapeutiques où la pharmacopée est très développée. La comparaison à un traitement de référence paraît satisfaisante, mais dans le cadre de développements européens, le médicament de référence peut ne pas être le même dans tous les pays et le laboratoire doit alors faire un choix. En outre, des comparaisons directes entre le médicament et son comparateur ne sont pas toujours réalisables. En effet, si les deux médicaments ont été développés dans des temps simultanés (totalement ou partiellement, dans les 3 ans environ), il est difficile d'obtenir qu'ils aient été comparés entre eux dans des essais cliniques.

Pour pallier ce manque d'essais comparatifs des traitements entre eux dans de très nombreux champs thérapeutiques, le recours à la comparaison indirecte est une méthode qui peut répondre à la question de la valeur relative des médicaments. Elle est de plus en plus employée par les laboratoires quand se pose la question de la hiérarchie en termes d'efficacité et/ou de sécurité de traitements concurrents. Devant cette utilisation croissante des comparaisons indirectes dans les dossiers déposés par les laboratoires, la HAS a publié

un rapport de synthèse en juillet 2009⁴ afin d'identifier les situations où les comparaisons indirectes peuvent être utilisées et les méthodes utilisables^k. Ainsi, selon la HAS, en l'absence de comparaison directe, une comparaison indirecte, réalisée sur des bases méthodologiques définies dans ce rapport, peut être prise en compte dans l'évaluation de l'efficacité relative de deux thérapies. Les comparaisons indirectes qui ne sont pas réalisées selon ce standard ne sont pas considérées dans l'évaluation. Il est aussi précisé que les résultats de comparaison directe ne doivent pas reléguer au second plan ceux des comparaisons indirectes et que ces deux estimations doivent être soigneusement confrontées et si possible combinées à l'aide d'une approche mixte (*mixed treatment comparison*). Selon la HAS, la mise en œuvre de cette approche doit 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). Il est souligné que les autres techniques valides sont aussi utilisables en l'absence, pour l'instant, de comparaisons approfondies des performances relatives des différentes méthodes.

Aussi, depuis 2009, la HAS envisage formellement le recours aux comparaisons indirectes afin d'évaluer le positionnement d'un nouveau traitement par rapport aux traitements existants et ainsi pallier le manque d'essais comparatifs des traitements actifs entre eux.

Nous avons souhaité déterminer si les comparaisons indirectes ont effectivement un poids dans l'évaluation du médicament par la Commission de la transparence. L'étude que nous avons menée dans le but de répondre à cette question est présentée dans le chapitre qui suit.

^k Ces méthodes sont exposées dans la partie 1 de cette thèse.

2 Place des comparaisons indirectes dans l'évaluation des médicaments par la Commission de la transparence : étude rétrospective 2009-2014

2.1 Objectif

L'objectif de notre étude rétrospective réalisée sur la période de juillet 2009 à octobre 2014 est de répondre aux questions suivantes :

- Typologie des comparaisons indirectes : Quelles méthodes de comparaison indirecte sont employées le plus couramment dans les études déposées par les laboratoires? Dans quel type de demande (inscription, renouvellement d'inscription etc.) ? Est-ce que les comparaisons indirectes prédominent dans certaines aires thérapeutiques ?
- Validité : Quel est le pourcentage de comparaisons indirectes proposées par les laboratoires considérées comme valides et donc acceptables par la CT ? Quels critères de validité sont le plus souvent remis en question ?
- Conséquences sur l'appréciation des médicaments : Ces comparaisons indirectes ont-elles eu un impact sur les avis de la Commission, notamment en termes d'ASMR?

2.2 Méthodes

Nous avons réalisé une requête par mots clés dans une base de données interne au SEM avec un filtre sur la date de mise à disposition de l'avis. Nous avons ainsi relevé les avis mis à disposition de juillet 2009 (date de publication du rapport de synthèse) à octobre 2014 comportant le mot « indirecte » ou « réseau » afin de sélectionner ceux mentionnant une comparaison indirecte. Nous avons éliminé les avis qui comportaient ces termes dans un contexte différent de celui des comparaisons indirectes (par exemple : « réseau » de soins). De plus, les rapports d'évaluation d'une classe de médicaments étaient exclus de l'analyse car il était difficile de relier un médicament à un niveau d'ASMR.

Pour chaque avis, les données suivantes ont été extraites :

- Date de publication de l'avis ;
- Type de demande (primo-inscription/renouvellement ?) ;
- Aire thérapeutique (classification ATC) ;

- Méthode de comparaison indirecte utilisée ;
- Comparaison indirecte jugée acceptable par la Commission ? oui/non et extrait de l'avis ;
- Critiques méthodologiques évoquées ;
- Comparaison indirecte citée dans les conclusions de l'avis ? oui/non et extrait de l'avis ;
- Niveau d'ASMR obtenu.

Lorsque la méthodologie de la comparaison indirecte était publiée, nous sommes allés rechercher l'article afin d'y trouver les renseignements quand cela était nécessaire. Concernant les critiques méthodologiques, les différents points ont été relevés grâce à une grille d'évaluation créée à partir des éléments de la partie 1/chapitre 3 (Annexe 2).

2.3 Résultats

2.3.1 Nombre d'avis relevés

Un total de 254 avis comportant le mot clé « indirecte » et/ou « réseau » parmi les avis rendus par la CT a été identifié entre juillet 2009 et octobre 2014. Sur ces 254 avis, nous en avons éliminés 131 pour lesquels ces mots clés n'étaient pas utilisés pour désigner une comparaison indirecte. Parmi les 123 avis restants, 24 avaient été relevés deux fois car ils comportaient à la fois le mot « indirecte » et le mot « réseau ». Après avoir exclu ces doublons, il restait 99 avis. Sur ces 99 avis, nous en avons retiré 39 de l'analyse pour les raisons suivantes :

- la demande a été retirée par le demandeur (2 avis),
- il s'agissait de la réévaluation de classes de médicaments (24 avis) (médicaments à base de somatropine, médicaments utilisés dans la maladie d'Alzheimer, médicaments utilisés dans l'incontinence urinaire, médicaments utilisés dans la sclérose en plaques, médicaments à base de diclofénac),
- le terme « indirecte » ne désignait pas une comparaison indirecte (5 avis qui avaient échappés au premier tri),
- il s'agissait de l'audition d'un laboratoire (1 avis),

- il était question de comparaison indirecte sur des données de pharmacocinétique (2 avis),
- il s'agissait de doublons de dossiers (3 avis),
- la date de l'avis ne correspondait pas à la période de l'analyse (2 avis).

Nous avons donc considéré les 60 avis restants, qui étaient répartis ainsi suivant les années :

- juillet 2009 – décembre 2009 (6 mois) : 2 avis. On peut donc extrapoler à 4 avis sur 12 mois.
- 2010 : 7 avis
- 2011 : 7 avis
- 2012 : 12 avis
- 2013 : 19 avis
- janvier - octobre 2014 (10 mois) : 13 avis. On peut donc extrapoler à 15,6 soit environ 16 avis sur 12 mois.

Cette répartition (avec les valeurs extrapolées) est représentée sur la Figure 23.

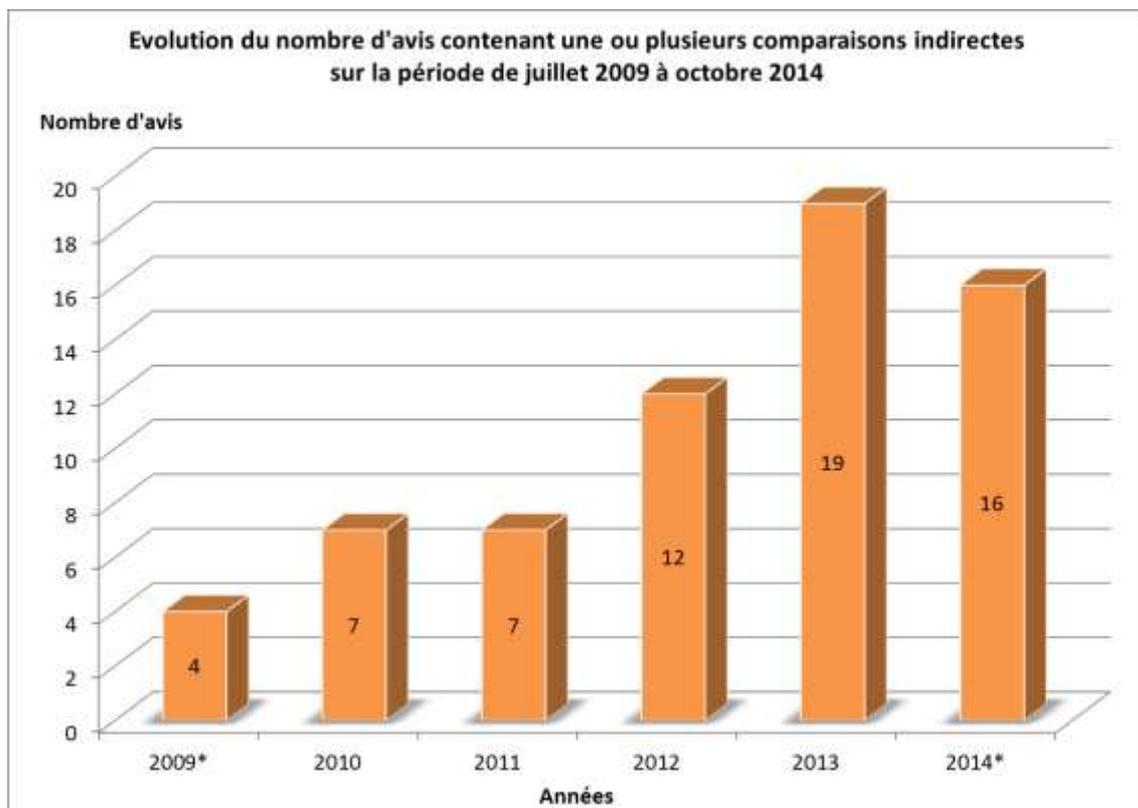


Figure 23 : Evolution du nombre d'avis contenant une ou plusieurs comparaisons indirectes sur la période de juillet 2009 à octobre 2014.

*: valeurs extrapolées (pour 2012 : 2 avis sur 6 mois, extrapolés à 4 avis sur 12 mois ; pour 2014 : 13 avis sur 10 mois extrapolés à 16 avis sur 12 mois).

Cette figure indique une utilisation croissante des comparaisons indirectes par les laboratoires depuis 2009 (sauf en 2014). Cependant elle peut aussi traduire une prise en compte croissante de ces méthodes par le SEM et la CT, étant donné que la méthodologie de notre étude consiste à tenir compte des avis dans lesquels les mots « indirecte » et « réseau » sont retranscrits à partir des dossiers déposés par les laboratoires.

2.3.2 Typologie des avis contenant des comparaisons indirectes

2.3.2.1 Méthodes de comparaison indirecte employées

Dans un premier temps, nous nous sommes intéressés à la méthode employée pour effectuer la comparaison indirecte, selon les terminologies définies dans la partie 1 de cette thèse.

Parmi les 60 avis considérés, 61 comparaisons indirectes ont été étudiées (1 avis comportait 2 comparaisons indirectes), et les méthodes utilisées étaient réparties ainsi :

- 13 comparaisons naïves (21%)
- 3 tests d'hétérogénéité (5%)
- 6 comparaisons indirectes sans précision (mais non naïves) (10%)
- 8 comparaisons indirectes ajustées selon la méthode de Bucher (13%)
- 28 méta-analyses en réseau (46%)
- 1 méta-régression (2%)
- 2 comparaisons indirectes utilisant la méthode de méta-analyse en réseau avec deux approches : la méthode de Bucher d'une part, et la méthode Bayésienne d'autre part. (3%).

La répartition des méthodes utilisées est représentée sur la Figure 24.

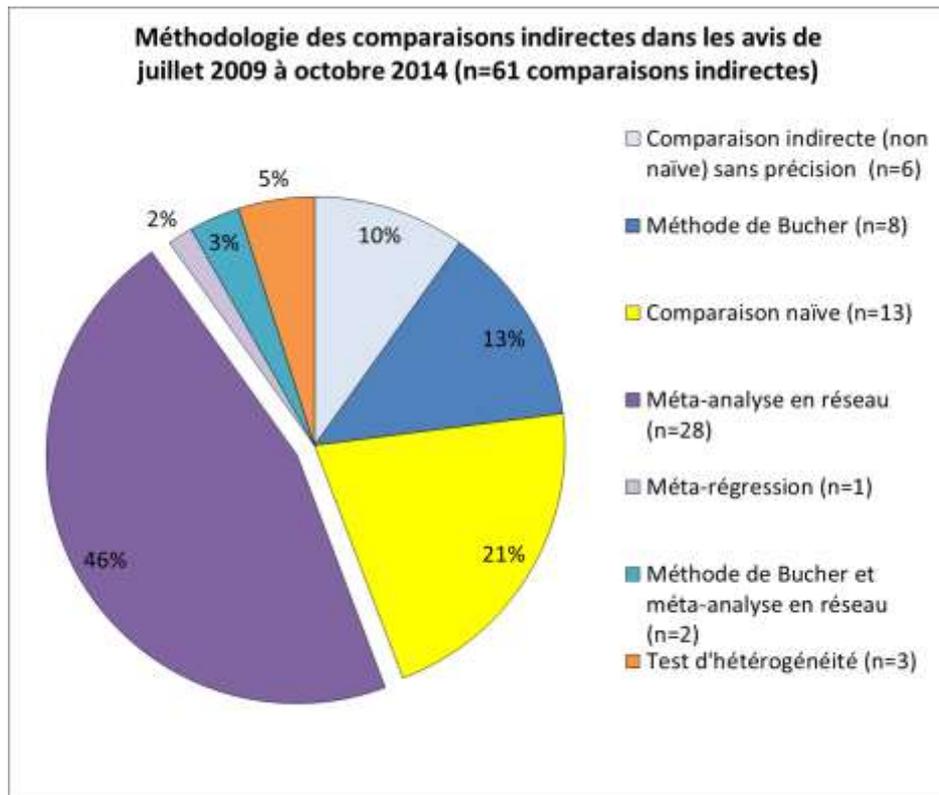


Figure 24: Méthodologie des comparaisons indirectes dans les avis de juillet 2009 à octobre 2014 (n=61 comparaisons indirectes).

On voit sur cette figure que la méthode la plus courante est la méta-analyse en réseau (46%), conformément aux recommandations de la HAS de 2009⁴.

Les comparaisons naïves figurent au 2^{ème} rang (21%) ce qui peut paraître surprenant puisqu'il s'agit d'une méthode inappropriée. Parmi ces 13 comparaisons naïves retrouvées, il s'agissait :

- de comparaisons des résultats de méta-analyses classiques (comparaison naïve d'estimations ponctuelles) (8 comparaisons) ;
- de comparaisons « historiques » c'est-à-dire comparant les résultats de plusieurs études dont certaines anciennes, ceci sans ajustement (comparaison naïve de bras actifs) (3 comparaisons) ;
- de comparaisons indirectes réalisées « en l'absence de méthodologie statistique appliquée à cette comparaison » (2 comparaisons).

La méthode de Bucher est parfois utilisée, alors que le recours à la méta-régression est très rare.

Lorsqu'on s'intéresse à la répartition des différentes méthodes suivant les années (Figure 25), on peut noter une tendance à la diminution de l'utilisation de la comparaison naïve et de la méthode de Bucher au profit de la méta-analyse en réseau.

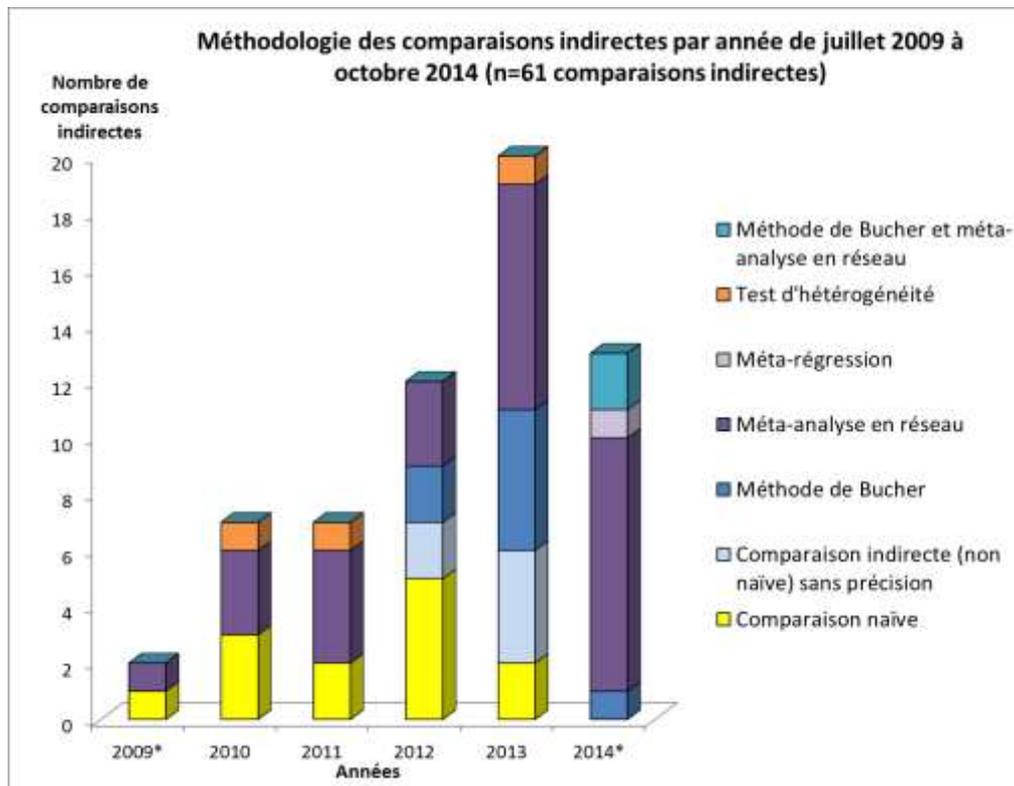


Figure 25: Méthodologie des comparaisons indirectes par année de juillet 2009 à octobre 2014. 2009*: période de juillet à décembre 2009. 2014* : période de janvier à octobre 2014.

2.3.2.2 Aires thérapeutiques concernées par des comparaisons indirectes

Nous avons cherché à déterminer si les comparaisons indirectes sont plus souvent employées dans certaines aires thérapeutiques, et si oui lesquelles.

La Figure 26 met en évidence que les aires thérapeutiques (définies par la classe ATC) les plus représentées sont :

- les immunosuppresseurs (tels que ROACTEMRA tocilizumab, CIMZIA certolizumab, SIMPONI golimumab, ORENCIA abatacept, KINERET anakinra, REMICADE infliximab),
- les agents antinéoplasiques (tels que VOTRIENT pazopanib, INLYTA axitinib, EFFALA acide 5-amino lévulinique, TAFINLAR dabrafenib), et

- les anti-thrombotiques (BRILIQUE ticagrelor, ELIQUIS apixaban, XARELTO rivaroxaban , ARIXTRA fondaparinux sodique, FRAGMINE daltéparine sodique).

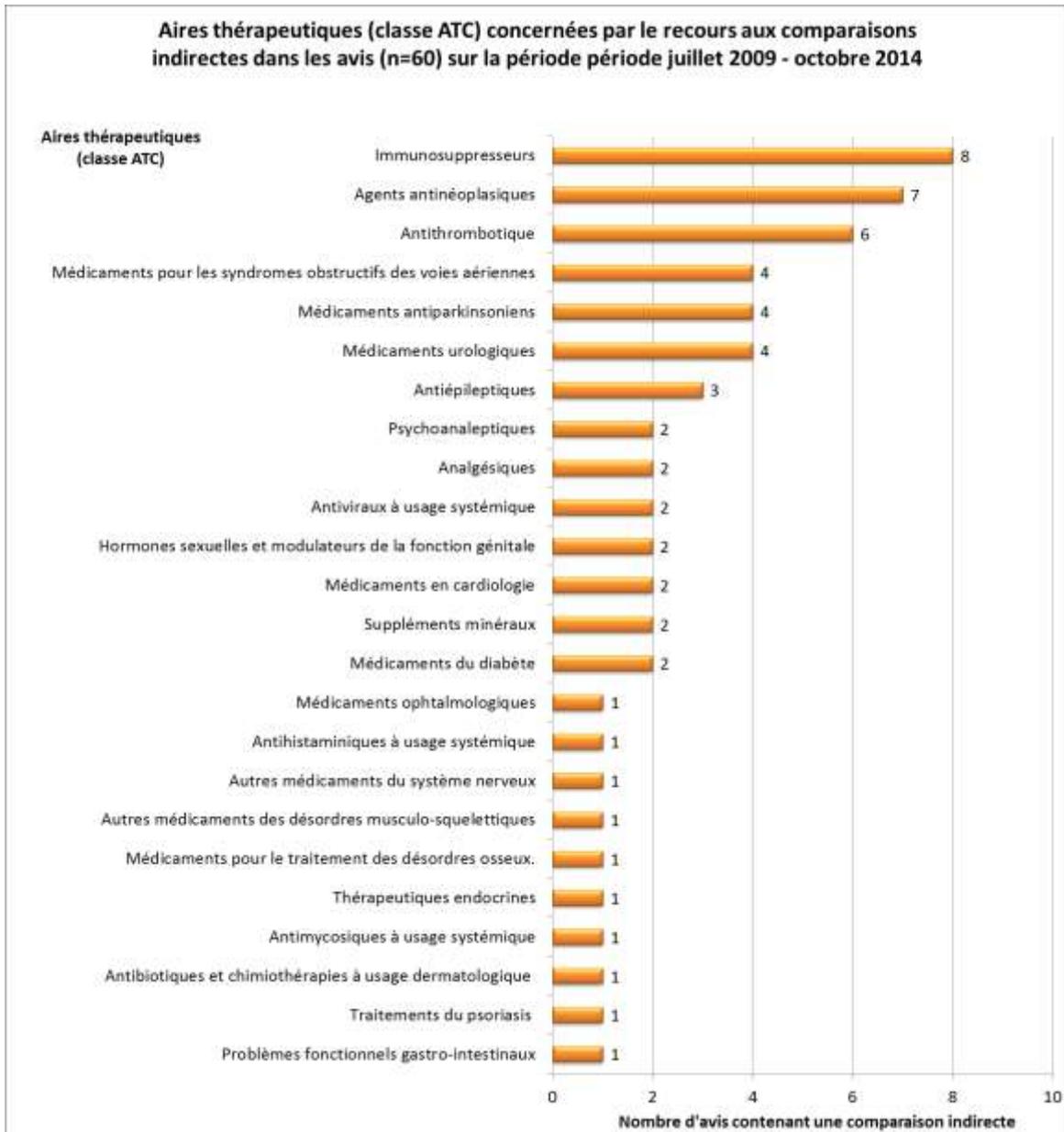


Figure 26 : Aires thérapeutiques (classes ATC) concernées par le recours aux comparaisons indirectes dans les avis (n=60), période juillet 2009-octobre 2014.

Les comparaisons indirectes concernent toutes les pathologies, mais les aires thérapeutiques les plus représentées ne sont pas celles où la pharmacopée est la plus développée comme on aurait pu s’y attendre. Cependant c’est le reflet des molécules mises sur le marché ces dernières années.

2.3.2.3 Motifs d'examen

Le relevé de données a permis de préciser le type de demande dans lequel on retrouve le plus souvent des comparaisons indirectes (Figure 27). Ainsi, c'est dans le cas d'inscription (53%) que les comparaisons indirectes sont les plus présentes, puis viennent ensuite les renouvellements d'inscription (23%), les réévaluations du SMR ou de l'ASMR (17%), et enfin les extensions d'indication (7%). Il nous a paru nécessaire de mettre ces résultats en perspective avec la répartition « habituelle » des dossiers traités par la CT. Nous avons choisi les données de l'année 2013 car ce sont les plus récentes⁷⁷ (Figure 28). Ainsi, en 2013, les deux types de demandes les plus fréquentes étaient en premier le renouvellement d'inscription (35%), puis en second l'inscription (28%). Cela nous montre que la répartition observée pour les comparaisons indirectes ne s'explique pas uniquement par la répartition habituelle des demandes à la CT. Sachant qu'au cours des renouvellements d'inscription, l'ASMR n'est pas réévaluée (seul le SMR l'est), cela peut expliquer qu'il est peu pertinent, eu égard aux critères en jeu dans l'évaluation, de présenter des données de comparaisons indirectes à l'occasion d'une inscription plutôt qu'à l'occasion du renouvellement quinquennal.

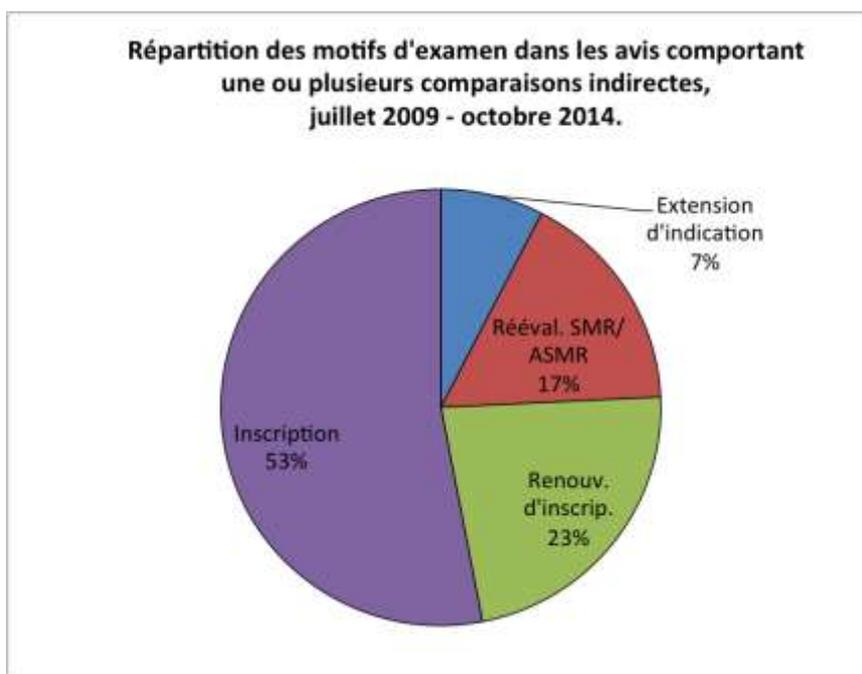


Figure 27 : Répartition des motifs d'examen dans les avis comportant une ou plusieurs comparaisons indirectes, juillet 2009-octobre 2014.

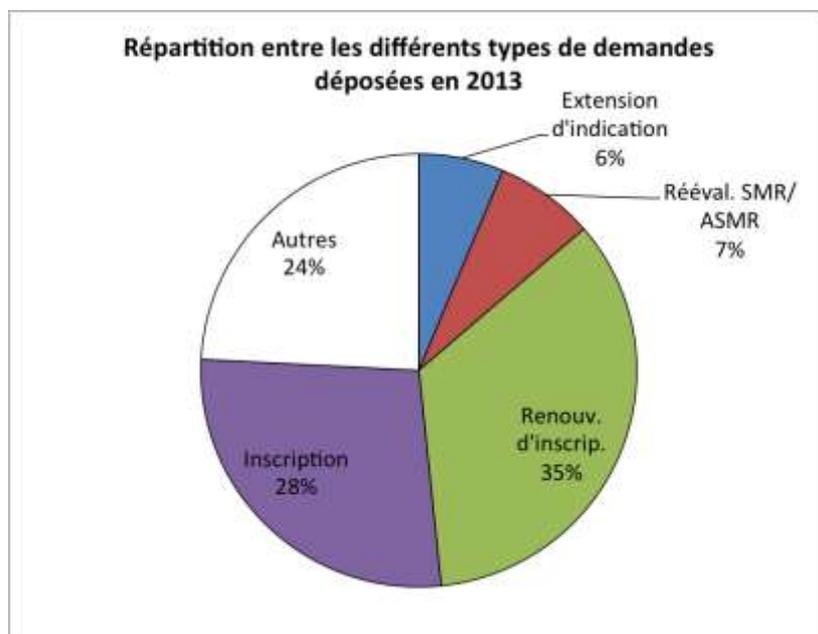


Figure 28 : Répartition entre les différents types de demandes déposées en 2013 auprès de la CT.

2.3.3 Validité des comparaisons indirectes et critiques méthodologiques

2.3.3.1 Validité des comparaisons indirectes

Sur les 61 comparaisons indirectes relevées dans les avis, 31 ont été considérées comme « acceptables » d'un point de vue méthodologique, et 30 n'ont pas été considérées comme exploitables en raison de diverses faiblesses méthodologiques. Cette répartition est illustrée sur la Figure 29. Ainsi environ 50% des comparaisons indirectes mentionnées dans les avis n'ont pas été retenues par la CT pour l'évaluation du médicament en raison d'objections méthodologiques.

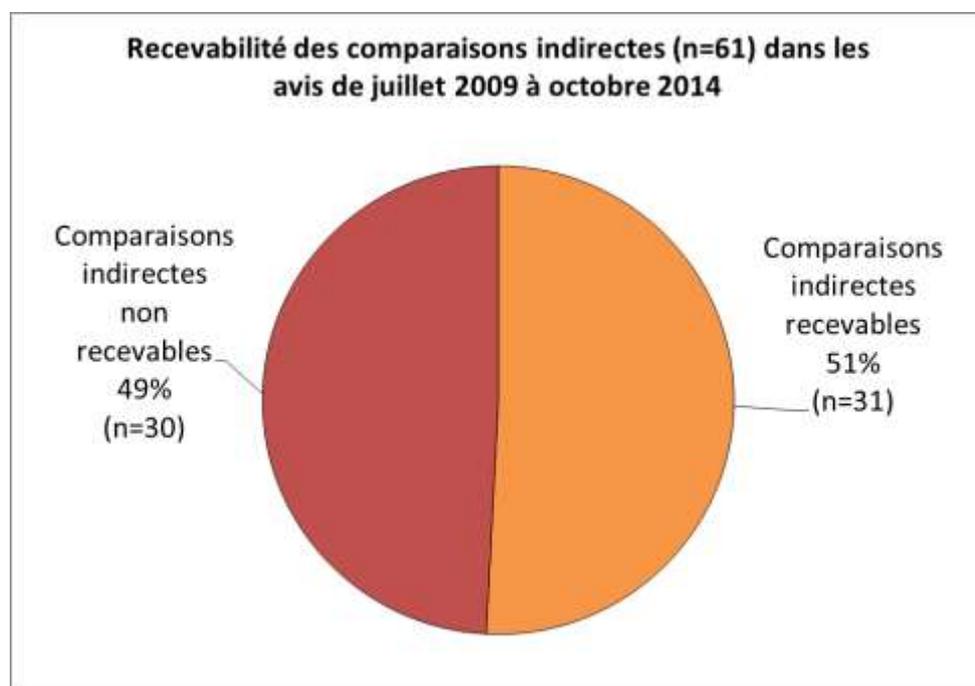


Figure 29: Recevabilité des comparaisons indirectes (n=61) dans les avis de juillet 2009 à octobre 2014.

2.3.3.2 Principales critiques méthodologiques

Différentes critiques ont été émises concernant certaines comparaisons indirectes qui ont été jugées de ce fait non recevables. La diversité de ces critiques est représentée sur la Figure 30, conformément aux différents items exposés dans la partie 1 chapitre 3 « Guide de lecture critique ». On peut noter que les comparaisons indirectes qui ont été jugées recevables ont parfois également reçu des critiques négatives de la part de la CT, mais nous

ne les avons pas représentées sur le graphique dans la mesure où elles n'ont pas empêché la prise en compte et l'interprétation de la comparaison indirecte.

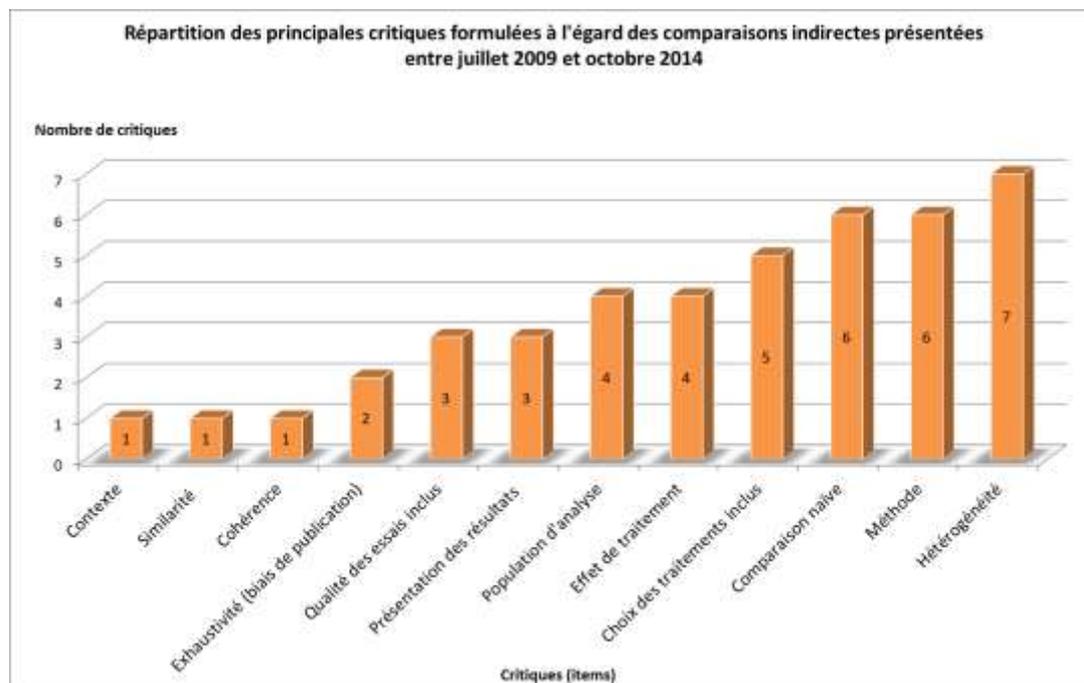


Figure 30 : Répartition des principales critiques formulées à l'égard des comparaisons indirectes présentées entre juillet 2009 et octobre 2014.

Contexte : La définition de la pathologie, les populations de patients, les modalités générales de prise en charge sont-elles susceptibles d'avoir évolué entre les moments où ont été réalisées les différentes études incluses dans la méta-analyse en réseau ? Similarité : Dans les conditions de l'essai de B, le traitement A aurait donné les mêmes résultats que ceux obtenus dans l'essai A, et vis versa. Cela suppose un équilibre dans les caractéristiques des patients ou de l'étude en terme de modificateur de l'effet relatif de traitement. Cette hypothèse se vérifie à l'aide des analyses en sous-groupe et du test d'interaction. Cohérence : L'hypothèse de cohérence est vérifiée s'il n'y a pas de différence entre les résultats des comparaisons directes et ceux des comparaisons indirectes, ou si le résultat de la comparaison indirecte ne varie pas en fonction du « chemin¹ » emprunté. Exhaustivité (biais de publication) : Un processus de recherche exhaustive et reproductible des essais de type méta-analyse a-t-il été suivi ? S'agit-il d'essais sélectionnés pour donner les résultats de comparaisons indirectes souhaités ou s'agit-il de tous les essais réalisés dans le domaine ? Qualité des essais inclus : Le rapport doit fournir les renseignements concernant les caractéristiques clés de chaque ECR, comme la méthode de randomisation, les perdus de vue etc. Présentation des résultats : présence d'une représentation graphique du réseau de preuves et d'un tableau récapitulatif avec les études incluses, les interventions, et les résultats observés avec chaque intervention dans chaque étude. Population d'analyse : Les populations des ECR qui constituent la base de la méta-analyse en réseau correspondent-elles à la population d'intérêt pour le décideur ? (Maladie d'intérêt, stade, gravité, antécédents de traitements, âge, sexe etc.) Effet de traitement : Les critères d'évaluation sont-ils pertinents ? (Clinical endpoint, intermediate endpoint : ex : mortalité cardiovasculaire, taux de cholestérol etc.) Choix des traitements inclus : Les médicaments inclus dans la comparaison indirecte (méta-analyse en réseau) correspondent-ils à ceux d'intérêt pour le décideur ? Tous les comparateurs pertinents ont été considérés ? Comparaison naïve : La méthode utilisée pour réaliser la comparaison indirecte est une comparaison naïve donc inappropriée. Méthode : La méthode statistique employée est suffisamment décrite pour en juger de l'acceptabilité (aperçu global des méthodes statistiques, justification du choix des critères d'évaluation, des estimations des effets relatifs, du choix d'un modèle à effets fixes ou aléatoires, du choix de la distribution *a priori* pour les paramètres du modèle si les analyses ont été effectuées dans un cadre bayésien, description des différentes analyses de sensibilités). Hétérogénéité : Les études rassemblées dans les méta-analyses doivent être homogènes (quantification de l'hétérogénéité grâce à l'indice de Higgins I²). Utilisation d'un modèle à effets aléatoires pour prendre en compte une certaine variabilité aléatoire de l'effet traitement d'un essai à un autre, ou bien d'un modèle à effets fixes si le véritable effet du traitement est commun à tous les essais.

¹ Voir Partie 1, chapitre 2.1.3.2.

Afin d'illustrer par des cas concrets ces différents points, nous avons choisi de présenter certains passages d'avis rapportant des comparaisons indirectes jugées non valides, et parfois, quand l'extrait méritait d'être cité, nous avons choisi ces passages parmi les avis rapportant des comparaisons indirectes jugées valides par la Commission, malgré quelques reproches sur la méthodologie.

2.3.3.2.1 L'hétérogénéité

Ainsi, la critique la plus fréquente concerne l'hétérogénéité de la comparaison indirecte qui limite son interprétation (7 critiques). Afin d'apporter un éclairage qualitatif sur ce point, voici quelques extraits d'avis :

- ❖ Avis INLYTA (axitinib) 2013, comparaison indirecte sans précision : *« les caractéristiques des patients inclus dans ces études diffèrent entre les études, le nombre de lignes et le type de traitements antérieurs diffèrent, certaines quantités d'effet sont issues de résultats d'analyses intermédiaires »*
- ❖ Avis VASTAREL (trimétazidine) 2011, méta-analyse en réseau : *« le choix a priori d'une méthode à effets aléatoires semble indiquer une hétérogénéité fréquemment retrouvée pour les effets thérapeutiques. Le rapport ne détaille malheureusement pas cette hétérogénéité pour chaque molécule et chaque critère de jugement considéré. On aurait aussi souhaité dans chaque cas une exploration détaillée des causes potentielles de l'hétérogénéité observée (ceci d'autant plus qu'aucune synthèse descriptive de la disparité clinique entre les essais n'a été formellement réalisée, même si l'on dispose de tableaux récapitulatifs en fin de rapport). »*
- ❖ Avis KINERET (anakinra) 2013, méta-analyse en réseau : *« Ces résultats sont à interpréter avec prudence étant donné les limites méthodologiques inhérentes à ce type d'analyse et l'hétérogénéité identifiée (liée à la diversité des populations incluses dans les essais cliniques retenus). »*

2.3.3.2.2 La méthodologie de la comparaison indirecte

La deuxième critique méthodologique la plus fréquente concerne la méthodologie de la comparaison indirecte (6 critiques). Nous avons choisis quelques passages d'avis (issus de

comparaisons indirectes qui ont été jugées valides ou non) montrant les incertitudes rencontrées.

2.3.3.2.1 Incertitude concernant la puissance nécessaire

En utilisant une comparaison indirecte, on veut savoir ce que le traitement A produit comme effet par rapport au traitement B, grâce à l'étude A vs placebo (étude 1) et l'étude B vs placebo (étude 2). Ainsi, le nombre de sujets nécessaires n'a pas été calculé pour la comparaison de traitements actifs mais pour la comparaison par rapport à un placebo. De plus, avec la comparaison indirecte, l'imprécision augmente car la taille de l'intervalle de confiance augmente. Aussi, la puissance des comparaisons indirectes est souvent insuffisante et le résultat est donc très souvent non significatif. Devant l'obtention d'un tel résultat, la bonne conduite serait de conclure qu'il n'est pas possible de conclure ; pourtant, bien trop souvent, les auteurs concluent à tort que les traitements sont équivalents. La Commission est très critique face à cette conclusion erronée (reproche évoqué dans 4 avis), comme le montre ces exemples :

- ❖ Avis TAFINLAR (dabrafénib) 2014, méthode de Bucher : *« étant donné le nombre de comparaisons testées (et non spécifiées dans un protocole a priori), la consommation du risque alpha n'est pas contrôlée. Pour les résultats non significatifs, c'est l'interprétation de cette absence de différence qui doit être discutée ici. De tels résultats « inconclusifs » sont fréquents en matière de comparaison indirecte, car cette approche (de type méta-analytique) est entièrement de nature rétrospective et il est donc impossible de garantir dans ce contexte que la puissance de ces comparaisons indirectes soit suffisante. »*
- ❖ Avis FYCOMPA (perampanel) 2013, méta-analyse en réseau : *« La méthodologie envisagée (de type méta-analytique) est de nature rétrospective et il est donc impossible de garantir dans ce contexte que la puissance des comparaisons indirectes soit suffisante pour montrer une supériorité d'un traitement par rapport à un autre. La puissance des comparaisons indirectes est en effet tributaire du nombre de sujets inclus dans chaque essai, effectifs qui ont été calculés avec comme seul objectif de garantir suffisamment de puissance aux comparaisons versus placebo. Cette limitation s'applique à la fois à l'évaluation de l'efficacité et de la tolérance relative. Pour cette dernière évaluation, cette limite devient encore*

plus problématique, les études n'ayant pas été construites avec une puissance suffisante pour l'évaluation des critères de tolérance. Ainsi, la conclusion d'une efficacité/tolérance « comparable » dans le contexte ci-dessus conduit à une sur-interprétation des résultats statistiques car cette conclusion est déduite de l'absence de différence statistiquement significative. »

- ❖ Avis SIMPONI (golimumab) 2014, méta-analyse en réseau (critique identique dans l'avis REMICADE (infliximab) 2014, méta-analyse en réseau) : *« les études utilisées pour les réaliser n'avaient pas été conçues pour comparer des traitements actifs entre eux mais pour les comparer au placebo. Il est en effet impossible de garantir dans ce contexte que la puissance de ces comparaisons indirectes ait été suffisante pour détecter une différence entre les 2 traitements et cela est particulièrement vrai pour les comparaisons en termes de tolérance qui n'étaient pas l'objectif principal des études versus placebo incluses dans la méta-analyse ».*

2.3.3.2.2 Incertitude concernant la validation des *a priori* avec la méthode Bayésienne

Comme nous l'avons abordé dans la partie 1 de cette thèse, il est très important que les *a priori* soient non informatifs dans la méthode Bayésienne pour ne pas orienter les résultats. Aussi nous avons trouvé quelques dossiers critiqués car il n'était pas précisé que les *a priori* étaient non informatifs. Cependant ces remarques n'ont pas été discriminantes puisque les comparaisons indirectes concernées ont tout de même été prises en compte.

- ❖ Avis BETMIGA (mirabegron) 2014, méta-analyse en réseau : *« la méthode bayésienne d'analyse utilisée est partiellement validée au niveau statistique (les a priori étaient non informatifs, l'indépendance des résultats par rapport aux a priori n'a pas été vérifiée). »*
- ❖ Avis FORXIGA (dapaglifozine) 2014, méta-analyse en réseau et méthode de Bucher : *« la méthode bayésienne d'analyse utilisée n'est pas validée au niveau statistique (on ne sait pas si les a priori étaient informatifs ou non, l'indépendance des résultats par rapport aux a priori n'a pas été vérifiée). »*

2.3.3.2.3 Incertitude relative à l'analyse de sensibilité

L'absence d'analyse de sensibilité ou une analyse de sensibilité partielle est critiquable :

- ❖ Avis VASTAREL (trimétazidine) 2011, méta-analyse en réseau : « *l'analyse de sensibilité n'a été réalisée que sur un seul critère : le plan expérimental des essais avec l'inclusion / exclusion systématique de tous les essais en cross-over.* »

2.3.3.2.3 Les hypothèses de similarité et de cohérence

Les hypothèses de similarité et de cohérence sont parfois discutées par la CT :

- ❖ Avis TAFINLAR (dabrafénib) 2014, méthode de Bucher, critique sur la similarité : « *Ces résultats ne doivent être considérés qu'à titre exploratoire, étant donné le risque d'existence d'éventuels facteurs d'interaction potentiels (i.e. homogénéité des populations des essais Break 3 et Brim 3 vis à vis d'éventuelles variables d'interaction).* »
- ❖ Avis XARELTO (rivaroxaban) 2012, méthode de Bucher, commentaire sur la similarité : « *L'hypothèse d'échangeabilité (hypothèse selon laquelle les essais sont suffisamment similaires et/ou les effets des médicaments versus leur comparateur commun sont suffisamment stables indépendamment des caractéristiques des patients pour garantir la reproductibilité des résultats) réalisée apparaît plausible : on dispose d'essais contemporains (publiés entre 2007 et 2010 pour les phases III), réalisés suivant des plans expérimentaux similaires et avec un profil similaire de patients inclus* ».
- ❖ Avis MULTAQ (dronédarone) 2011, méta-analyse en réseau, critique sur la cohérence : « *L'analyse de l'hypothèse de cohérence (i.e. cohérence des résultats sur l'ensemble du réseau de comparaison indirecte) n'a pas été réalisée.* »

2.3.3.2.4 Autres critiques

Les critiques ont également porté sur :

- le choix des traitements inclus (5 critiques) : réseau pauvre, choix discutable d'exclure *a priori* de la méta-analyse certaines molécules.
- les effets traitement (4 critiques) : certaines quantités d'effet correspondent à des résultats issus d'analyses intermédiaires, utilisation d'un score discutable.

- la population d'analyse (4 critiques) : populations différentes en termes de prise en charge, populations ne correspondant pas à l'AMM exacte (notamment le stade de la maladie).
- la présentation des résultats (3 critiques) : schéma en réseau non disponible, étude fournie uniquement sous forme de résumé.
- la qualité des essais inclus (3 critiques) : analyse de la qualité des essais inclus et du niveau de preuve de leurs résultats incomplète.

2.3.3.3 Le cas des comparaisons indirectes naïves

Comme nous l'avons évoqué précédemment, notre étude a révélé 13 comparaisons indirectes naïves. Parmi celles-ci, 11 ont été jugées inacceptables dont 6 qui ont été clairement critiquées (Figure 30) (exemples : «*exposition au biais de sélection et confusion*», «*ne répondent pas aux pré-requis nécessaires à la comparaison indirecte*», «*absence d'ajustement du fait des comparaisons multiples*»). Ainsi, 2 comparaisons indirectes naïves ont été jugées comme acceptables par la Commission. Nous avons souhaité détailler ces 2 cas.

❖ Avis VESICARE (solifénacine) du 7 octobre 2009

Dans cet avis, la comparaison indirecte est une comparaison naïve d'estimations ponctuelles. En effet elle se base sur une méta-analyse classique, qui a comparé la tolérance des traitements antimuscariniques dans l'hyperactivité vésicale^{82,83}. Dans cette méta-analyse la solifénacine et l'oxybutynine ont été comparés au placebo en ce qui concerne les arrêts de traitement pour effets indésirables. L'analyse qui en ressort est que «*les arrêts de traitement pour effets indésirables lors des études randomisées versus placebo avec la solifénacine plaident en faveur d'une tolérance meilleure de la solifénacine lorsqu'on la compare indirectement à l'oxybutynine (RR =1,16, NS pour la solifénacine 5 mg, RR = 1,53 [1,02 ; 2,3], p=0,04 pour la solifénacine 10 mg ; RR = 1,91 [1,18 ; 3,1], p=0,01 pour des posologies d'oxybutynine de 7,5 à 10 mg par jour, RR = 1,89 [1,23 ; 2,9], p<0,01 pour une posologie de 15 mg d'oxybutynine).*» Cette affirmation, qui n'a pourtant pas de fondement méthodologique puisqu'aucune méthode de comparaison indirecte approuvée n'a été utilisée, n'a pas été critiquée et a été reprise dans l'avis.

On peut toutefois noter que ce même raisonnement méthodologique a été présenté dans l'avis CERIS (chlorure de trospium) du 5 janvier 2011 et dans l'avis TOVIAZ (fésotérodine) du 15 février 2012 à partir des mêmes méta-analyses, et qu'il a cette fois été critiqué et non accepté, preuve d'une évolution de la prise en compte des résultats de comparaisons indirectes. Dans ces deux avis les critiques ont été les suivantes : « *Cette méta-analyse comporte certains biais : non exhaustivité de la recherche des données, absence d'ajustement du fait des comparaisons multiples et de test d'hétérogénéité. (...) Les conclusions que l'on peut tirer des comparaisons (...) indirectes avec la solifénacine sont peu informatives pour des raisons méthodologiques.* »

❖ Avis SIFROL (pramipexole) du 19 décembre 2012

Dans cet avis le terme « comparaison indirecte » est aussi utilisé à tort pour désigner le raisonnement intuitif qui consiste à comparer les différents résultats de méta-analyses entre eux sans aucun ajustement (comparaison naïve d'estimations ponctuelles). La méta-analyse Cochrane en cause^{84,85} a inclus les études contrôlées et randomisées disponibles en décembre 2008 pour le lisuride, le pergolide et la cabergoline (agonistes dérivés de l'ergot de seigle), le ropinirole (ADARTREL), le pramipexole (SIFROL), la rotigotine (NEUPRO) et le sumanirole (agonistes non dérivés de l'ergot de seigle). Les résultats de 38 études randomisées ont été pris en compte, 35 études versus placebo (6 954 patients inclus) et 3 études versus L-DOPA. On peut lire dans l'avis « *selon les analyses en sous-groupe (comparaisons indirectes), les agonistes les plus actifs sur les symptômes au score IRLS sont des agonistes ergotés avec une réduction supplémentaire moyenne de -11,5 points [IC 95% : -15,1 à -7,8] pour la cabergoline (2 études) et de - 11,7 [IC 95% : -14,8 à -8,6] pour le pergolide (1 étude).* »

Il s'agit d'un raisonnement purement intuitif qui consiste à comparer l'estimation ponctuelle de l'efficacité de A vs placebo à celle de B vs placebo, hors comme nous l'avons vu dans la partie 1, cette méthode a des limites.

Après avoir déterminé le pourcentage de comparaisons indirectes considérées comme acceptables par la Commission de la transparence et les raisons qui n'ont pas permis d'accepter un certain nombre de comparaisons indirectes, nous avons tenté de définir la place donnée à ces méthodes dans l'attribution de l'ASMR du médicament.

2.3.4 Retentissement de la comparaison indirecte sur les conclusions de l'avis

2.3.4.1 Comparaisons indirectes acceptables reprises dans les conclusions de l'avis

Sur les 31 comparaisons indirectes dont la méthodologie a été acceptée par la Commission, 15 n'ont pas été reprises dans les conclusions de l'avis ce qui laisse à penser qu'elles n'ont pas influencé le raisonnement de la Commission. Parmi les 16 qui ont été retrouvées dans les conclusions de l'avis :

- 5 n'étaient pas « suffisantes » pour pouvoir hiérarchiser les traitements entre eux (l'avis TECFIDERA (dimethyl fumarate) se trouve dans ce cas, on détaillera ce dossier dans la partie 2.3.4.2.2) ou bien la méta-analyse ne montrait de différence statistiquement significative ;
- 11 ont permis une hiérarchisation.

Cette répartition est illustrée sur la Figure 31.

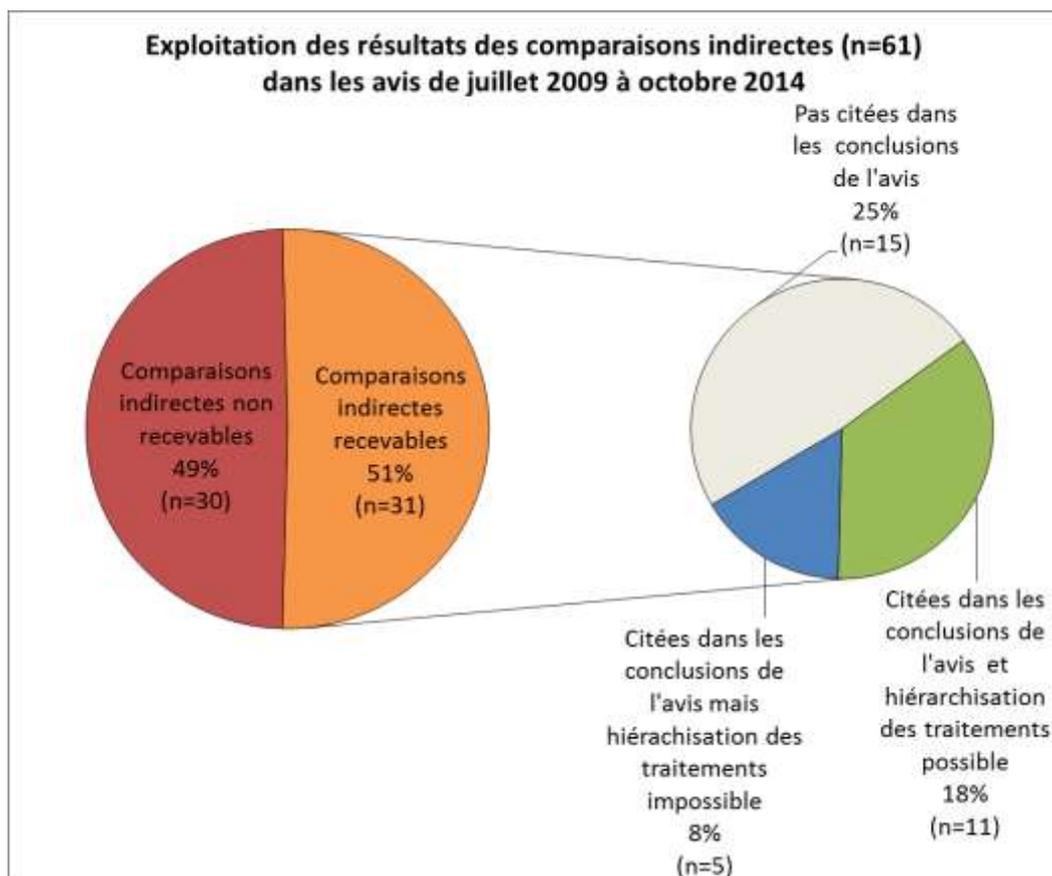


Figure 31 : Exploitation des résultats des comparaisons indirectes (n=61) dans les avis de juillet 2009 à octobre 2014.

Ainsi si l'on considère les 61 comparaisons indirectes relevées dans notre étude seulement 11 se retrouvent dans les conclusions de l'avis soit seulement 18%.

2.3.4.2 Impact des comparaisons indirectes dans la prise de décision

2.3.4.2.1 Concordance avec le niveau d'ASMR

Nous nous sommes risqués à rechercher un éventuel lien entre le résultat de la comparaison indirecte et le niveau d'ASMR attribué par la CT. Sur les 11 dossiers, nous en avons éliminé un qui concernait un renouvellement d'inscription pour lequel la Commission n'a pas statué sur l'ASMR (avis TASMAR (tolcapone) du 2 février 2011), et 2 dossiers concernés par des comparaisons naïves (SIFROL (pramipexole) et VESICARE (solifénacine)) dont nous avons déjà parlé (cf partie 2.3.3.3). La description des 8 dossiers restants est donnée dans le Tableau 9.

Date avis	Dossier	DCI	Typologie de la CI	Conclusion de la comparaison indirecte	Avis de la commission	
					ASMR	Extrait des conclusions de l'avis mentionnant la comparaison indirecte
9-sept.-09	ROACTEMRA	tocilizumab	R	Le tocilizumab n'est pas significativement supérieur aux autres biothérapies.	ASMR V par rapport aux anti-TNF alpha.	La place de ROACTEMRA par rapport aux anti-TNF alpha est difficile à préciser compte tenu de l'absence d'étude l'ayant directement comparé à ces médicaments et du fait du faible recul sur son efficacité et sa tolérance à long terme comparativement aux anti-TNF. Cependant, selon les résultats d'une méta-analyse de comparaison indirecte, l'efficacité de ROACTEMRA serait comparable à celle des anti-TNF.
16-nov.-11	PROLIA	denosumab	R	Les résultats de cette méta-analyse suggèrent l'absence de différence d'efficacité entre le denosumab et les autres molécules disponibles notamment pour la prévention des fractures du col fémoral.	IV	Au vu des données cliniques disponibles sur l'incidence fracturaire issues d'études versus placebo et de comparaisons indirectes ainsi que des études sur critère intermédiaire (variation DMO) versus comparatif actif, il ne peut être attendu de PROLIA qu'un impact supplémentaire faible sur la morbidité des patients traités au regard des alternatives thérapeutiques existantes.
18-janv.-12	ELIQUIS	apixaban	B	L'apixaban n'apparaît pas inférieur au dabigatran etexilate ou au rivaroxaban et ce pour aucun des critères d'efficacité ou de tolérance hémorragique retenus.	IV par rapport à l'énoxaparine (comparaison directe)	Une comparaison indirecte, de bonne qualité méthodologique tend à montrer que l'apixaban (ELIQUIS) est au moins aussi efficace dans ces deux indications que le rivaroxaban (XARELTO) et que le dabigatran (PRADAXA).
1-févr.-12	SIMPONI	golimumab	R	Une méta-analyse de comparaison indirecte n'a pas mis en évidence de différence statistique entre le GLM et les autres anti-TNF chez les patients en échec des traitements de fond classiques de la PR pour les réponses ACR 20, 50 et 70. Aucune comparaison de ce type n'a été réalisée dans la population des patients naïfs de MTX, dans laquelle SIMPONI dispose aussi d'une AMM.	V	Compte tenu, dans la population des patients en échec du méthotrexate (MTX), de : - la démonstration de la supériorité de SIMPONI 50 mg en association au méthotrexate en termes de réponse clinique (ACR 20), - des résultats d'une méta-analyse de comparaison indirecte suggérant l'absence de différence d'efficacité par rapport aux autres anti-TNF [...]
14-mars-12	ORENCIA	abatcept	R	Les résultats de cette méta-analyse de comparaison indirecte sont en faveur d'une efficacité de l'abatcept comparable à celle des autres biothérapies sur le critère principal HAQ.	V	Toutefois, en l'absence de données comparatives directes par rapport aux anti-TNF alpha ou autre biothérapie (anti IL 6) déjà indiqués en situation d'échec au MTX (absence de comparaison directe, non infériorité d'après une méta-analyse), il n'est pas attendu d'ORENCIA un impact supplémentaire sur la morbidité et la qualité de vie des patients par rapport aux anti-TNF alpha.
12-juin-13	ELIQUIS	apixaban	B	<u>En termes d'efficacité</u> : une absence de différence entre l'apixaban, le dabigatran etexilate et le rivaroxaban pour prévenir la survenue des AVC et des ES que ce soit avec ou sans correction du biais. <u>En termes de tolérance</u> , l'apixaban pourrait exposer dans une moindre mesure les patients au risque d'hémorragiques majeures et gastro-intestinales que dabigatran etexilate 150mg et rivaroxaban.	V	<u>Intérêt de santé publique</u> : Au vu des données disponibles, issues notamment d'un essai comparatif en double aveugle (amélioration en termes d'efficacité et de tolérance versus warfarine) et d'une comparaison indirecte suggérant une amélioration en termes de tolérance versus les spécialités XARELTO et PRADAXA, il est attendu un impact supplémentaire modéré de l'apixaban (ELIQUIS) sur la morbi-mortalité des patients traités par rapport à la prise en charge actuelle. <u>ASMR</u> : Les comparaisons indirectes, faites à partir des trois études RELY, ROCKET AF et ARISTOTLE dont la méthodologie et les caractéristiques des patients inclus diffèrent, ne permettent pas de hiérarchiser ces trois médicaments.
16-oct-13	KINERET	anakinra	R	Aucune différence n'a été mise en évidence entre les biothérapies en termes de taux de réponse ACR50, excepté pour l'étanercept qui s'est montré supérieur à l'anakinra (OR 0,34; IC 95% [0,14 ; 0,81]) et pour l'adalimumab qui s'est montré supérieur à l'anakinra (OR 2,20; IC 95% [1,01 ; 4,75]).	V	Une revue systématique de la littérature (revue Cochrane) de comparaison des biothérapies dont l'anakinra (5 études incluses), dans le traitement de la polyarthrite rhumatoïde, montre, sous réserve d'une hétérogénéité substantielle des taux de réponse ACR50 non différents entre les biothérapies, excepté pour l'étanercept et pour l'adalimumab pour lesquels les taux de réponse sont supérieurs à celui de l'anakinra.
11-juin-14	EYLEA	afibercept	BR	Les résultats obtenus avec les deux approches sont concordants et n'ont pas permis de mettre en évidence une différence significative entre afibercept et ranibizumab.	V	Une étude de comparaison indirecte de l'afibercept au ranibizumab a été réalisée en utilisant deux approches, la méthode Bayésienne à effet aléatoire et la méthode de Bucher. Les résultats de la comparaison indirecte, concordants entre les deux méthodes de calcul, n'ont pas permis de mettre en évidence une différence statistiquement significative entre afibercept et ranibizumab en termes de gain d'acuité visuelle ≥ 15 lettres à 6 mois.

Tableau 9 : Concordance entre la conclusion de la comparaison indirecte (CI) avec l'ASMR.

B : Méthode de Bucher. R : Méta-analyse en réseau. En vert : une possible concordance. En rouge : discordance. En gris : pas de lien.

Nous pouvons supposer un possible impact de la comparaison indirecte sur le niveau d'ASMR (6 dossiers concernés, colorés en vert) lorsque celle-ci n'a pas mis en évidence de différence d'efficacité entre le médicament et ses comparateurs, et que la Commission s'est prononcée pour une absence d'amélioration du service médicale rendu (ASMR V). A l'inverse, pour le dossier PROLIA, le résultat de la comparaison indirecte et le niveau d'ASMR sont en contradiction (dossier coloré en rouge) puisque l'ASMR est de niveau IV (mineure) alors que la comparaison indirecte n'a pas révélé de différence entre le médicament et les autres molécules disponibles. Concernant le dossier ELIQUIS (coloré en gris), on ne peut pas confronter le résultat de la comparaison indirecte et le niveau d'ASMR puisque l'ASMR a été prononcée sur la base du résultat de l'essai clinique direct comparant ELIQUIS à l'énoxaparine.

Ainsi, dans 6 des 8 dossiers, la comparaison indirecte pourrait avoir joué un rôle dans l'attribution du niveau d'amélioration du service médical rendu du médicament. Toutefois, cette conclusion est à prendre avec beaucoup de précautions dans la mesure où la fixation du niveau d'ASMR repose sur les résultats de l'ensemble des études présentées dans le dossier concernant l'efficacité et la tolérance. C'est ce faisceau d'arguments qui permet à la Commission de se prononcer, et non des résultats isolés.

2.3.4.2.2 Analyse du poids des comparaisons indirectes dans les décisions de la CT au travers de deux avis rendus en 2014

Après avoir analysé attentivement les 60 avis qui composent notre étude, il nous a paru évident que la CT donne davantage de crédit aux résultats des comparaisons directes *versus* un comparateur actif quand elles existent, qu'aux résultats de comparaison indirecte. Afin de s'en convaincre, voici deux avis récents emblématiques de ce constat.

❖ Avis TECFIDERA⁸⁶ (diméthyl fumarate), 7 mai 2014

Il s'agit de la demande d'inscription sécurité sociale et collectivités de TECFIDERA (diméthyl fumarate), nouveau médicament administré par voie orale, indiqué en traitement de fond chez les patients adultes atteints de sclérose en plaques de forme rémittente. Le dossier du laboratoire a reposé notamment sur :

- deux études de phase III, DEFINE et CONFIRM randomisées en double aveugle, de supériorité versus placebo ayant évalué l'efficacité du dimethyl fumarate pendant deux ans chez des patients atteints de SEP-RR ; L'étude CONFIRM comportait un bras contrôle actif traité par acétate de glatiramère.
- une méta-analyse en réseau dans laquelle l'efficacité et la tolérance du dimethyl fumarate a été comparée à celle des autres traitements de la SEP-RR ;

Dans l'étude DEFINE, le pourcentage estimé de patients avec poussées (critère principal) a été de 27 % dans le groupe traité par le dimethyl fumarate versus 46 % dans le groupe placebo (HR : 0,51 ; IC 95 % [0,40 à 0,66]). Dans l'étude CONFIRM, le taux annualisé de poussées (critère principal) a été significativement diminué par rapport au placebo : 0,22 (IC 95 % [0,18 à 0,28]) dans le groupe dimethyl fumarate versus 0,40 (IC 95 % [0,33 à 0,49]) dans le groupe placebo ($p < 0,02$). L'analyse groupée des deux études a permis de mettre en évidence une diminution du risque de progression du handicap à 2 ans avec une progression de l'EDSS confirmée à 12 et à 24 semaines. L'étude CONFIRM comportait un bras de validation actif traité par acétate de glatiramère. Dans le cadre d'une comparaison réalisée *a posteriori*, il n'a pas été observé de différence d'efficacité entre le dimethyl fumarate et l'acétate de glatiramère sur le taux annualisé de poussée (ratio du taux annualisé de poussées : 0,78 ; IC 95 % [0,59 à 1,05]) et le risque de progression du handicap (HR : 0,85 ; IC 95 % [0,56 à 1,29]).

L'objectif de la méta-analyse en réseau était de comparer de façon indirecte l'efficacité et la tolérance de dimethyl fumarate 240 mg x 2/jour à celles des autres traitements immunomodulateurs de la SEP-RR. Au total, 27 études randomisées ont été incluses dans l'analyse. Il s'agissait majoritairement d'études versus placebo avec 18 études versus placebo et 9 études incluant des comparaisons directes entre deux traitements immunomodulateurs. Les deux études ayant évalué le dimethyl fumarate étaient les études DEFINE et CONFIRM. Ainsi, la seule comparaison directe disponible entre le dimethyl fumarate et un traitement actif était celle versus l'acétate de glatiramère issue de l'étude CONFIRM. Le réseau de comparaison pour le taux annualisé de poussées était fourni et est présenté sur la Figure 32.

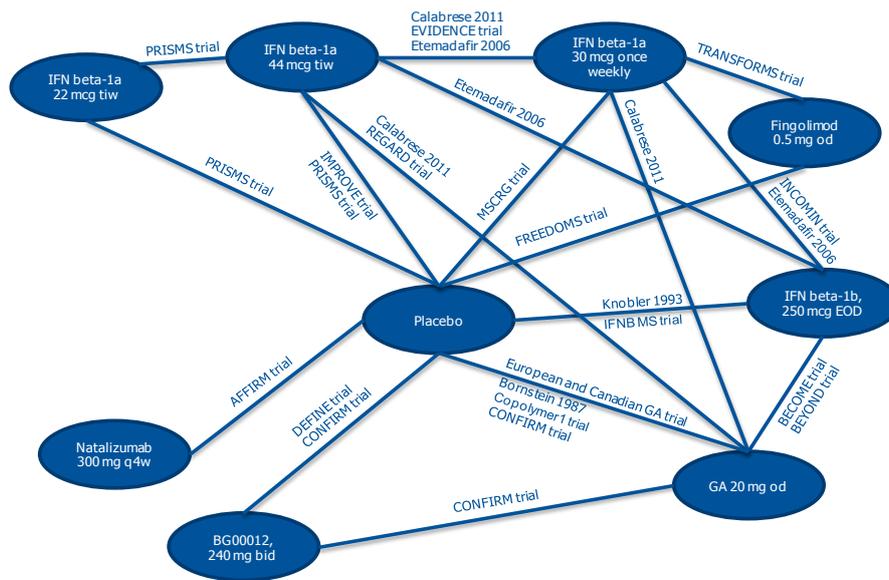


Figure 32 : Réseau de comparaisons de la MTC pour le taux annualisé de poussées dans l’avis TECFIDERA du 07 mai 14.

Dans cette méta-analyse, l’efficacité du dimethyl fumarate 240 mg x 2/jour sur le taux annualisé de poussées a été supérieure à celle du placebo, des interférons bêta-1a et bêta-1b, de l’acétate de glatiramère et du tériflunomide. Aucune différence n’a été observée sur le risque de progression du handicap entre le dimethyl fumarate et les autres traitements.

Bien que cette méta-analyse ait été jugée de bonne qualité méthodologique, la CT ne l’a pas considérée comme preuve suffisante seule pour conclure à une différence : « *L’interprétation des résultats de cette méta-analyse en réseau doit rester prudente compte-tenu du peu de comparaisons directes disponibles et de la disparité des caractéristiques entre les études. Cette méta-analyse en réseau bien que de bonne qualité ne peut remplacer les conclusions issues d’études de comparaison directe bien conduites.* »

Aussi, la CT s’est prononcée en faveur d’une absence d’ASMR, malgré les résultats positifs de la comparaison indirecte : « *En l’absence d’une étude de supériorité versus un traitement actif, TECFIDERA n’apporte pas d’amélioration du service médical rendu (ASMR V, inexistante) dans la prise en charge de la sclérose en plaques de forme rémittente récurrente. Les résultats de la méta-analyse en réseau montrant par comparaison indirecte une diminution du taux annualisé de poussées avec TECFIDERA 240 mg x 2/jour comparativement aux interférons bêta, à l’acétate de glatiramère et au tériflunomide ne peuvent être*

considérés comme suffisants pour conclure à l'efficacité supérieure de TECFIDERA à la dose de 240 mg x 2/jour par rapport à ces traitements.»

❖ Avis TIVICAY⁸⁷ (dolutégravir), 28 mai 2014

TIVICAY (dolutégravir) est un médicament antirétroviral (ARV) appartenant à la classe des inhibiteurs d'intégrase (INI). Le dossier concerne la demande d'inscription de TIVICAY 50 mg sur la liste des spécialités remboursables aux assurés sociaux et agréées aux collectivités et divers services publics. Il comporte notamment 6 études de phase III réalisées chez des patients adultes naïfs de traitement ou prétraités ainsi qu'une méta-analyse en réseau dont l'objectif était de comparer l'efficacité et la tolérance de TIVICAY par rapport aux autres 3^{èmes} agents de référence en initiation de traitement chez les patients naïfs : inhibiteurs de protéase boostés, inhibiteurs non nucléosidiques de la transcriptase inverse, et INI.

Il est difficile de conclure sur un éventuel poids de la méta-analyse en réseau dans le niveau d'ASMR de TIVICAY car les résultats de cette méta-analyse n'ont pas été repris dans les conclusions de l'avis. Cependant, on peut lire, dans le chapitre « Autres données » de l'avis, le passage suivant : *« Selon les auteurs, « cette méta-analyse en réseau suggère que le dolutégravir pourrait avoir un profil favorable par rapport aux autres 3^{èmes} agents de référence (ATV/r [ritonavir-boosted atazanavir], LPV/r [ritonavir-boosted lopinavir], RPV[rilpivirine], and EVG/c[cobicistat-boosted elvitegravir]) même si les études de comparaisons directes restent la référence pour évaluer les différentes options de traitement. En résumé, cette analyse met en évidence le dolutégravir comme une option de première ligne pour les patients naïfs de traitement antirétroviral. »*

2.4 Résumé et discussion

Notre étude a sélectionné, grâce aux mots clés « indirecte » et « réseau », les avis mis à disposition dans une base de données interne au SEM entre juillet 2009 et octobre 2014 comportant des comparaisons indirectes. A partir de ces avis relevés, nous avons confirmé un recours croissant aux comparaisons indirectes par les laboratoires. Cependant, au vue de la méthodologie employée (extraction de données issues des avis et non des dossiers déposés par les laboratoires), cela peut aussi montrer que les avis détaillent de plus en plus les données cliniques mises à dispositions par les laboratoires dans leur dossier.

Au travers de cette étude rétrospective, nous avons établi une typologie de la méthodologie des comparaisons indirectes utilisées par les laboratoires dans les dossiers déposés auprès de la Commission de la transparence de 2009 à 2014. La méta-analyse en réseau s'est révélée la méthode la plus employée, conformément aux recommandations de la HAS de 2009⁴. La comparaison naïve, méthode non appropriée pour réaliser une comparaison indirecte en raison des biais auxquels elle expose (notamment exposition aux biais de sélection et confusion), était encore très utilisée de 2009 à 2013. Son utilisation semble cependant se raréfier progressivement, au profit de la méta-analyse en réseau. Quelques aires thérapeutiques sont davantage concernées par le recours aux comparaisons indirectes (immunosuppresseurs, antinéoplasiques, antithrombotiques), mais contrairement à ce que l'on pouvait attendre, il ne s'agit pas de celles où la pharmacopée est la plus développée. Il semble que les laboratoires font davantage appel à ces méthodes lors de l'inscription d'un nouveau médicament.

Environ 50% des comparaisons indirectes relevées dans les avis ont été jugées invalides et inacceptables d'un point de vue méthodologique par la CT. Ceci a été principalement motivé par l'hétérogénéité, suivie des critiques sur la méthodologie de la comparaison indirecte elle-même. Sur ce dernier point, on peut noter que de nombreux reproches ont été fait quant à l'insuffisance de puissance des comparaisons indirectes pour montrer une supériorité d'un traitement par rapport à un autre ; ainsi que l'absence de vérification de l'indépendance des résultats par rapport aux *a priori* et des *a priori* non informatifs avec la méthode Bayésienne. Les comparaisons naïves d'estimations ponctuelles, notamment à partir de méta-analyses classiques, semblent désormais reconnues et récusées. Toutes les critiques de validité semblent encore de nos jours freiner l'exploitation des résultats de comparaisons indirectes pour l'évaluation de l'efficacité et de la tolérance du médicament en vue de son accès au remboursement. Ainsi, seulement 18% des comparaisons indirectes relevées dans notre étude étaient mentionnées dans les conclusions de l'avis, ce qui laisse à penser la faible part de ces méthodes sur les conclusions de la CT.

Bien qu'il ait été difficile de rattacher les résultats des comparaisons indirectes valides avec un niveau d'ASMR, on peut remarquer que sur les 61 comparaisons indirectes recueillies, seulement 6 (soit environ 10%) étaient valides et ont permis de hiérarchiser les

traitements, et allaient dans le sens de l'ASMR attribué, sans pouvoir dire si elles ont été l'élément décisionnel de l'avis de la CT. Comme le montre l'avis TIVICAY, les études de comparaisons directes restent la référence pour évaluer la valeur relative des différentes options de traitement en 2014.

3 Conclusion

La méta-analyse en réseau, même de bonne qualité, ne semble pas pouvoir remplacer les conclusions issues d'études de comparaison directe bien conduites. Notamment, il semble encore prématuré de pouvoir conclure à une efficacité supérieure sur la seule base d'une comparaison indirecte. Celle-ci, quand elle permet de conclure (ce qui est assez rare du fait du manque de puissance), trouve plutôt sa place comme « validation » de la cohérence externe d'une comparaison directe. Les comparaisons indirectes se positionnent alors comme des informations complémentaires permettant de conforter, quand cela est possible, la démonstration faite par une comparaison directe bien conduite.

Ainsi nous venons de considérer le rôle de la Commission de la transparence dans l'évaluation de la pertinence de la prise en charge du médicament par la sécurité sociale, notamment au vu de son SMR ainsi que de l'ASMR qu'il est susceptible d'apporter par rapport aux traitements déjà disponibles (évaluation médico-technique), et la place limitée des comparaisons indirectes dans cette évaluation. Toutefois, il n'y a de pertinence, au sens global du terme, en matière d'admission au remboursement, que si l'efficacité du médicament est aussi prise en compte⁷⁷. En conséquence, la détermination de l'efficacité a été récemment confiée à la HAS dans le cadre de sa mission médico-économique et, selon les cas, l'évaluation médico-économique des médicaments est conduite par la Commission d'évaluation économique et de santé publique (CEESP) parallèlement à l'évaluation médico-technique de la CT pour ajouter une dimension nouvelle à ses avis. Aussi, nous nous sommes interrogés sur la place que pouvaient occuper les comparaisons indirectes dans l'évaluation médico-économique des médicaments par la CEESP.

PARTIE 3

Place des comparaisons indirectes dans l'évaluation médico-économique des médicaments.

1 L'évaluation médico-économique du médicament en France

Longtemps fondée sur l'efficacité clinique, l'évaluation du médicament, étape indispensable pour son accès au marché, intègre depuis peu une nouvelle composante, un critère d'efficience, qui passe par le calcul d'un ratio différentiel coût-résultat.

1.1 Principes et objectifs de l'évaluation médico-économique.

L'évaluation médico-économique permet d'évaluer le résultat de santé et les coûts générés par un produit de santé, comparativement à l'ensemble de ses alternatives médicalement pertinentes (Figure 33).

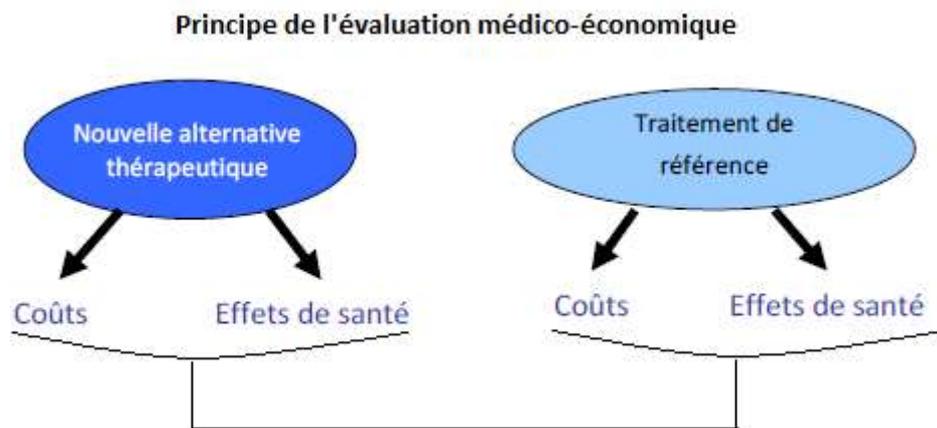


Figure 33 : Principe de l'évaluation médico-économique.

Lorsque le produit de santé est plus efficace et moins cher que ses comparateurs, l'interprétation est évidente ; le produit est efficient. En revanche, lorsqu'il est plus efficace et plus cher, l'interprétation du résultat du calcul économique en termes d'efficience n'est plus immédiate et la question de la qualification de l'efficience se pose.

L'évaluation médico-économique se traduit par un ratio différentiel coût-résultat (RDCR ou *incremental cost effectiveness ratio* – ICER en anglais) qui est un rapport entre le différentiel de coût et le différentiel d'efficacité entre les deux stratégies comparées (Figure 34).

Le ratio différentiel coût-résultat (RDCR) dans l'évaluation médico-économique

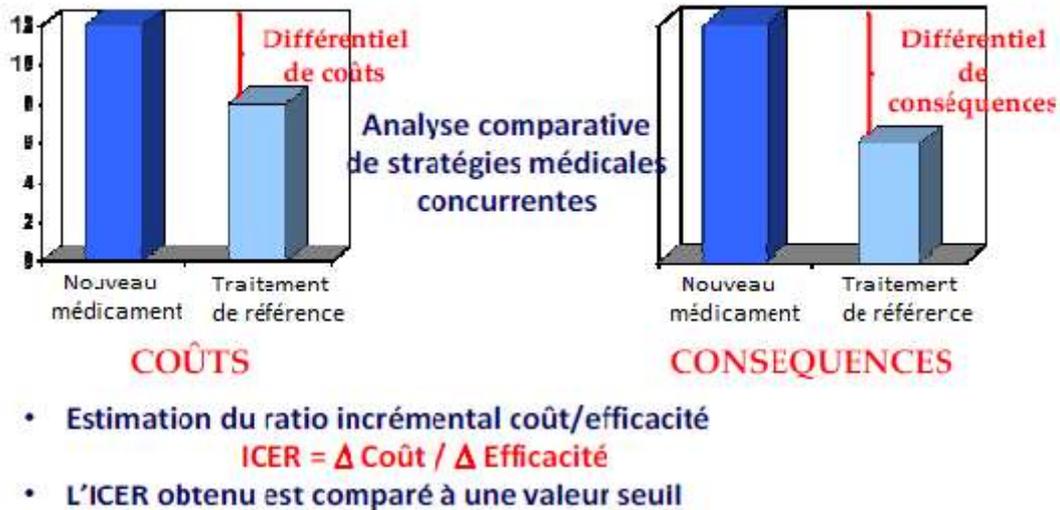


Figure 34 : Le ratio différentiel coût-résultat (RDCR) dans l'évaluation médico-économique.

Le RDCR s'exprime en coût par année de vie gagnée ou en coût par année de vie pondérée par la qualité de vie («*Quality-Adjusted Life Years*», QALY). Le QALY est un indicateur économique qui permet de rassembler les gains provenant de la morbidité (gain en qualité) et la réduction de la mortalité (gain en quantité)⁸⁸. Il consiste à multiplier le temps passé dans un état de santé donné par un coefficient (l'utilité) rendant compte de la valeur accordée à cet état. L'utilité est un score de 0 (mort) à 1 (parfait état de santé) qui correspond aux préférences élicitées dans différentes populations pour des états de santé donnés, grâce à des questionnaires (questionnaire EQ-5D notamment). Ainsi, une année au cours de laquelle l'intervention thérapeutique permet de prolonger l'espérance de vie effective mais affecte les conditions de vie (par exemple, en évitant le décès au prix d'un handicap) sera comptée entre 0 et 1. Par exemple, un an en bonne santé (utilité égale à 1) est équivalent à $1 \times 1 = 1$ QALY ; de la même manière, deux années dans un état malade (supposons une utilité égale à 0.5) sont équivalentes à $2 \times 0.5 = 1$ QALY.

Dans de nombreux pays, il existe une règle d'interprétation du RDCR fondée sur l'identification d'une valeur de référence : le produit de santé est dit efficace (ou coût-efficace) si le RDCR est inférieur à cette valeur. L'Organisation mondiale de la santé (OMS), par exemple, estime ainsi un produit rentable si le RDCR est inférieur à trois fois le produit intérieur brut du pays considéré⁸⁹. Cependant, ce principe de comparaison du RDCR à une

valeur seuil pour décider du remboursement selon un principe binaire (efficient versus non efficient) n'est pas appliqué en France, comme nous le verrons plus loin.

L'objectif de l'évaluation médico-économique est de hiérarchiser les différentes options de traitement envisageables en vue d'une allocation optimale des ressources⁹⁰. La question à laquelle tente de répondre une évaluation médico-économique peut se verbaliser ainsi : « Etant donné le bénéfice clinique et le coût supplémentaire de ce médicament par rapport à la stratégie de référence, peut-on considérer que l'utilisation de ce médicament est coût-efficace ? ». En France, l'évaluation médico-économique est un outil supplémentaire d'aide à la décision en matière de politique de santé publique. Elle vise à éclairer les décideurs qui doivent arbitrer entre différents produits de santé susceptibles d'être financés collectivement, dans un contexte où coexistent des besoins en santé potentiellement illimités, et des ressources budgétaires limitées. La croissance continue des dépenses de santé et plus particulièrement des dépenses liées aux produits de santé, en lien avec des prix très élevés, sont en partie à l'origine de l'introduction réglementaire de l'évaluation médico-économique dans le dispositif de décision de prix⁷⁶, comme nous allons le voir dans le chapitre qui suit.

1.2 Le contexte de l'évaluation médico-économique en France

Alors qu'elle existe depuis les années 1990 en Royaume-Uni, l'évaluation médico-économique des produits de santé innovants et coûteux n'a été introduite en France qu'en 2012 dans les procédures de fixation du prix.

1.2.1 Le décret du 2 octobre 2012

La loi de financement de la Sécurité sociale (LFSS) du 21 décembre 2011 explicite que la HAS « réalise ou valide notamment les études médico-économiques sur les stratégies de soins, de prescription ou de prises en charges les plus efficaces ». Le décret du 2 octobre 2012 (n°2012-1116, article R.161-71-1 du code de la sécurité sociale), applicable un an après sa publication, précise les conditions de cette évaluation médico-économique des produits de santé. L'évaluation médico-économique devient alors requise dans le cadre :

- d'une procédure d'inscription ou de renouvellement d'inscription sur la liste des médicaments remboursés par la Sécurité sociale ou sur la liste d'agrément aux collectivités,
- si une ASMR I, II ou III (majeure, importante ou modérée) est sollicitée ou obtenue par le laboratoire^m,
- et que le produit (ou la technologie) est susceptible d'avoir un « impact significatif » sur les dépenses de l'assurance maladie.

L'estimation de l'impact du produit sur les dépenses tient elle-même compte « *de son incidence sur l'organisation de soins, les pratiques professionnelles ou les conditions de prise en charge des malades et, le cas échéant, de son prix.* ». Après concertation entre la HAS, le CEPS et les industriels, il a été décidé⁹¹ que l'impact sur les dépenses serait considéré comme significatif dès lors que le chiffre d'affaires annuel lié au produit est estimé égal ou supérieur à 20 millions d'euros, la deuxième année de pleine consommation pour les primo-inscription ou l'année précédant la demande pour les renouvellements, ou qu'un impact significatif du produit sur les dépenses d'assurance maladie est revendiqué par l'industriel.

L'évaluation médico-économique telle que décrite par le décret du 2 octobre 2012 est confiée à la Commission d'Evaluation Economique et de Santé Publique de la HAS. Le décret précise le rôle et les objectifs de cette commission qui doit notamment rendre un avis d'efficience sur les médicaments et les dispositifs médicaux.

1.2.2 La Commission évaluation économique et de santé publique (CEESP) de la HAS et les avis d'efficience

1.2.2.1 La Commission évaluation économique et de santé publique (CEESP) de la HAS

La Commission évaluation économique et santé publique a été créée à l'occasion de la mise en œuvre de la Loi de financement de la sécurité sociale pour 2008 qui a confié à la HAS une mission médico-économique sur les stratégies de santé. Cette mission a été confirmée et élargie par la Loi de financement de la sécurité sociale pour 2012, en incluant les produits de santé et créant une commission réglementée pour mener à cette mission. Le

^m Un projet visant à étendre l'évaluation de l'efficience aux ASMR IV est récemment évoqué⁹⁰.

Collège de la HAS, qui détermine la composition de cette commission, a choisi de conserver la composition et le fonctionnement de la CEESP. Elle est ainsi composée de trente-trois membres permanents choisis par le Collège de la HAS pour leur compétence dans le domaine de l'évaluation économique et de la santé publique, ayant une voix délibérative, dont un Président nommé parmi les membres du Collège de la HAS et deux vice-présidents. Les membres de la Commission actuelle, nommés pour un mandat de trois ans, sont issus d'horizons géographiques, professionnels et disciplinaires variés : économistes, cliniciens, épidémiologistes, spécialistes en santé publique, en sciences humaines et sociales, représentants d'associations de patients et d'usagers. Des avis d'efficience (efficience prévisible, pour la première inscription d'un produit ou d'une stratégie de santé, et efficience constatée lors de sa réévaluation) sont validés par la Commission à l'attention du CEPS et publiés sur le site de la HAS. Pour rendre ses avis, la CEESP s'appuie sur les travaux du service évaluation économique et santé publique ainsi que sur les comptes rendu des deux sous-commissions. Les sous commissions sont composées chacune d'une partie des membres de la CEESP en fonction de leur domaine d'expertise. Concernant l'activité non programmée :

- La sous-commission économie est systématiquement mobilisée dans l'expertise technique des modélisations déposées par les industriels.
- La sous-commission sciences humaines et sociales peut également être mobilisée, à la demande d'un membre de la CEESP, dans l'expertise des dossiers lorsque le produit évalué soulève des enjeux particuliers (éthiques, sociaux...).

Le travail de la CEESP, pour les avis rendus dans le cadre de l'appréciation de l'efficience d'un produit de santé pour lequel un remboursement a été demandé, s'appuie sur trois types de fondations :

- Un cadre de référence qui renseigne les industriels sur les attentes de la HAS vis-à-vis des modèles d'évaluation qu'ils peuvent utiliser pour apprécier l'efficience de leur produit. Ces recommandations sont regroupées dans un document intitulé « *Choix méthodologique pour l'évaluation médico-économique de la HAS* », publié en 2011⁹² (nous le détaillerons dans le chapitre 1.3) ;

- Une organisation délibérative reposant sur un examen en sous-commission technique des dossiers d'évaluation suivi par leur examen et discussion au cours de la réunion mensuelle de la Commission qui donnera son avis ;
- Un examen uniquement du dossier déposé par l'industriel : la CEESP ne construit pas de nouveau «modèle économique» (ce terme sera explicité au chapitre 1.3.5).

La procédure formalisée et publique qui entoure l'avis d'efficience comprend quatre étapes :

- 1) dans un premier temps, le dossier soumis à l'avis de la CEESP par l'industriel doit être déclaré recevable par le collège de la HAS, au vu des conditions citées dans le décret de 2012 ;
- 2) l'analyse méthodologique du modèle médico-économique utilisé par le laboratoire est ensuite effectuée par les chefs de projets du SEESP avec possibilité de poser des questions à l'industriel, selon une procédure formalisée ;
- 3) un projet d'avis d'efficience est validé par la CEESP ;
- 4) l'avis définitif est rendu au vu des observations éventuelles de l'industriel, et adressé au CEPS.

1.2.2.2 Les avis d'efficience

L'avis d'efficience est très synthétique dans sa forme. Il présente dans un premier temps l'objectif et le contexte de l'étude, dans un second temps la conclusion sur la conformité méthodologique de l'étude selon les recommandations du guide HAS 2011⁹² puis dans un troisième temps la conclusion de la CEESP sur l'efficience⁷⁶. L'analyse critique détaillée est annexée à l'avis (annexe 2 de l'avis). Les réserves qui peuvent être émises sur chacun des points expertisés par la CEESP sont listées dans l'annexe 4 et sont hiérarchisées en trois niveaux :

- Réserve mineure (-) : élément de non-conformité aux recommandations en vigueur, cependant l'impact attendu sur les conclusions est faible ;
- Réserve importante (+) : élément de non-conformité aux recommandations en vigueur, avec un impact attendu important sur les conclusions, en termes d'incertitude ;
- Réserve majeure (++) : élément de non-conformité aux recommandations en vigueur qui invalide tout ou partie de l'étude économique.

Cinq cas de figure peuvent se présenter selon les conclusions de la CEESP :

- la méthode de l'étude est considérée comme conforme aux recommandations de la HAS, aucune réserve n'est soulevée ;
- la méthode de l'étude est considérée comme acceptable avec des réserves mineures au regard des recommandations de la HAS, pouvant générer une incertitude faible sur le résultat de l'étude ;
- la méthode de l'étude est considérée comme acceptable malgré des réserves importantes. Ces réserves ne remettent pas en cause la validité de l'étude, mais elles augmentent l'incertitude attachée au résultat. En effet, elles questionnent sur la fiabilité des résultats quantitatifs énoncés (RDCR) ou l'interprétation qui en est proposée par l'auteur en termes d'efficacité ;
- la méthode de l'étude est considérée comme non conforme, en raison de réserves majeures qui invalident l'étude.
- la méthode de l'étude est considérée comme non suffisamment documentée pour que la CEESP puisse apprécier sa conformité. Cela revient à conclure à la non-conformité de l'étude.

Les avis d'efficacité ne portent aucun jugement sur le niveau d'efficacité du produit examiné (c'est-à-dire sur la valeur du RDCR), ils jugent uniquement la conformité de l'analyse médico-économique présentée et donc la validité du RDCR. Si aucune réserve majeure n'a été émise, la conclusion de la CEESP mentionne le résultat quantitatif de l'évaluation médico-économique, c'est-à-dire le RDCR. Une seule réserve majeure suffit à déclarer qu'une appréciation valable ne peut pas être portée sur la valeur d'efficacité du produit pour lequel le remboursement est demandé, ce qui aura ensuite des répercussions sur la fixation du prix du produit par le CEPS.

1.2.3 La place de l'évaluation médico-économique dans l'admission au remboursement

Depuis le 2 octobre 2013 (date de mise en application du décret du 2 octobre 2012), la HAS rend deux avis distincts (l'un rendu par la CT, l'autre par la CEESP) lorsqu'un laboratoire pharmaceutique souhaite que son produit fasse l'objet d'un remboursement par l'assurance maladie une fois qu'il a obtenu une AMM et qu'il répond aux conditions fixées par le décret (Figure 35).

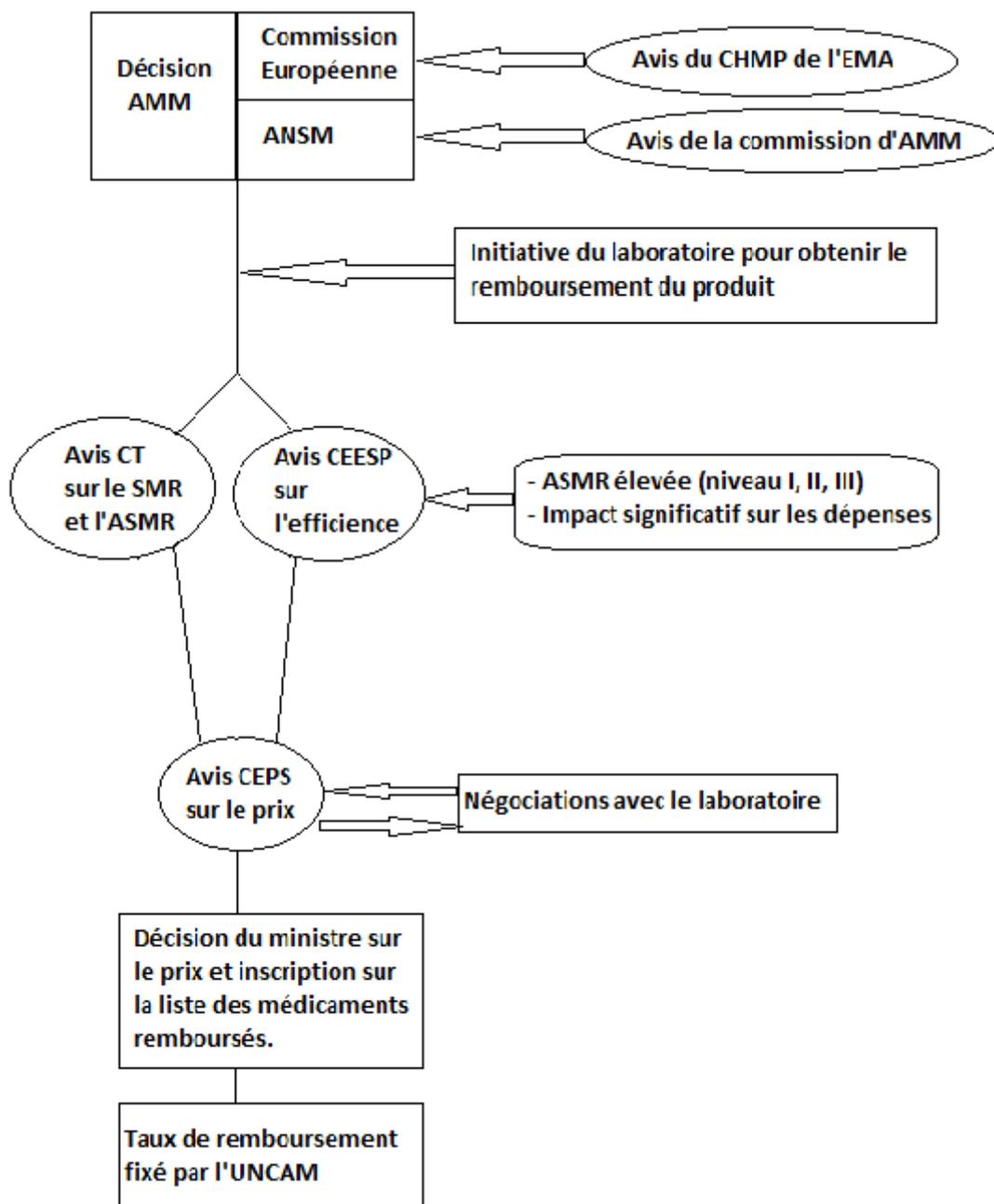


Figure 35: Procédure de remboursement et de fixation du prix pour les médicaments⁹⁰.

Le délai de traitement d'un dossier, c'est-à-dire son examen par la CT et la CEESP, est de 90 jours à compter de sa date de réception par la HAS, ceci afin de respecter le délai total d'évaluation de 180 jours donnés à la HAS et au CEPS pour la mise sur le marché d'un médicament. La HAS s'est engagée à respecter ce délai réglementaire, ce qui impose un parallélisme des deux évaluations, à la fois, médico-technique et médico-économique.

L'analyse médico-économique n'intervient pas directement pour l'admission au remboursement, mais elle est prise en compte lors de la négociation du prix des produits de santé par le CEPS. L'avis de la CEESP n'est d'ailleurs pour l'instant rendu public sur le site internet de la HAS qu'après la prise de décision du CEPS sur le prix du produit de santé. En cas de réserve majeure dans l'avis d'efficacité, le CEPS peut fixer un prix éventuellement inférieur au plafond européen malgré une ASMR I à III. Le recul n'est pas suffisant pour connaître l'influence des avis d'efficacité dans le processus de négociation des prix.

Les avis d'efficacité ont vocation à aider les décideurs (CEPS, ministère) à affiner leurs prises de décision concernant le prix. Ils ne constituent pas un avis de remboursement ou non comme c'est le cas en Royaume-Uni où le National Institute for Clinical Excellence (NICE) considère que le produit est efficace et donc admissible au remboursement si son RDCR est compris entre 20.000 et 30.000 livres/QALY (avec certaines exceptions). En France, aucune valeur seuil de référence pour le RDCR n'est à ce jour spécifiée. Ainsi, il est possible de déterminer le coût du gain en santé produit par une innovation mais il n'est pas possible de dire si ce coût est acceptable pour la collectivité. La HAS n'est pas légitime pour édicter seule cette valeur, mais elle a récemment souhaité contribuer à cette réflexion au moyen d'une revue de la littérature publiée en décembre 2014⁹³ permettant de faire un état des connaissances sur ce sujet. La vice-présidente de la CEESP a récemment expliqué que la commission souhaiterait « évoluer vers une qualification des RDCR », même en l'absence de valeur de référence au-delà de laquelle les pouvoirs publics refuseraient de rembourser un médicament⁹⁴.

1.3 Les points clés d'une évaluation médico-économique

Nous allons détailler dans cette partie certains points clés d'une évaluation médico-économique qui nous seront ensuite utiles pour analyser les données de l'étude rétrospective que nous avons menée sur un certain nombre d'avis d'efficacité (chapitre 2). Certains de ces points seront complétés au regard des recommandations dictées par la HAS dans son guide publiée en 2011⁹² et intitulé « *Choix méthodologiques pour l'évaluation économique à la HAS* » qui explicite son cadre de référence méthodologique en matière

d'évaluation économiqueⁿ. L'exhaustivité des recommandations de la HAS est résumée en Annexe 3.

1.3.1 Le choix du type d'étude

Il existe différents types d'étude médico-économique suivant la manière dont l'efficacité du médicament est rapportée (Figure 36). Elle peut être rapportée en unité physique telle que le nombre d'années de vies gagnées, le nombre de récurrences évitées etc. et dans ce cas il s'agit d'une « analyse coût-efficacité » (ACE). L'efficacité peut également être mesurée par le nombre d'année de vies pondérées par la qualité de vie ou QALY (*Quality Adjusted Life Years*), on parle alors d'une « analyse coût-utilité » (ACU). Lorsque l'efficacité est exprimée elle aussi en unité monétaire, on parle d'« analyse coût-bénéfice » (ACB). Ces trois types d'analyses sont regroupés sous le terme « analyses coût-résultat » (ACR). Dans une « minimisation de coûts » on fait l'hypothèse d'une égalité d'efficacité entre les stratégies comparées et on met en concurrence uniquement les coûts (ce cas sera évoqué dans le chapitre 2 au travers de l'avis d'efficience OSPUMIT). Cette méthode est considérée comme une sous-catégorie de l'évaluation coût-résultat en situation d'équivalence entre les résultats.

Type d'étude	Coût	Efficacité	Exemples
Cout-efficacité	€	Unités physiques	années de vie gagnées, nombre de récurrences évitées, nombre de diagnostics corrects etc ...
Coût-utilité	€	QALY (Quality Adjusted Life Years)	Années de vie pondérées par la qualité de vie
Minimisation des coûts	€	Hypothèse d'équivalence d'efficacité	-
Coût-bénéfice	€	Unités monétaires	Bénéfices exprimés en valeurs monétaires (€) (Peu utilisé en France)

Figure 36 : Les différents types d'évaluation médico-économiques.

ⁿ La publication en 2016 d'une deuxième version du guide méthodologique de l'évaluation économique de la HAS a été annoncée.

Parmi les méthodes disponibles, la HAS privilégie les ACR, et plus précisément l'analyse coût-utilité et l'analyse coût-efficacité qui sont, à ce jour, les plus abouties d'un point de vue méthodologique et les plus pratiquées. Le guide précise les modalités suivantes :

- *« Si la qualité de vie liée à la santé est identifiée comme une conséquence importante des interventions étudiées, l'analyse de référence est de type coût-utilité et le critère de résultat à privilégier est la durée de vie ajustée sur la qualité. L'analyse coût-utilité est systématiquement accompagnée d'une analyse coût-efficacité qui utilise la durée de vie sans pondération comme critère de résultat.*
- *Si la qualité de vie liée à la santé n'est pas identifiée comme une conséquence importante des interventions étudiées, l'analyse de référence est de type coût-efficacité et le critère de résultat à privilégier est la durée de vie.*

1.3.2 Le choix de la population d'analyse

La population d'analyse est défini dans le guide comme « *l'ensemble des individus dont la santé est affectée par les interventions étudiées de manière directe ou de manière induite* » (le terme « induite » défini par exemple l'entourage, comme c'est le cas avec les vaccins). De manière générale la population d'analyse est donc la population de l'indication de l'AMM du médicament. En présence d'une hétérogénéité sur les résultats, le guide précise que la démarche économique peut justifier une analyse de sous-populations spécifiques.

1.3.3 Le choix du comparateur et le recours aux comparaisons indirectes

L'évaluation économique étant une démarche comparative, elle implique le choix de comparateurs. Selon le guide, « *l'analyse de référence identifie toutes les interventions qui sont concurrentes avec l'intervention évaluée* ». Cette recommandation est parfois difficile à respecter, étant donné la possibilité d'un nombre élevé de comparateurs et le manque de comparaisons directes. Un choix pourra alors être réalisé mais les arguments sur lesquels repose cette sélection doivent être justifiés. Une autre solution consiste à avoir recours à des comparaisons indirectes. Le guide précise qu' « *En l'état actuel des connaissances, deux approches peuvent être mises en avant : la comparaison indirecte ajustée, simple et*

transparente, et les méthodes bayésiennes en réseau. Cette dernière semble la mieux à même de dessiner la hiérarchie complète du champ thérapeutique étudié. Les comparaisons dites « naïves » d'estimations ponctuelles ou de bras actifs issus d'essais contrôlés randomisés distincts ne sont pas admissibles. »

1.3.4 L'horizon temporel

L'horizon temporel est la période de temps durant laquelle les données d'efficacité et les données de coût vont être relevées et analysées pour le calcul du RDCR. Il dépend de l'histoire naturelle de la maladie, de la chronologie des interventions, de l'occurrence des résultats associés aux interventions et de l'occurrence des coûts. La HAS recommande d'adopter un horizon temporel suffisamment long pour intégrer l'ensemble des différentiels de coût et de résultat attendus. Cet horizon temporel peut correspondre à la durée de l'étude clinique à partir de laquelle sont recueillies les données d'efficacité et de coûts. Cependant, il arrive bien souvent que la durée de l'étude soit trop courte et le recours à une « modélisation » est alors nécessaire.

1.3.5 La modélisation

Lorsque les données issues des essais cliniques ne sont pas suffisantes (comparaisons directes non disponibles, horizon temporel trop court etc.) il est possible de recourir à une modélisation qui est une représentation simplifiée et formalisée d'un phénomène réel que l'on cherche à analyser. Les modèles sont des outils analytiques issus d'autres disciplines appliqués au domaine de l'évaluation économique. Ils ont pour but de synthétiser les données disponibles (issues des essais cliniques, d'études observationnelles ou d'avis d'experts) pour estimer les données de coût et d'efficacité des médicaments en situation d'information imparfaite. Le modèle de Markov, couramment utilisé, permet de modéliser l'évolution d'une cohorte de patients dans le temps (Figure 37). Les patients se trouvent toujours dans un des états de santé préalablement définis par l'analyste (ex : état malade, état survie sans progression, état décès etc.). Cette cohorte transite à chaque cycle d'un état à un autre (évolution qui correspond à un évènement clinique majeur) selon une probabilité de transition. Un nombre déterminé à l'avance de cycles de Markov définit l'horizon temporel du modèle.

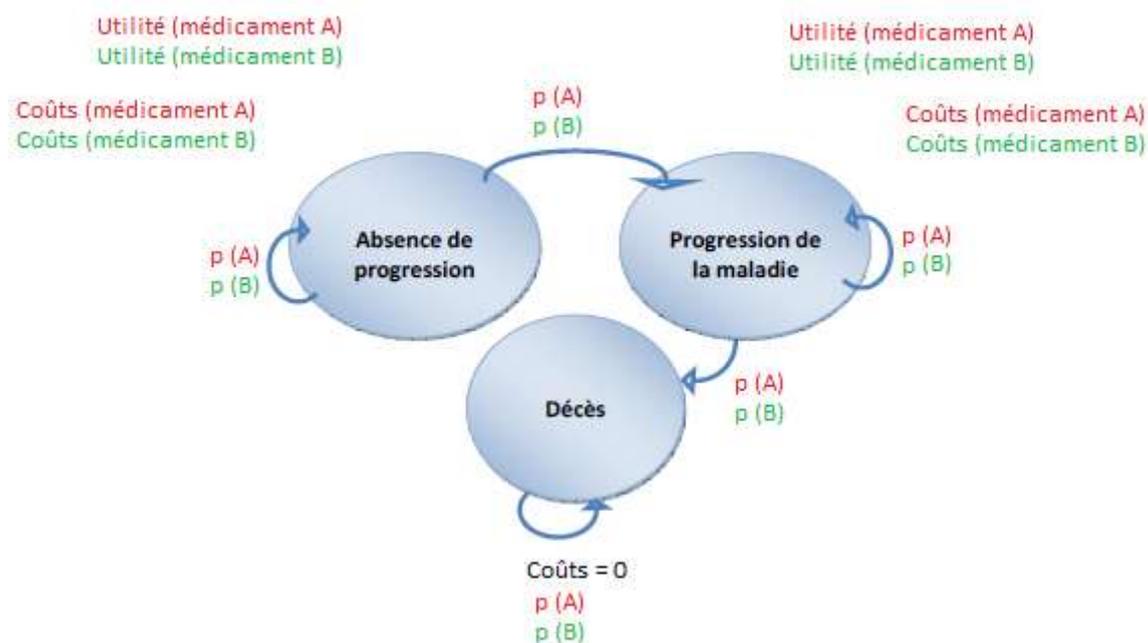


Figure 37 : Exemple d'un modèle de Markov.

Ce modèle est constitué de 3 états de santé (« Absence de progression », « Progression de la maladie » et « Décès ») associés chacun à des coûts et une utilité. Les probabilités de transition sont notées $p(A)$ et $p(B)$.

Les valeurs des probabilités de transitions sont définies à partir du pourcentage d'efficacité de chaque médicament (par exemple survie sans progression et survie globale) retrouvé dans les essais cliniques ou dans la littérature. Lorsque ces données ne sont pas issues du même essai qui a comparé A et B de manière directe, il est nécessaire d'effectuer une comparaison indirecte pour obtenir des probabilités de transition « ajustées » versus placebo pour chaque traitement. Une fois que le modèle a « tourné » durant n cycles définis au préalable, on additionne pour chaque traitement les coûts et les données d'utilités, ce qui permet de calculer le RDCR.

La HAS préconise l'utilisation d'un modèle, et de justifier les choix méthodologiques relatifs à la structure du modèle.

1.3.6 L'expression des résultats

La représentation de l'efficacité des interventions étudiées dans un repère coût-résultat est recommandée par la HAS. Dans un tel système orthonormé, l'axe des abscisses correspond au coût net total de l'intervention par patient et l'axe des ordonnées à la valeur

attribuée pour représenter les effets de la technologie sur la santé (mortalité, morbidité, qualité de vie, selon les cas) (Figure 38). Chaque point représente une thérapie. Tous les points qui se situent en bas à droite d'un autre (c'est-à-dire dès que la thérapie représente plus de coûts pour moins de bénéfiques, comme c'est le cas pour l'intervention 2 par rapport à l'intervention 4 sur la Figure 38) sont dit dominés. La notion de dominance stricte est élargie à la dominance généralisée : une intervention est dominée au sens de la dominance généralisée si elle est strictement dominée par une combinaison, supposée linéaire en coût et en résultat, de deux autres interventions. Ainsi, sur la Figure 38 l'intervention 5 est dominée au sens de la dominance généralisée car elle est strictement dominée par des combinaisons des interventions 6 et 7. Les points non dominés sont reliés et forment la frontière d'efficience. Celle-ci peut être définie comme l'ensemble des points qui relient les interventions existantes telles qu'aucun des points n'est dominant et telles que la frontière n'est pas redéfinie.

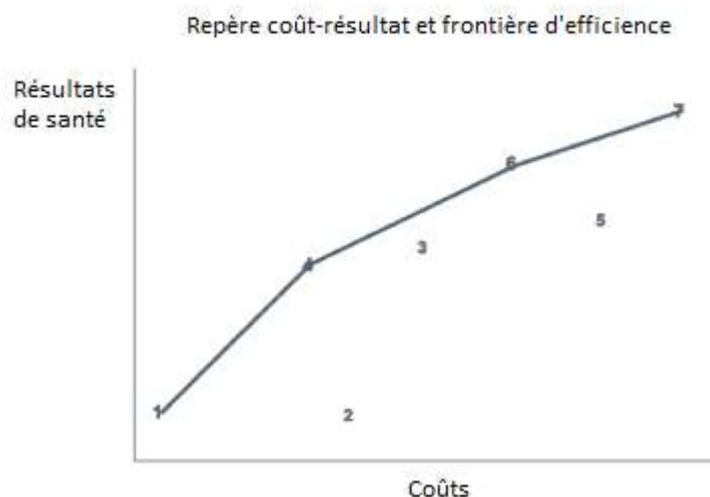


Figure 38 : Représentation graphique de la frontière d'efficience dans un repère coût-résultat. La frontière d'efficience est symbolisée par la ligne brisée. L'intervention 2 est strictement dominée par l'intervention 4. L'intervention 5 est dominée au sens de la dominance généralisée car elle est strictement dominée par la combinaison des interventions 6 et 7.

Sur la Figure 38, la ligne brisée visualise la frontière d'efficience au sens de la dominance généralisée : toutes les interventions situées au-dessous de la frontière d'efficience sont dominées. Les interventions situées sur la frontière d'efficience (au sens de la dominance généralisée) sont qualifiées d'efficentes *a priori*. Toutefois, les RDCR, calculés

pour deux interventions consécutives sur la frontière d'efficacité, augmentent lorsque le coût et le résultat des interventions augmentent. S'il existe un seuil d'acceptabilité (ce qui n'est pas le cas en France comme nous l'avons déjà vu), les interventions situées sur la frontière, mais dont le RDCR (par rapport à l'intervention précédente) est supérieur au seuil, sont en principe rejetées. L'intervention efficace, compte tenu du seuil, est celle dont le RDCR est inférieur à celui-ci.

La HAS précise que les interventions de santé constituant la frontière d'efficacité doivent être identifiées et un RDCR doit être calculé pour chacune d'elles, avec le détail des différentiels de coût et de résultat.

1.3.7 Les analyses de sensibilités

La caractérisation de l'incertitude fait partie de l'évaluation économique. La HAS demande que les sources d'incertitude suivantes soient identifiées : l'incertitude sur les paramètres, l'incertitude sur la structure du modèle et l'incertitude sur les choix méthodologiques. Aussi, elle précise que :

- la réalisation d'une analyse de sensibilité « probabiliste » doit être privilégiée lorsque les distributions théoriques ou empiriques des paramètres sont connues ou peuvent être estimées ;
- une analyse de sensibilité « déterministe univariée » doit être systématiquement réalisée sur les paramètres susceptibles d'influencer les conclusions du modèle.

Une analyse de sensibilité permet de tester la robustesse des résultats de RDCR aux variations des données d'efficacité et/ou de coût. Le plus souvent les deux types d'analyses sont réalisés.

1.3.7.1 Analyse de sensibilité déterministe

Dans une analyse de sensibilité déterministe, on choisit les paramètres susceptibles de faire varier le RDCR, on sélectionne pour chacun un intervalle de valeurs plausibles, et on observe pour chaque valeur les variations du RDCR qu'elles engendrent. Le degré de variation du résultat donne une indication sur la sensibilité du paramètre. Si le RDCR reste inchangé, cela renforce la robustesse du résultat de l'évaluation. Une analyse de sensibilité

déterministe est dite univariée lorsqu'un seul paramètre est ajusté alors que les autres paramètres restent constants (à la différence d'une analyse multivariée où l'on teste l'incertitude de toutes les variables simultanément). La représentation graphique d'une analyse de sensibilité déterministe univariée est le graphique de Tornado, qui est une forme spécifique d'histogramme permettant de comparer l'incertitude générée par chacun des paramètres, les autres paramètres étant fixés à leur valeur retenue dans l'analyse dite de référence. Dans le diagramme de Tornado illustré sur la Figure 39, l'axe y rouge représente la mesure du RDCR lorsque tous les paramètres sont fixés à la valeur retenue dans l'analyse de référence, qui est indiquée entre crochets. Les barres horizontales représentent la valeur prise par le ratio différentiel lorsque l'on fait varier les valeurs du paramètre entre deux bornes, qui sont indiquées entre parenthèses.

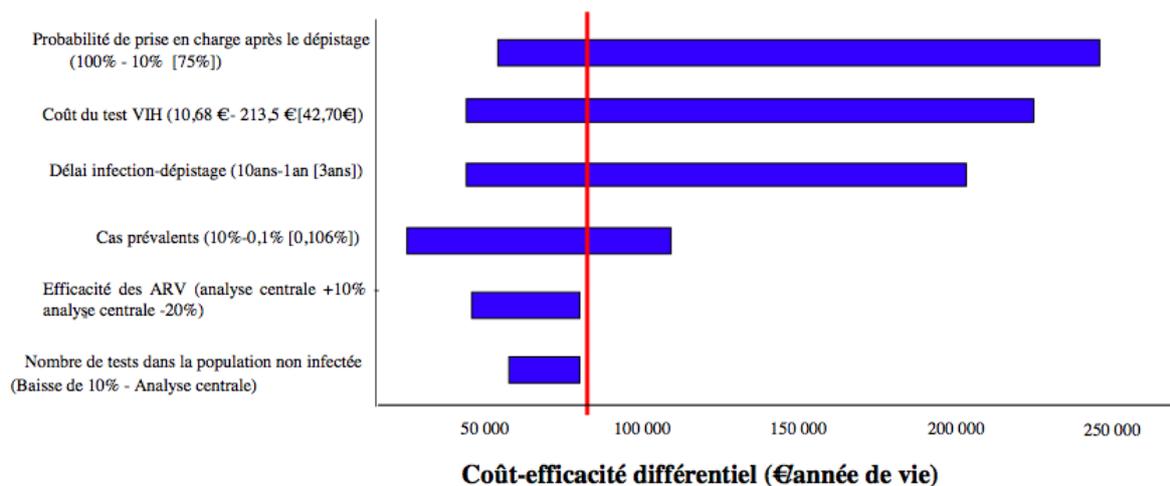


Figure 39: Présentation d'une analyse de sensibilité univariée sous la forme d'un graphique de Tornado⁹⁵.

1.3.7.2 Analyse de sensibilité probabiliste

Dans une analyse de sensibilité probabiliste, on attribue à chaque paramètre susceptible de faire varier le RDCR, une distribution de probabilité rendant compte de l'incertitude sur le paramètre. Le calcul du RDCR se fait n fois (généralement 1 000) en tirant à chaque fois au sort une valeur dans la distribution de probabilité de chaque paramètre. Il est ainsi obtenu la distribution du RDCR après avoir pris en compte l'incertitude sur tous les paramètres, à partir de laquelle peuvent être calculées différentes statistiques comme la

moyenne et les probabilités que le RDCR soit inférieur à une valeur donnée. L'analyse probabiliste permet donc de faire varier simultanément tous les paramètres sur lesquels une incertitude existe et d'estimer la variation du RDCR quand toute l'incertitude du modèle est quantifiée et prise en compte dans son calcul. Le résultat de l'analyse probabiliste peut être présenté sous la forme d'un nuage de points, chacun de ces points étant une possibilité concrète du RDCR (Figure 40). Plus la zone de répartition des points est dense, plus l'on peut supposer que le rapport coût-résultat effectif est dans cette zone.

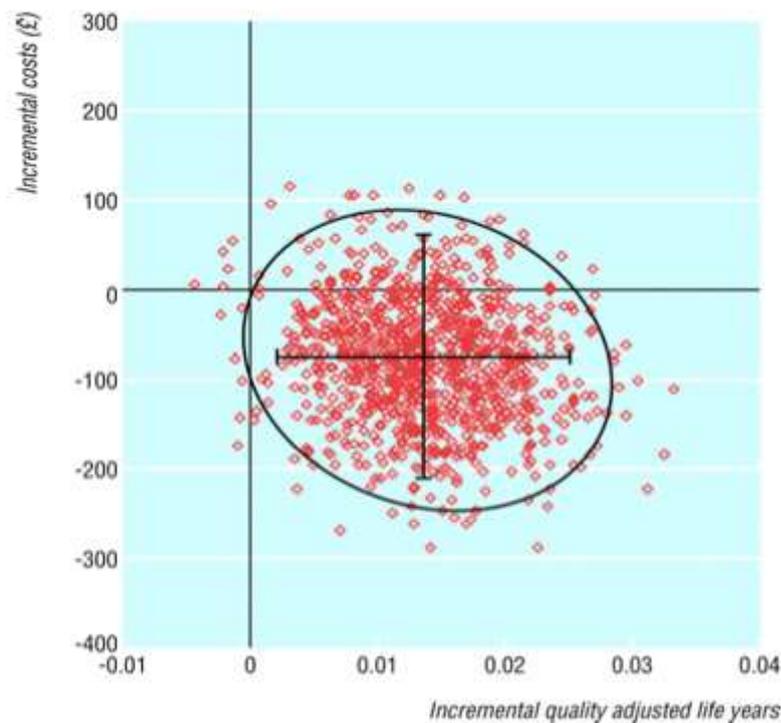


Figure 40: Présentation d'une analyse de sensibilité probabiliste sous la forme d'un nuage de points et d'ellipse de confiance⁹⁶.

Dans l'exemple ci-dessus, les intervalles d'incertitude associés respectivement au différentiel de coût et de résultat sont représentés par les traits verticaux et horizontaux. L'ellipse représente l'intervalle d'incertitude à 95 % associé au rapport coût-résultat. L'intervention évaluée domine le comparateur (elle est plus efficace et moins chère), mais il existe une incertitude concernant les coûts puisque l'intervalle d'incertitude à 95% inclut la possibilité que le coût de l'intervention évaluée soit supérieur au coût du comparateur.

La courbe d'acceptabilité permet également d'illustrer le résultat d'une analyse probabiliste. Elle illustre graphiquement la problématique du choix d'une intervention par rapport à une autre en fonction du seuil d'acceptabilité que l'on retient. Étant donné une valeur seuil définie comme le ratio différentiel maximal acceptable (axe X), la courbe d'acceptabilité indique la probabilité (axe Y) que le ratio différentiel réel soit inférieur à cette valeur (Figure 41).

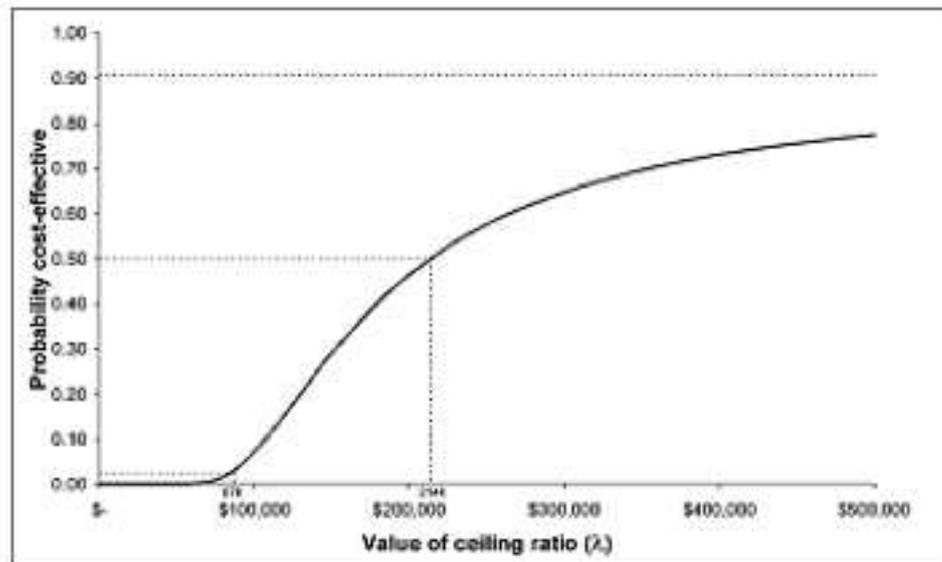


Figure 41 : Présentation de l'incertitude par une courbe d'acceptabilité⁹⁷.

Dans l'exemple ci-dessous, on observe que le ratio différentiel réel associé à l'intervention évaluée a une probabilité de 50% d'être en dessous de \$200 000. La probabilité d'être en dessous d'un seuil de \$100 000 est nulle : le ratio différentiel réel de l'intervention n'est jamais inférieur à ce seuil.

Depuis la mise en place du décret, une quinzaine d'avis d'efficience ont été rendus par la CEESP. Compte-tenu de l'obligation de prise en compte de l'exhaustivité des comparateurs dans les évaluations médico-économiques (voir chapitre 1.3.3), nous avons souhaité évaluer l'importance du recours aux comparaisons indirectes par les laboratoires pharmaceutiques, ainsi que l'appréciation de la CEESP sur ces méthodes. Notre étude rétrospective présentée dans le chapitre qui suit a considéré 14 avis d'efficience.

2 Place des comparaisons indirectes dans l'évaluation médico-économique des médicaments : étude rétrospective

1.1 Objectifs

Les objectifs de notre étude rétrospective sont les suivants :

- évaluer, depuis la mise en place du décret du 2 octobre 2012, le pourcentage d'avis d'efficience ayant nécessité le recours à des comparaisons indirectes et les indications thérapeutiques concernées.
- appréhender les raisons pour lesquelles une comparaison indirecte a été ou n'a pas été utilisée.
- Pour les avis sans comparaisons indirectes : apprécier le caractère exhaustif des comparateurs considérés dans l'analyse d'efficience via les réserves émises par la CEESP vis à vis des stratégies comparées.
- Pour les avis avec comparaisons indirectes : relever les critiques méthodologiques et éventuellement les réserves qui ont été émises par la CEESP à l'égard des comparaisons indirectes présentées par les industriels.

1.2 Méthodes

Nous avons considéré les avis d'efficience enregistrés dans la base de données interne du Service Evaluation Economique et Santé Publique (SEESP) de la HAS et clos en date du 15 décembre 2014. Dans chaque avis d'efficience, les données suivantes ont été relevées :

- indication thérapeutique ;
- stratégies comparées ;
- utilisation ou non d'une comparaison indirecte et méthodologie employée ;
- prise en compte ou non de tous les comparateurs, et si non, quels étaient les comparateurs manquants ;
- qualité de la comparaison indirecte : critiques méthodologiques et réserves éventuelles.

1.3 Résultats

1.3.1 Nombre d'avis relevés

Entre octobre 2013 (date de mise en œuvre du décret du 2 octobre 2012), et le 15 décembre 2014, soit une durée de 1 an et 2 mois, nous avons listés **14 avis d'efficience** clos. Ces avis concernaient les spécialités suivantes (par ordre chronologique) : DEFITELIO (défibrotide), KADCYLA (trastuzumab emtansine), LEMTRADA (alemtuzumab), ZOSTAVAX (vaccin vivant atténué du virus varicelle-zona), SOVALDI (sofosbuvir), TIVICAY (dolutégravir), VECTIBIX (panitumumab), TECFIDERA (diméthyle fumarate), OPSUMIT (macitentan), ROTARIX (vaccin à rotavirus vivant atténué), ROTATEQ (vaccin à rotavirus vivant atténué), OLYSIO (siméprévir), ADEMPAS (riociguat), et XOFIGO (dichlorure de radium-223).

En date de l'analyse, 5 des 14 avis étaient rendus publics sur le site de la HAS (KADCYLA⁹⁸, LEMTRADA⁹⁹, SOVALDI¹⁰⁰, TIVICAY¹⁰¹ et ADEMPAS¹⁰²).

1.3.2 Nombre d'avis concernés par le recours à une comparaison indirecte

Sur les 14 avis relevés, seulement 5 avis (OPSUMIT, ADEMPAS, KADCYLA, LEMTRADA et TECFIDERA) mentionnaient le recours à une comparaison indirecte, soit 36% des avis d'efficience relevés (Figure 42).

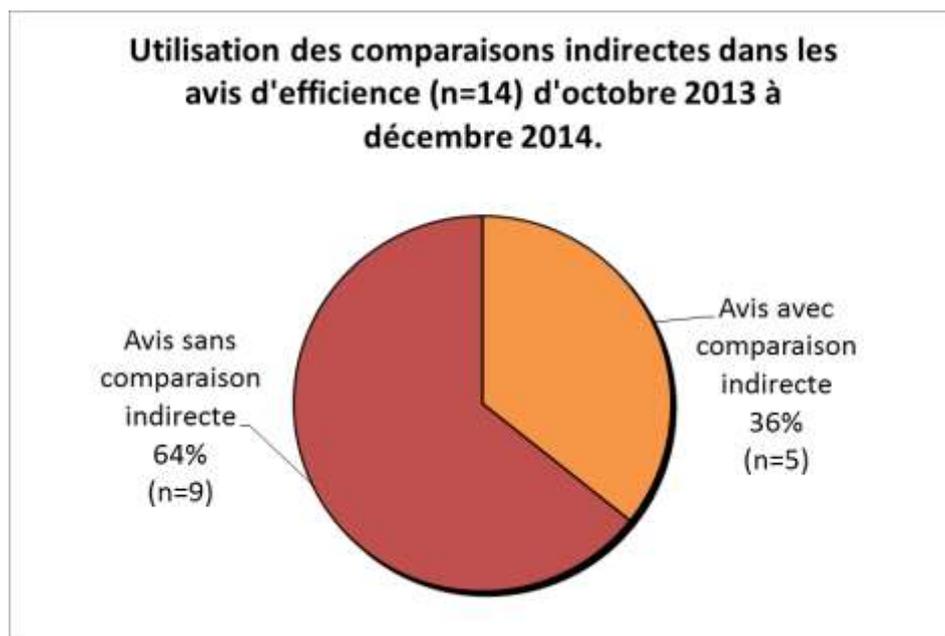


Figure 42 : Utilisation des comparaisons indirectes dans les avis d'efficience (n=14) d'octobre 2013 à décembre 2014.

Nous avons souhaité comprendre les raisons qui pouvaient motiver un laboratoire à utiliser ou ne pas utiliser la méthodologie des comparaisons indirectes dans son avis d'efficacité, et les conséquences que cela pouvaient avoir sur les réserves émises par la CEESP concernant les stratégies comparées.

1.3.3 Analyse des avis ne contenant pas de comparaison indirecte

Le but de cette analyse est d'appréhender les éléments qui peuvent dispenser un laboratoire de faire appel à une comparaison indirecte pour prendre en compte tous les comparateurs existants. Nous évaluerons également le caractère exhaustif des comparateurs considérés au moyen des réserves émises à propos des stratégies comparées.

1.3.3.1 Analyse quantitative

Les 9 avis sans comparaisons indirectes ont été considérés, puis pour chacun, nous avons relevé les motifs pour lesquels il n'y a pas eu de comparaison indirecte. Ces motifs sont les suivants (Figure 43) :

- trop d'hétérogénéité pour pouvoir réaliser une comparaison indirecte : 5 avis (SOVALDI, OLYSIO, XOFIGO, TIVICAY, ROTATEQ), soit 36% des avis et 56% des avis sans comparaisons indirectes ;
- pas de comparateurs actifs et/ou données de comparaisons directes disponibles : 4 avis (ROTARIX, ZOSTAVAX, DEFITELIO, VECTIBIX), soit 28% des avis et 44 % des avis sans comparaisons indirectes.

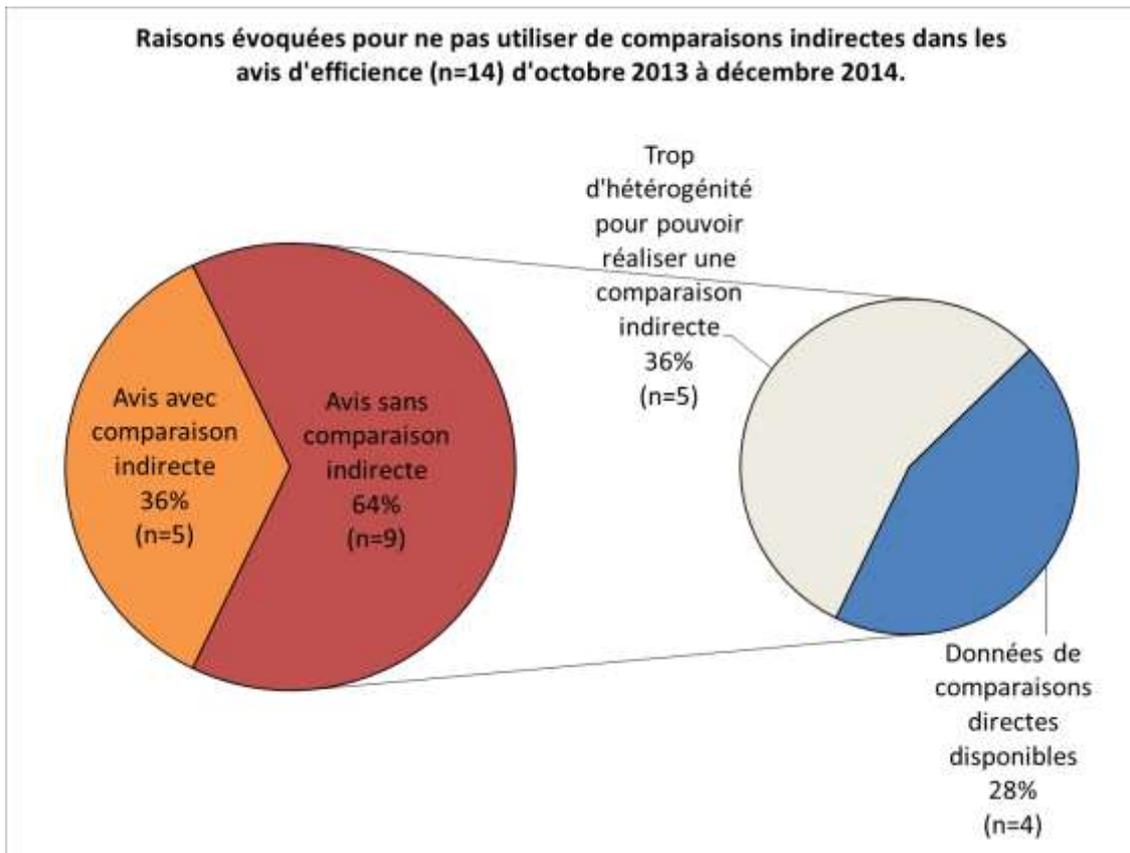


Figure 43 : Raisons évoquées pour ne pas utiliser de comparaisons directes (n=14) dans les avis d'efficience d'octobre 2013 à décembre 2014.

Notre allons à présent envisager individuellement chaque avis d'efficience.

1.3.3.2 Analyse qualitative

L'analyse des 9 avis d'efficience sans comparaison indirecte est résumée dans le Tableau 10.

date	dossier	DCI	Indication	Stratégies comparées	Raisons invoquées pour ne pas utiliser une CI	Prise en compte des tous les comparateurs? Si non : comparateurs manquants	Réserve(s) en rapport avec les stratégies comparées		
							mineure	importante	majeure
15/04/2014	SOVALDI	sofosbuvir	Hépatite C	nombreuses	HETEROGENEITE, pas assez d'essais comparatifs.	NA	NA	Robustesse des données d'efficacité (données non comparatives et effectifs faibles dans les essais cliniques)	NA
14/10/2014	OLYSIO	siméprévir	Hépatite C	nombreuses	HETEROGENEITE, pas assez d'essais comparatifs.	NA	NA	Robustesse des données d'efficacité de l'étude COSMOS (étude de phase Ila, données non comparatives, effectifs faibles)	NA
22/07/2014	ROTARIX	vaccin à Rotavirus humain vivant atténué	Vaccination	absence de vaccination	Le HCSP recommande la vaccination par l'un ou l'autre des deux vaccins disponibles en les considérant de même niveau d'efficacité. L'industriel n'a donc pas retenu ROTATEQ comme stratégie alternative, estimant qu'il n'y avait pas lieu de différencier les deux vaccins dans la stratégie thérapeutique.	Il aurait été pertinent de proposer comme comparateur à ROTARIX, le vaccin ROTATEQ (concurrent de ROTARIX).	Documentation insuffisante de l'absence de prise en compte du vaccin ROTATEQ comme comparateur,	NA	NA
16/09/2014	ROTATEQ	vaccin à Rotavirus humain vivant atténué	Vaccination	absence de vaccination	HETEROGENEITE entre les essais et pas lieu de se comparer puisque les deux vaccins Rotarix et Rotateq sont équivalents	*renforcement des pratiques de prescription des solutés de réhydratation orale * promotion de l'allaitement maternel *prévention des infections nosocomiales chez les jeunes enfants hospitalisés	Absence de comparateurs actifs	NA	NA
15/04/2014	ZOSTAVAX	vaccin vivant atténué varicelle zona	Vaccination	abs de stratégie vaccinale	Stratégie vaccinale recommandée par le HCSP versus absence de stratégie vaccinale.	oui : Le choix de comparer une stratégie avec et sans vaccin est cohérent.	NA	NA	NA
11/03/2014	XOFIGO	dichlorure de radium 223	Cancer de la prostate	placebo	HETEROGENEITE des essais en termes de population et de critères d'inclusion.	non: *deux autres radioisotopes (METASTRON et QUADRAMET) : pas acceptable. Remarque: *cabazitaxel, abiraterone : pas de données directes et trop d'hétérogénéité pour indirecte. Recevable.	NA	Absence de comparateurs actifs.	NA
27/05/2014	TIVICAY	dolutégravir	VIH	raltégravir (ISENTRESS) (essai direct SAILING)	HETEROGENEITE dans le profil des patients inclus rendant difficile l'exploitation et l'interprétation des résultats de la comparaison indirecte.	▯ *elvitégravir (STRIBILD)	NA	Choix du comparateur justifié par rapport aux données disponibles, mais non représentatif de l'ensemble des alternatives thérapeutique disponibles en vie réelle.	NA
18/02/2014	DEFITELIO	déflibrotide	Maladie veino-occlusive hépatique	traitement de support	Comparaison basée sur l'étude pivot	oui	NA	NA	Les données de l'étude pivot exploitée (2005-01) ne permettent pas de montrer d'une façon adéquate l'efficacité du déflibrotide.
27/05/2014	VECTIBIX	panitumumab	Cancer colorectal métastatique	bevacizumab (AVASTIN) (essai direct PEAK)	Comparaison indirecte en cours	*cetuximab (ERBITUX)	Absence du comparateur «chimiothérapie seule»	Absence du comparateur cetuximab (ERBITUX)	Données individuelles cliniques de faible niveau de preuve pour estimer les probabilités de transition.

Tableau 10 : Résumé de l'analyse des avis d'efficience sans comparaison indirecte.

En marron : trop d'hétérogénéité pour pouvoir réaliser une comparaison indirecte. En bleu : données de comparaisons directes disponibles. En vert : prise en compte de tous les comparateurs. En rouge : manque de certains comparateurs. CI : Comparaison Indirecte.

1.3.3.2.1 Cas des traitements du VHC (avis SOVALDI¹⁰⁰ du 15/04/14 et OLYSIO du 14/10/14) : comparaison naïve des bras actifs

La problématique a été similaire dans les deux avis concernant des médicaments indiqués dans le traitement de l'hépatite C : SOVALDI (sofosbuvir) et OLYSIO (siméprévir). En effet, dans cette pathologie, la réalisation d'une comparaison indirecte s'avère très délicate en raison de la forte hétérogénéité des études cliniques disponibles eu égard aux différents génotypes du virus, aux différents stades de fibroses, aux divers statuts possibles des patients (naïf ou prétraités). De plus, les études sont souvent des études de phase II non comparatives. Pour toutes ces raisons, les données d'efficacité utilisées pour l'analyse médico-économique sont issues des divers essais cliniques disponibles avec une méthodologie qui s'apparente à une comparaison naïve des bras actifs. Cette manière de procéder a occasionné une réserve importante de la part de la CEESP qui a critiqué la « *robustesse des données d'efficacité* ». Ces deux analyses d'efficience reposent donc sur des données d'efficacité cliniques peu fiables dont il est nécessaire de tenir compte dans les analyses de sensibilité.

1.3.3.2.2 Cas des vaccins (avis ROTARIX du 22/07/14, ROTATEQ du 16/09/14 et ZOSTAVAX du 15/04/14) : comparaison à l'absence de stratégie vaccinale

Concernant les vaccins ROTARIX et ROTATEQ, tous les deux indiqués pour la prévention des gastro-entérites dues à une infection à rotavirus, ils sont considérés comme équivalents en terme d'efficacité par le HCSP (Haut Conseil de la santé publique). En l'absence de différenciation des deux vaccins, leur efficience peut être considérée comme équivalente dès lors que le coût des schémas vaccinaux est similaire. Aussi, les deux laboratoires ont choisi la même stratégie : ils ont comparé leur vaccin à l'absence de stratégie vaccinale, et non au vaccin concurrent. Ces évaluations d'efficience ne requièrent donc pas de comparaisons indirectes dans la mesure où les essais de phase III fournissent des résultats d'efficacité versus placebo (assimilé à l'absence de stratégie vaccinale).

Le laboratoire de ROTARIX a fondé son argumentaire sur l'absence de données d'efficacité comparatives entre les deux vaccins ; la CEESP a cependant critiqué, dans une

réserve mineure, une « *documentation insuffisante de l'absence de prise en compte du vaccin ROTATEQ comme comparateur* ». En ce qui concerne ROTATEQ, le laboratoire a présenté les résultats d'une méta-analyse qui n'a pas mis en évidence de différence entre ROTATEQ et ROTARIX, et a de plus évoqué une impossibilité de réaliser une comparaison indirecte en raison de l'hétérogénéité entre les essais (« *impossibilité de comparer l'efficacité clinique entre les deux vaccins à partir des essais respectivement menés avec chaque vaccin versus placebo en raison des différences constatées entre les essais en termes de pays participants ayant des contextes épidémiologiques différents, de critères d'éligibilité des populations incluses, des périodes de suivi, des définitions de la sévérité des gastro-entérite à rotavirus et des critères de jugement différents retenus*»). La CEESP a pourtant critiqué dans une réserve mineure « *l'absence de comparateurs actifs* » et a suggéré qu'il aurait pu être intéressant de voir apparaître l'une des stratégies concurrentes non médicamenteuses suivantes :

- « *renforcement des pratiques de prescription des solutés de réhydratation orale telles que préconisées par le HCSP,*
- *promotion de l'allaitement maternel telle que préconisée dans le rapport « Plan d'action Allaitement maternel » issu du Programme National Nutrition Santé,*
- *prévention des infections nosocomiales chez les jeunes enfants hospitalisés. »*

L'exemple de ROTAREQ et ROTARIX montre que la comparaison au placebo seul, même accompagnée de justifications de la part du laboratoire, ne convainc pas la CEESP quand il existe des comparateurs pertinents car ce choix limite les conclusions sur l'efficacité. Nous verrons plus loin que l'analyse de XOFIGO va dans le sens de ce constat.

En ce qui concerne le vaccin ZOSTAVAX, indiqué dans la prévention du zona et des douleurs post-zostériennes, c'est le seul vaccin dans cette indication. La construction de son analyse médico-économique n'a pas fait appel à une comparaison indirecte car les résultats des essais directs *versus* placebo étaient disponibles. Le comparateur considéré dans l'évaluation est donc l'absence de stratégie vaccinale et ce choix a été jugé acceptable par la CEESP.

1.3.3.2.3 Avis XOFIGO et TIVICAY : trop d'hétérogénéité pour réaliser une comparaison indirecte

Dans les avis XOFIGO et TIVICAY, un seul comparateur a été considéré par l'industriel au regard des données disponibles. Cependant, comme nous allons le détailler, ce choix limite la portée des conclusions en termes d'efficacité et est critiqué par la CEESP qui attribue alors une réserve importante.

❖ XOFIGO (avis du 11/03/14)

Le XOFIGO (dichlorure de radium-223) est un produit radiopharmaceutique indiqué dans le traitement du cancer de la prostate avec métastases osseuses symptomatiques et sans métastases viscérales connues. Le comparateur retenu par l'industriel dans le modèle médico-économique est le placebo, auquel le XOFIGO a été comparé dans l'essai de phase III ALSYMPCA. Selon l'industriel, il existe deux comparateurs médicalement pertinents disponibles sur le marché français dans la même indication (cabazitaxel [JEVTANA] et acétate d'abiratéron [ZYTIGA]) mais le développement clinique de ces deux traitements s'est déroulé parallèlement au XOFIGO si bien qu'il n'était pas possible de réaliser une comparaison directe. De plus, toujours selon le laboratoire, il n'était pas non plus possible de réaliser une comparaison indirecte en raison de l'hétérogénéité des essais en termes de population (proportion de patients avec des métastases osseuses et/ou des métastases viscérales différentes) et des critères d'évaluation (définition de la survie sans progression différente notamment (Figure 44).

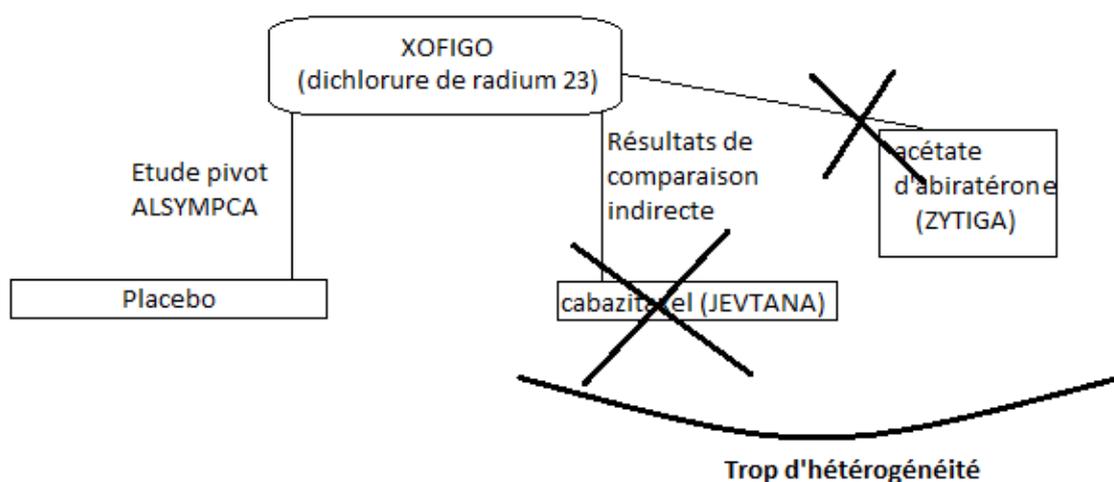


Figure 44 : Problème posé dans le choix des comparateurs dans l'avis d'efficacité XOFIGO (avis du 11/03/2014).

La CEESP a estimé que sur le fondement de ces éléments, il n'était effectivement « *pas possible, compte-tenu des données disponibles, de comparer directement ou indirectement le dichlorure de radium-233 au cabazitaxel et à l'abiratéron* ». Néanmoins, elle a reproché au laboratoire de ne pas avoir effectué une comparaison versus le chlorure de strontium 89Sr (METASTRON) et le diphosphonate marqué au samarium 153Sm (QUADRAMET), deux radio-isotopes indiqués dans le traitement des douleurs causées par les métastases osseuses du cancer de la prostate, ou bien de ne pas en avoir tenu compte dans le cadre des traitements palliatifs. En effet, ces deux traitements étaient disponibles à la date de réalisation de l'essai ALSYMPCA. La CEESP a précisé que « *le fait que l'efficacité démontrée de ces deux traitements soit limitée à un effet antalgique, comme l'évoque les industriels, n'est pas un argument recevable. Il est en effet pertinent de comparer un traitement ayant un impact sur la qualité de vie et sur la survie d'une part avec un traitement n'ayant un impact que sur la qualité de vie d'autre part (plutôt que le placebo) dès lors que des données sur des critères de jugement comparables sont disponibles et que les traitements sont « en concurrence » avec l'intervention étudiée (HAS, 2011).* »

La CEESP réaffirme ainsi la limite forte à l'évaluation réalisée si le comparateur choisi est uniquement le placebo. En effet, dans la pratique, il n'est pas envisageable dans la plupart des cas de proposer à ces patients un traitement palliatif uniquement. La CEESP insiste sur l'importance du choix du comparateur dans la mesure où l'absence d'un comparateur est susceptible de modifier de façon très importante le niveau de RDCR calculé. Pour ces raisons, la Commission a donné au XOFIGO une réserve importante au regard de « *l'absence de comparateurs actifs* ».

❖ TIVICAY¹⁰¹ (avis du 27/05/14)

Le dolutégravir (TIVICAY) est un inhibiteur de l'intégrase du virus de l'immunodéficience humaine (VIH), indiqué dans le traitement de l'infection par le VIH en association avec d'autres médicaments antirétroviraux. L'étude évalue l'efficacité de dolutégravir en comparaison au raltégravir (ISENTRISS, un autre inhibiteur de l'intégrase du VIH), en association à un traitement de fond optimisé (TFO). Selon l'industriel, le choix du comparateur est fondé sur le fait que seul un essai clinique, l'essai SAILING (essai clinique de non infériorité comparant l'efficacité de dolutégravir versus raltégravir), a pu être identifié. Une revue de la littérature a cependant été mise en œuvre par le laboratoire dans la

perspective de conduire des comparaisons indirectes à partir d'une méta-analyse en réseau (la méta-analyse en réseau évoquée faisait partie du dossier déposé à la CT et cette dernière n'en a pas tenu compte dans ses conclusions, comme nous l'avons vu dans la partie 2, chapitre 2.3.4.2.2). Les auteurs précisent que le réseau identifié est disjoint et que les études présentent une très forte hétérogénéité dans le profil des patients inclus, notamment dans l'essai SAILING et celui comparant elvitégravir (VITEKTA, inhibiteur de l'intégrase du VIH) au raltégravir, rendant difficile l'exploitation et l'interprétation des résultats de la comparaison indirecte. Par ailleurs, les auteurs précisent que l'intégration de l'élvitégravir, non commercialisé à ce jour, nécessiterait l'élaboration d'hypothèses génératrices d'incertitudes. Enfin, concernant les inhibiteurs de protéase, les auteurs expliquent, qu'outre les difficultés posées par les comparaisons indirectes, le fait que ces traitements soient inclus dans le traitement de fond optimisé rendrait l'interprétation des résultats difficiles s'ils étaient également considérés comme des comparateurs de dolutégravir (Figure 45).

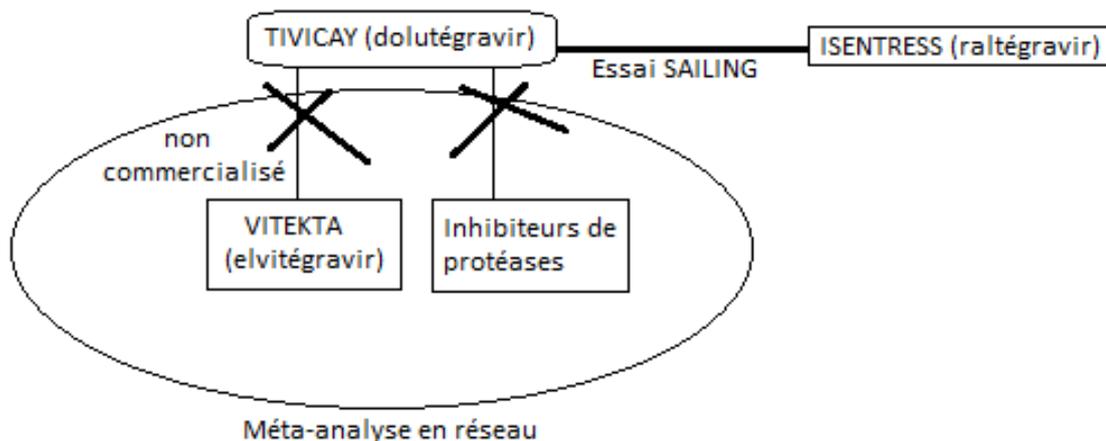


Figure 45 : Stratégies comparées considérées ou non dans l'avis d'efficacité TIVICAY (27/05/2014).

Selon la CEESP, tout comme dans l'avis XOFIGO, si le choix du comparateur est justifié au regard des données disponibles, la prise en compte d'un seul comparateur va nécessairement limiter la portée des conclusions du modèle par rapport aux options de traitement existantes en vie réelle. Aussi, une réserve importante a été émise en ce qui concerne le « *choix du comparateur justifié par rapport aux données disponibles, mais non représentatif de l'ensemble des alternatives thérapeutiques disponibles en vie réelle.* ».

1.3.3.2.4 Avis DEFITELIO et VECTIBIX : un seul comparateur, évaluation fondée sur la comparaison directe disponible

❖ DEFITELIO (avis du 18/02/14)

DEFITELIO (défibrotide) est un médicament orphelin indiqué dans le traitement de la maladie veino-occlusive (MVO) hépatique sévère^o dans le cadre de la transplantation de cellules souches hématopoïétiques. Il n'existe pas d'alternative thérapeutique au défibrotide dans le traitement de la MVO. Aussi, pour l'analyse d'efficacité, la stratégie alternative envisagée par le laboratoire est la pratique courante dans l'indication (intitulée «*Best Supportive Care*» ou « traitement de support»). Ce choix a été jugé cohérent avec les recommandations méthodologiques par la CEESP, dans la mesure où le comparateur représente bien la prise en charge actuelle de la MVO hépatique sévère et qu'il n'existe pas d'autres alternatives. L'analyse d'efficacité se fonde donc sur les résultats de l'étude pivot de phase III (étude 2005-01) évaluant l'efficacité du défibrotide dans le traitement de la MVO hépatique sévère en comparaison aux données d'efficacité d'une cohorte historique de patients non traités.

Ainsi, dans le cas du défibrotide, le laboratoire n'a pas eu recours à une comparaison indirecte dans la mesure où la seule alternative à ce traitement est le traitement de support, et que l'étude de phase III fournissait les données d'efficacité nécessaires. Ce choix a été jugé acceptable par la CEESP. On peut cependant noter que la qualité des données cliniques a été jugée insuffisante (étude de phase III avec une cohorte historique, dont le critère principal de jugement est évalué à l'issue d'une période de 100 jours) ce qui a occasionné une réserve majeure : «*Les données de l'étude pivot exploitée (2005-01) ne permettent pas de montrer d'une façon adéquate l'efficacité du défibrotide.*». Le RDCR revendiqué par l'industriel n'a donc pas été validé.

❖ VECTIBIX (avis du 27/05/14)

L'étude expertisée évalue le RDCR du panitumumab (VECTIBIX) versus le bevacizumab (AVASTIN), en 1ère ligne de traitement et en association à un protocole de

^o La MVO est une atteinte des cellules endothéliales sinusoidales hépatiques ; cette complication est principalement engendrée par la chimiothérapie myéloblastive administrée aux patients avant la transplantation de cellules souches hématopoïétiques. Le défibrotide jouerait un rôle tant dans la protection des cellules endothéliales que dans la restauration de l'équilibre thrombo-fibrinolytique.

chimiothérapie (mFOLFOX 6), chez des patients atteints de cancer colorectal métastatique exprimant le gène RAS non muté (type sauvage). Cette restriction d'utilisation aux patients avec un statut RAS non muté a fait l'objet d'une demande de modification de l'AMM, car le VECTIBIX possédait déjà une indication dans le traitement des patients atteints d'un cancer colorectal métastatique.

La non-prise en compte des autres alternatives thérapeutiques dans le modèle est justifiée ainsi par l'industriel :

- *Concernant les protocoles de chimiothérapie non ciblée* : le rapport précise que cette alternative n'est pas retenue comme pertinente au vu des recommandations thérapeutiques actuellement en vigueur, et à la lumière des faibles parts de marché. La CEESP a fait remarquer que la chimiothérapie non ciblée a pourtant été choisie par l'industriel comme comparateur dans l'essai PRIME^P sur lequel s'est appuyée la demande de modification de l'AMM auprès de l'EMA ainsi que la demande de modification des conditions d'inscription auprès de la CT en faveur d'une ASMR III. De plus, elle a rappelé que l'argument sur les parts de marché n'est en soi pas suffisant : ce n'est pas parce qu'une stratégie est moins utilisée en pratique qu'elle n'est pas sur la frontière d'efficacité. Les recommandations françaises précisent qu'une chimiothérapie ou la biothérapie seule sont recommandées uniquement chez des patients spécifiques, qui ne représentent donc pas l'ensemble des patients RAS sauvage. Il aurait été ainsi justifié, selon la CEESP, de considérer cette option dans une analyse en sous-groupe. L'absence du comparateur « chimiothérapie seule » est reprochée à travers une réserve mineure.
- *Concernant le cetuximab (ERBITUX), de la même classe thérapeutique que panitumumab* : cette alternative n'est pas retenue dans l'analyse en raison de la non disponibilité des données chez les patients avec un statut RAS non muté, au moment de la préparation du dossier par l'industriel. Celui-ci précise cependant qu'une comparaison indirecte est en cours, s'appuyant sur toutes les données pertinentes publiées concernant la population RAS non muté. La CEESP a fait remarquer que cela

^P L'étude PRIME a évalué l'efficacité et la tolérance de VECTIBIX en association au protocole FOLFOX-4 versus FOLFOX-4 seul en première ligne de traitement.

aurait justifié une demande de suspension du dossier afin de présenter une évaluation médico-économique complète. L'absence du comparateur cetuximab (ERBITUX) a occasionné une réserve importante.

Au total, dans l'analyse d'efficacité de VECTIBIX (panitumumab), les données d'efficacité utilisées sont issues d'une analyse en sous-groupe de l'essai PEAK, essai randomisé de phase II réalisé en ouvert ayant évalué VECTIBIX en association au protocole FOLFOX-6 versus AVASTIN en association au protocole FOLFOX-6 dans la population K-RAS non muté (Figure 46).

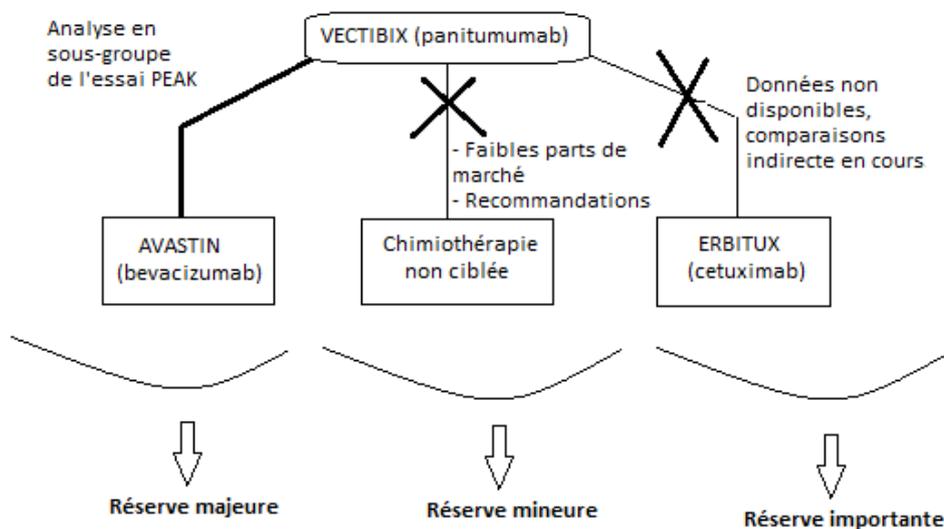


Figure 46 : Stratégies comparées considérées ou non dans l'avis d'efficacité VECTIBIX (27/05/2014) et réserves occasionnées.

Cependant, s'agissant de l'analyse en sous-groupe d'un essai en ouvert, la force de la preuve sur l'effet traitement est faible compte tenu des biais de l'étude, biais de sélection et biais de suivi notamment. Aussi, la CEESP a émis une réserve majeure concernant « *les données individuelles cliniques de faible niveau de preuve pour estimer les probabilités de transition* ».

1.3.4 Analyse des avis contenant des comparaisons indirectes

Nous nous sommes attachés dans cette analyse à relever la méthodologie des comparaisons indirectes utilisées dans les avis d'efficacité concernés, et à évaluer si cette

méthodologie avait occasionné ou non des réserves de la part de la CEESP. Nous avons également analysé les indications thérapeutiques des dossiers déposés.

1.3.4.1 Analyse quantitative : méthodologie des comparaisons indirectes, indications thérapeutiques, réserves.

1.3.4.1.1 Méthodologie des comparaisons indirectes

Parmi les 5 avis d'efficacité mentionnant des comparaisons indirectes, nous avons relevé :

- 2 avis utilisant la méthode de Bucher
- 2 avis utilisant la comparaison indirecte en réseau (MTC : « *Mixed treatment comparison* »)
- 1 avis qui utilise le terme comparaison indirecte de manière inappropriée (avis OPSUMIT).

Cette répartition est illustrée sur la Figure 47.

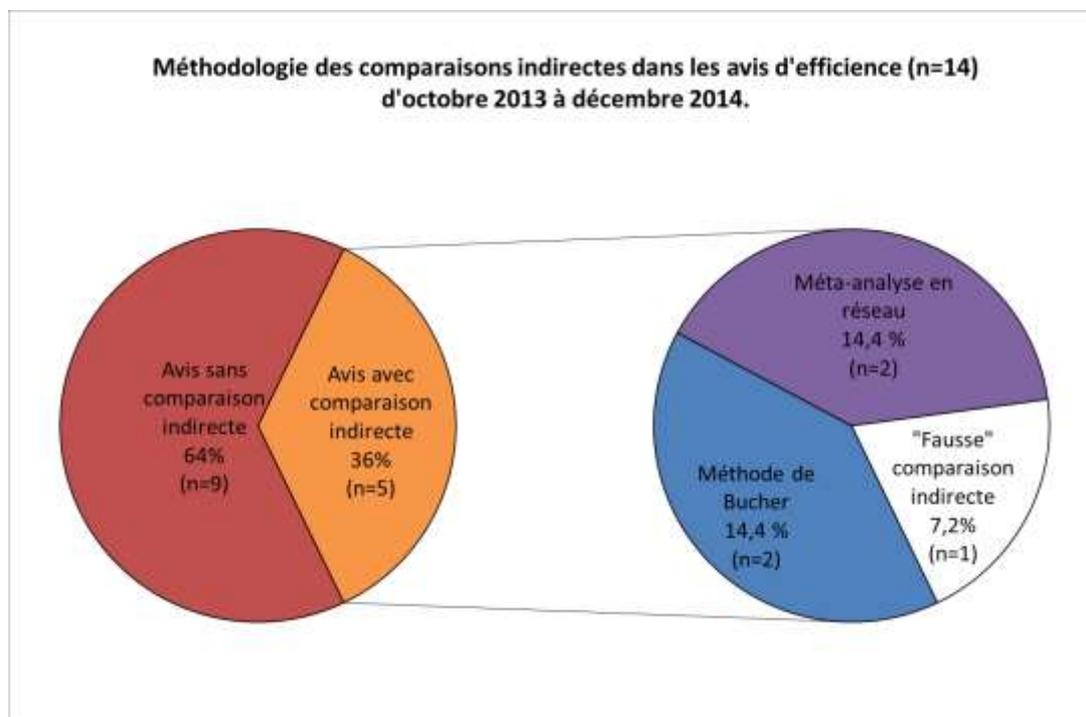


Figure 47 : Méthodologie des comparaisons indirectes dans les avis d'efficacité (n=14) d'octobre 2013 à décembre 2014.

Ainsi, la méthode de Bucher et la méta-analyse en réseau sont employées en proportion équivalente. Ces deux méthodes répondent aux recommandations de la HAS quant aux méthodes de comparaisons indirectes à utiliser dans une évaluation médico-économique.

1.3.4.1.2 Indications thérapeutiques

Les indications thérapeutiques identifiées dans ces 5 avis mentionnant une comparaison indirecte sont :

- des maladies rares (2 avis) : Hypertension artérielle pulmonaire (HTAP) idiopathique et Hypertension pulmonaire thromboembolique chronique (HTP-TEC) ;
- la sclérose en plaques (SEP-RR) (2 avis) ;
- l'oncologie : cancer du sein (1 avis).

Ces indications sont regroupées sur la Figure 48.

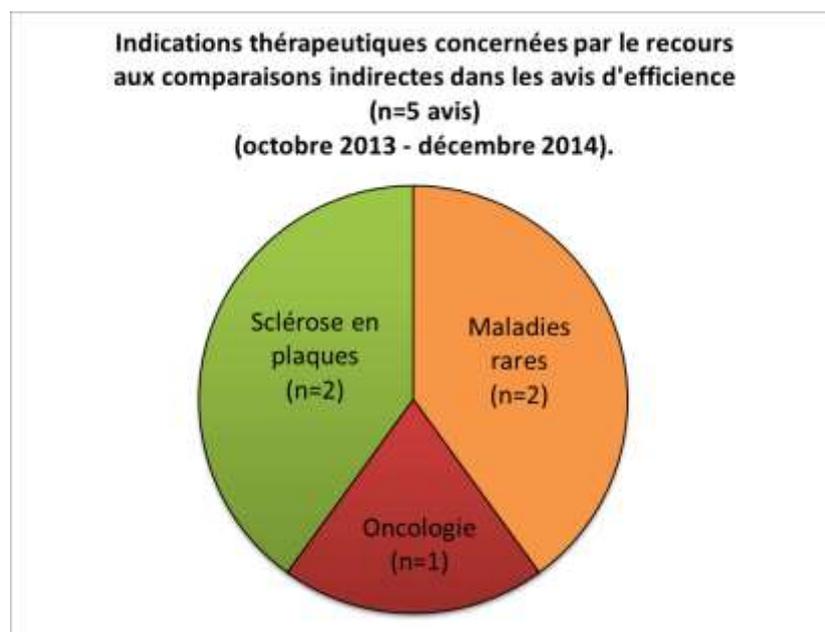


Figure 48 : Indications thérapeutiques concernées par le recours aux comparaisons indirectes dans les avis d'efficacité (n=5) (octobre 2013 – décembre 2014).

Les indications thérapeutiques qui nécessitent une comparaison indirecte pour réaliser les études d'efficacité sont les maladies où il y a beaucoup de stratégies comparées,

avec parfois des développements concomitants (oncologie, sclérose en plaques), et les maladies rares.

1.3.4.1.3 Réerves émises à l'encontre des comparaisons indirectes.

Sur les 5 avis mentionnant des comparaisons indirectes, nous ne considérons ici que les 4 avis contenant réellement une comparaison indirecte (exclusion de l'avis OPSUMIT, que nous détaillerons plus loin). Sur ces 4 avis :

- aucune réserve majeure n'a concerné les comparaisons indirectes ;
- une réserve importante a été émise à propos d'une comparaison indirecte dans l'un des avis (LEMTRADA) ;
- une réserve mineure a été émise à propos d'une comparaison indirecte dans un autre des avis (TECFIDERA).

Cette distribution est représentée sur la Figure 49.

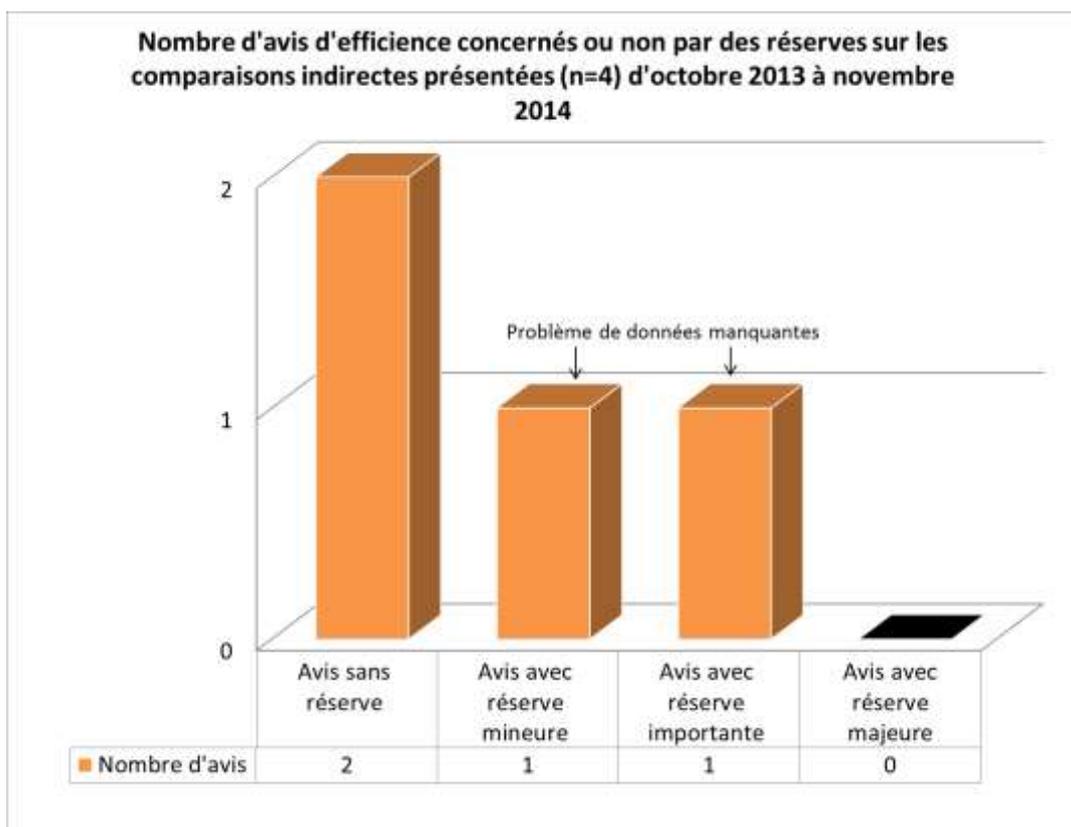


Figure 49 : Nombre d'avis d'efficace concernés ou non par des réserves sur les comparaisons indirectes présentées (n=4) d'octobre 2013 à novembre 2014.

Les deux réserves émises étaient en lien avec des données manquantes.

Au total, toutes les comparaisons indirectes présentées ont par conséquent été jugées acceptables par la CEESP (aucune réserve majeure invalidant l'analyse). Les deux réserves émises étaient en lien avec des données manquantes dans le rapport des comparaisons indirectes présenté par l'industriel.

1.3.4.2 Analyse qualitative

L'analyse des 5 avis d'efficacité ainsi que les réserves portant sur les comparaisons indirectes est résumée dans le Tableau 11. Nous allons à présent explorer ces avis.

Date	Dossier	DCI	Indication	Stratégies comparées	Rationnel pour l'utilisation d'une comparaison indirecte	Méthode de CI	Prise en compte des tous les comparateurs? Si non : quels sont les comparateurs manquants?	Qualité de la comparaison indirecte	Réserve portant sur la comparaison indirecte		
									mineure	importante	majeure
24/06/14	OPSUNIT	macitentan	Hypertension artérielle pulmonaire (HTAP) idiopathique	bosentan (TRACLEER)	Absence de comparaisons directes entre macitentan 10mg (OPSUMIT) et bosentan (TRACLEER)	Méthode non valide	OUI	La méthodologie n'est pas suffisamment détaillée pour juger de la qualité des comparaisons indirectes effectuées.	NA	NA	*Analyse coût-résultat fondée sur une hypothèse de supériorité non démontrée. *L'effet traitement de bosentan (TRACLEER) est estimé sur les données du bras macitentan 3mg de l'essai Séraphin.
14/10/14	ADEMPAS	riociguat	Hypertension pulmonaire thromboembolique chronique (HTP-TEC)	*bosentan (TRACLEER) *absence d'alternative thérapeutique *sildénafil : analyse exploratoire	Absence d'essai direct vs bosentan.	BUCHER : Les odds ratios de chaque bras actif sont estimés par comparaisons indirectes (méthode de Bucher) à partir de deux essais : CHEST-1 (ADEMPAS vs placebo) et BENEFIT (bosentan vs placebo).	OUI	Acceptable	NA	NA	NA
11/03/14	KADCYLA	trastuzumab emtansine	Cancer du sein	*lapatinib + capecitabine (comp.directe, étude pivot) = étude de référence *trastuzumab + capecitabine (Méthode de Bucher) = étude complémentaire car trop incertitudes *Capecitabine monothérapie (Méthode de Bucher)	Absence d'essais direct	Comparaison indirecte par la méthode de BUCHER, de trastuzumab emtansine versus trastuzumab + capecitabine, (représentant 20% des prescriptions), et versus capecitabine monothérapie (représentant moins de 5% des prescriptions en France).	OUI	Interprétation avec prudence=> analyse complémentaire.	NA	NA	NA
15/04/14	LEMTRADA	alemtuzumab	Scléroses en plaque (SEP-RR)	1ère ligne : l'interféron-β-1a (REBIF 44mcg) interféron-β-1a 30 mcg(AVONEX) l'interféron-β-1b 250 mcg (BETAFERON/EXTAVIA) l'acétate de glatiramère 20 mg (COPAXONE) 2ème ligne : fingolimod 0,5 mg (GILENYA) natalizumab 300 mg (TYSABRI) A venir : fumarate de dimé-thyle 240 mg (TECFIDERA) térfunomide 14 mg (AUBAGIO).	Prise en compte simultanément de l'ensemble de ces stratégies	Réseau MTC	OUI	La façon dont cette analyse a été conduite par les auteurs n'a pas été suffisamment décrite pour en juger la qualité et la robustesse.	NA	Présentation insuffisante des données cliniques pour juger de leur robustesse : essais intégrés dans le MTC et méthode utilisée, homogénéité des populations pour des analyses en sous-groupe.	NA
24/06/14	TECFIDERA	dimethyl fumarate	Scléroses en plaque (SEP-RR)	Placebo 1ère ligne : interféron-β-1a (REBIF 44 mcg) interféron-β-1a 30 mcg (AVONEX) interféron-β-1b 250 mcg (EXTAVIA) acétate de glatiramère 20 mg (COPAXONE) 2nd ligne : fingolimod 0.5 mg (GILENYA) natalizumab (TYSABRI).	Prise en compte simultanément de l'ensemble de ces stratégies	Réseau MTC	NON Les alternatives à venir, au moment de la réalisation de l'étude, que sont le térfunomide (AUBAGIO) et l'alemtuzumab (LEMTRADA), ne sont pas prises en compte	Méthode utilisée adaptée et de bonne qualité.	Les risques relatifs de progression du handicap et de survenue d'une poussée ne peuvent pas être retrouvés dans le rapport complet des comparaisons indirectes.	NA	NA

Tableau 11 : Résumé de l'analyse des avis d'efficience avec des comparaisons indirectes.

En bleu : Méthode de Bucher. En violet : Méta-analyse en réseau (MTC). En vert : positif. En rouge : négatif.

1.3.4.2.1 Avis OPSUNIT du 24/06/14 : utilisation d'une méthode invalide

L'étude expertisée évalue le rapport différentiel coût-résultat (RDCR) du macitentan 10mg (OPSUNIT) versus le bosentan (TRACLEER) tous deux commercialisés par le même laboratoire, chez des patients atteints d'HTAP. L'essai Séraphin (phase III), qui a comparé OPSUNIT à un placebo comportait trois bras : OPSUNIT 10mg, OPSUNIT 3mg, et placebo. Devant l'absence de données de comparaison directe entre OPSUNIT et TRACLEER (ce qui peut paraître surprenant puisqu'il s'agit du même laboratoire), l'industriel a fait le postulat que TRACLEER avait une moins bonne efficacité que OPSUNIT 10mg, et qu'il pouvait donc être considéré comme équivalent au bras OPSUNIT 3mg. L'étude d'efficience a donc comparé le bras OPSUNIT 10 mg au bras OPSUNIT 3mg (assimilé à TRACLEER) (Figure 50).

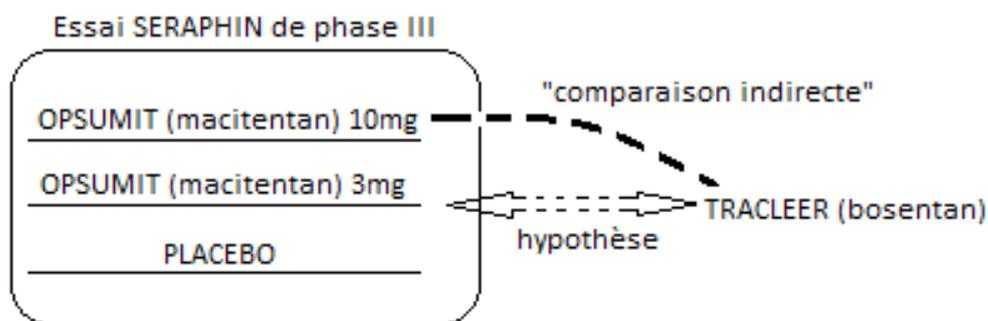


Figure 50 : Principe de la « fausse » comparaison indirecte décrite dans l'avis d'efficience OPSUNIT du 24/06/2014.

L'efficacité du TRACLEER a été supposée équivalente au bras OPSUNIT 3 mg de l'essai de phase III SERAPHIN.

La méthodologie de cette « fausse » comparaison indirecte n'a pas été considérée acceptable par la CEESP pour deux raisons :

- « le choix d'une analyse de type coût-utilité repose sur une hypothèse de supériorité de macitentan 10mg (OPSUNIT) versus bosentan (TRACLEER). En l'absence de comparaison directe, la justification de cette hypothèse repose sur des comparaisons indirectes dont la qualité méthodologique est incertaine ;
- l'utilisation des données du bras macitentan 3mg de l'essai Séraphin pour alimenter le bras comparateur du modèle repose sur des justifications dont le niveau de preuve n'est pas suffisant. »

Ces remarques ont fait l'objet d'une réserve majeure. La CEESP a précisé que l'hypothèse de supériorité de macitentan 10mg (OPSUNIT) versus bosentan (TRACLEER)

étant incertaine, une analyse de référence en minimisation de coût aurait été préférable. Le RDCR calculé n'a pas été repris dans les conclusions, seules les données de coûts ont été reprises.

1.3.4.2.2 Avis ADEMPAS et KADCYLA : utilisation de la méthode de Bucher

❖ Avis ADEMPAS¹⁰² du 14/10/14

L'étude expertisée évalue l'efficacité du riociguat (ADEMPAS) dans le traitement de l'hypertension pulmonaire thromboembolique chronique (HTP-TEC). A ce jour, il n'existe pas de traitement médicamenteux ayant obtenu une AMM dans l'indication concernée. Cependant, le laboratoire indique que selon les données d'un registre français, les traitements les plus souvent administrés en 1^{ère} intention sont le bosentan (TRACLEER) et le sildénafil en monothérapie, et leur association en bithérapie. Les deux comparateurs retenus dans l'analyse de référence sont le bosentan et l'absence d'alternative thérapeutique. En l'absence de comparaison directe, la comparaison au bosentan est effectuée grâce à une comparaison indirecte qui fait appel à la méthode de Bucher : les odds ratios de chaque bras actif sont estimés à partir de deux essais: CHEST-1 (riociguat vs placebo) et BENEFIT (bosentan vs placebo) (Figure 51).

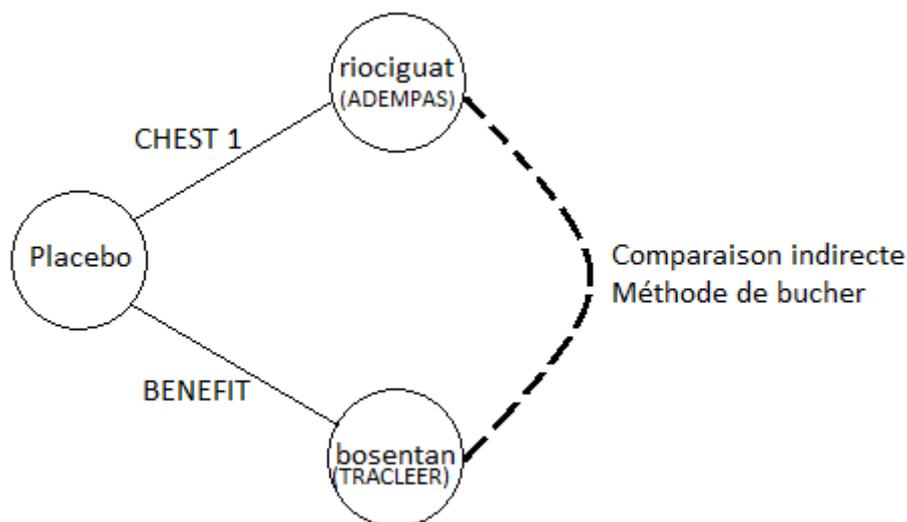


Figure 51 : Comparaison indirecte effectuée avec la méthode de Bucher dans l'avis d'efficacité ADEMPAS du 14/10/14.

Le sildénafil en monothérapie ou en association au bosentan a été écarté comme traitement comparateur dans l'analyse de référence en raison d'une part, des faibles niveaux de prescription par rapport au bosentan en monothérapie, et d'autre part, de l'absence de données cliniques évaluant le sildénafil en monothérapie ou en association dans l'HTP-TEC. Une analyse exploratoire est cependant présentée par l'industriel, intégrant un bras sildénafil et un bras sildénafil associé au bosentan. La CEESP a rappelé à cette occasion, comme dans l'avis VECTIBIX, que la faible part de marché n'est pas à elle seule un argument recevable pour exclure une alternative des comparateurs dans une analyse médico-économique. En revanche, l'absence de données sur l'effet traitement du sildénafil utilisé hors AMM ne peut être contestée, et cela justifie donc son intégration dans une analyse exploratoire, fondée sur des hypothèses.

La comparaison indirecte réalisée n'a pas occasionné de réserves de la part de la Commission⁹.

❖ Avis KADCYLA⁹⁸ du 11/03/14: comparaison indirecte en analyse complémentaire

L'étude KADCYLA soumise évalue l'efficacité du trastuzumab emtansine (KADCYLA) en monothérapie dans le traitement de patients adultes atteints d'un cancer du sein.

La stratégie alternative envisagée dans l'analyse de référence est le comparateur de l'étude pivot (lapatinib + capécitabine), décrit comme correspondant au traitement de référence du traitement de 2^e ligne du cancer du sein métastatique et représentant 30% des traitements pour cette indication en France (Figure 52). Une revue systématique de la littérature a par ailleurs été conduite afin d'identifier la disponibilité de données cliniques d'intérêt pour la comparaison de trastuzumab emtansine par rapport aux alternatives thérapeutiques utilisées en pratique courante en France. Cette revue systématique a permis de sélectionner trois études cliniques randomisées permettant de réaliser une comparaison indirecte par la méthode de Bucher, de trastuzumab emtansine versus trastuzumab + capécitabine (20% des prescriptions en France) et versus capécitabine monothérapie (moins de 5% des prescriptions). Les autres alternatives thérapeutiques utilisées en pratique

⁹ On peut cependant souligner qu'une réserve majeure globale a été émise sur la présentation des résultats ce qui a invalidé l'étude et l'ensemble des résultats et conclusions soumis à la Commission.

courante dans la population de l'indication ne peuvent pas être évaluées comparativement à trastuzumab emtansine compte tenu de l'absence de données directes ou indirectes.

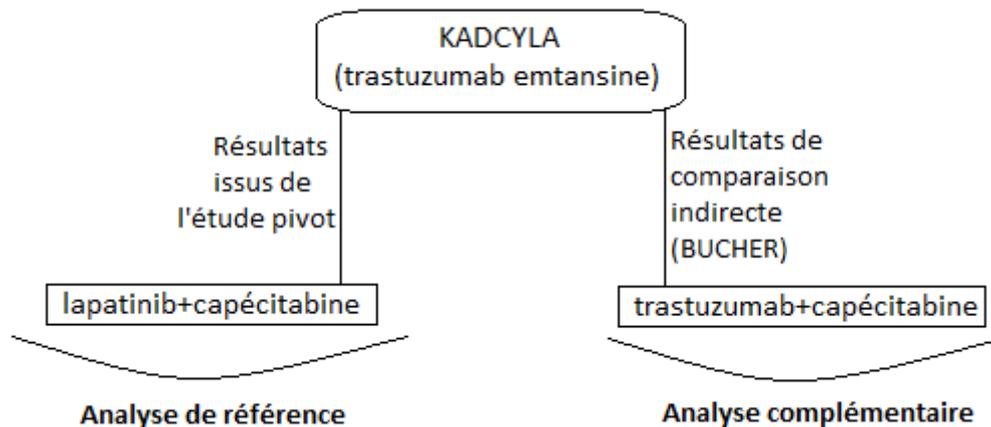


Figure 52 : Analyse du dossier d'efficacité KADCYLA (avis d'efficacité du 11/03/2014).

Dans le dossier soumis, l'analyse de l'efficacité de trastuzumab emtansine versus trastuzumab + capécitabine est présentée comme une analyse complémentaire (Figure 52) à l'analyse de référence compte tenu de « l'incertitude entourant les résultats de la comparaison indirecte et des hypothèses considérées pour réaliser cette analyse ». Selon le rapport, l'effet du traitement est estimé sur la base d'une « comparaison indirecte de deuxième ordre » (ce terme n'est pas défini par le laboratoire) à partir de deux études cliniques. La CEESP n'a pas tenu compte de cette analyse complémentaire dans la mesure où l'analyse de référence avait un meilleur comparateur, et que le terme « comparaison indirecte de 2^{ème} ordre » n'était pas défini, laissant penser qu'il ne permettait pas de tirer des conclusions à haut niveau de preuve.

1.3.4.2.3 Avis LEMTRADA et TECFIDERA (sclérose en plaques) : utilisation de la méta-analyse en réseau

L'objectif de ces deux avis d'efficacité est de présenter le rapport coût-efficacité des spécialités LEMTRADA (alemtuzumab) et TECFIDERA (diméthyl fumarate) par rapport aux alternatives médicamenteuses pour le traitement de la forme active de la sclérose en plaques rémittente (SEP-RR). Ces alternatives étant nombreuses, les deux laboratoires ont

fait appel chacun à une méta-analyse en réseau pour pouvoir réaliser l'étude. Certaines critiques ou au contraire certains points positifs ont été émis par la CEESP concernant ces méta-analyses en réseau. Nous les avons relevés à l'aide de la grille de lecture (Annexe 2) ; ils sont regroupés dans le Tableau 12 et seront détaillés dans les deux paragraphes qui suivent.

Date	Dossier	DCI	Analyse de la recherche et essais inclus	Homogénéité et similarité		Méthode statistique		
			Qualité des études incluses	Hétérogénéité	Similarité (modificateurs)	Méthode	Cohérence	Présentations des résultats
15/04/2014	LEMTRADA	alemtuzumab	Les designs des études n'ont pas été présentés.		Les éventuels facteurs modulateurs de l'effet des traitements et leurs répartitions dans les essais n'ont pas été présentés (par exemple, l'existence d'un traitement précédent).		D'autres méta-analyses ont été réalisées, notamment celle de la Cochrane (2013). Les auteurs auraient pu discuter les résultats des différentes méta-analyses réalisées et expliquer les différences observées.	
24/06/2014	TECFIDERA	dimethyl fumarate	La qualité des études est analysée au travers de deux indicateurs (qualité recrutement/randomisation, score de Jadad) et d'une évaluation qualitative.	Pour analyser l'impact de facteurs confondants (type d'aveugle, taille de la population dans chaque bras) ou de covariables (durée de l'étude, âge, sexe, ancienneté de la maladie, taux de poussée et niveau de handicap à l'entrée) sur les critères de résultat, une analyse de sensibilité et une évaluation de l'hétérogénéité sont réalisées.				Les résultats présentés dans le rapport technique concernant les risques relatifs de progression du handicap et de survenue d'une poussée de chaque traitement par rapport au placebo ne peuvent être retrouvés dans le rapport complet des comparaisons indirectes qui présentent les résultats du diméthyle fumarate par rapport aux autres traitements

Tableau 12 : Résumé des critiques (en rouge) et points positifs (en vert) émis à l'encontre des méta-analyses en réseau présentées dans les avis d'efficacité LEMTRADA et TECFIDERA.

❖ Avis LEMTRADA⁹⁹ du 15/04/14 :

Dans le « projet d'avis » LEMTRADA (nous verrons plus loin que la demande a ensuite été retirée par le laboratoire), le modèle comporte le bras de référence (alemtuzumab) et un bras comparateur. Les stratégies comparées sont des traitements médicamenteux suivants : l'interferon- β -1a (REBIF), l'interferon- β -1a (AVONEX), l'interferon- β -1b (BETAFERON/EXTAVIA), l'acétate de glatiramère (COPAXONE), le fingolimod (GILENYA), le natalizumab (TYSABRI), le fumarate de diméthyle (TECFIDERA), et le tériflunomide (AUBAGIO). L'estimation de l'effet des différents traitements repose principalement sur des données de comparaison indirecte en réseau (« *mixed treatment comparison* »). Cependant, la façon dont cette analyse a été conduite n'a pas été suffisamment décrite par les industriels dans le rapport pour pouvoir juger la qualité et la robustesse des données, ce qui a occasionné une réserve importante (« *Présentation insuffisante des données cliniques pour juger de leur robustesse : essais intégrés dans le MTC et méthode utilisée, homogénéité des populations pour des analyses en sous-groupe.* »). En effet, les études intégrées dans la revue de la littérature ont été listées mais les méthodologies des études n'ont pas été présentées. De plus, les éventuels facteurs modulateurs de l'effet des traitements et leurs répartitions dans les essais n'ont pas été décrits (par exemple, l'existence d'un traitement précédent). Enfin, d'autres méta-analyses ont été réalisées, mais les auteurs n'ont pas discuté les résultats et expliqué les différences observées (hypothèse de cohérence). La CEESP a d'autre part souligné la prise en compte des alternatives à venir non encore disponibles (fumarate de diméthyle (TECFIDERA) et teriflunomide (AUBAGIO) qui répond de façon satisfaisante aux attentes de la HAS.

On peut noter que le laboratoire a retiré sa demande d'inscription sur la liste des spécialités pharmaceutiques agréées à l'usage des collectivités et divers services publics suite à un projet d'avis adopté par la Commission de la transparence le 23 avril 2014 qui prévoyait un service médical rendu par LEMTRADA insuffisant dans l'indication. Le projet d'avis d'efficience du CEESP a donc été également retiré. Il est publié sur le site de la HAS avec le procès-verbal de la CEESP du 15 avril 2014.

L'analyse de l'avis d'efficience LEMTRADA démontre que la réalisation d'une méta-analyse en réseau dans le but de saisir tous les comparateurs pertinents répond certes aux recommandations de la HAS, mais à condition que l'exhaustivité des données soient

présentée et justifiée. Cette condition a été plus fidèlement respectée dans le rapport TECFIDERA.

❖ Avis TECFIDERA du 24/06/14

Tout comme dans l'avis d'efficience LEMTRADA, le laboratoire a réalisé des comparaisons indirectes entre un placebo et les différents traitements afin de prendre en compte simultanément l'ensemble des stratégies. La méta-analyse en réseau présentée a également été soumise dans le dossier pour la CT, comme nous l'avons vu précédemment dans la partie 2, chapitre 2.3.4.2.2. Les stratégies comparées sont les mêmes que dans l'avis LEMTRADA, à la différence que seuls les traitements actuellement sur le marché sont considérés dans l'analyse, ce qui a été désapprouvé par la CEESP au travers d'une réserve importante (« *Choix d'exclure tériflunomide (AUBAGIO) et l'alemtuzumab (LEMTRADA) comme comparateur non conforme aux recommandations de la HAS* »). Les auteurs justifient ce choix en indiquant que le CEPS recommande de ne pas inclure les traitements non remboursés dans l'analyse. La CEESP a alors rappelé que le fait de ne pas avoir obtenu l'accès au remboursement ne signifie pas que les produits ne doivent pas être pris en compte, et qu'en l'absence de prix connus, des hypothèses peuvent être envisagées pour prendre en compte l'ensemble des traitements (par exemple : prix de l'ATU, prix pratiqué dans un autre pays européens, prix d'un comparateur etc.).

L'effet des différents traitements est documenté à partir d'une comparaison indirecte en réseau réalisée à partir d'une revue systématique de la littérature. Le rapport technique présente les risques relatifs de la maladie sous les différents traitements par rapport au placebo, celui-ci étant considéré comme équivalent à l'évolution naturelle de la maladie. Cette réduction du risque est appliquée aux probabilités de transition entre les états sans traitement (évolution naturelle de la maladie) afin d'estimer les probabilités de transitions associées à chacun des traitements comparés. Les analyses de sensibilité probabilistes retiennent les intervalles de confiance de la méta-analyse en réseau pour tester la sensibilité du modèle aux paramètres d'efficacité.

La méthode utilisée pour réaliser ces comparaisons indirectes est bien décrite par les auteurs :

- le design des études ainsi que leurs principales caractéristiques (traitements évalués, durée de suivi, population, pays) sont détaillés,
- la qualité des études est analysée au travers de deux indicateurs (notamment score de Jadad) et d'une évaluation qualitative.
- une analyse de sensibilité et une évaluation de l'hétérogénéité sont réalisées pour analyser l'impact de facteurs confondants (type d'aveugle, taille de la population dans chaque bras) ou de covariables (durée de l'étude, âge, sexe, ancienneté de la maladie, taux de poussée et niveau de handicap à l'entrée) sur les critères de résultats.

Une réserve mineure a néanmoins été émise car les résultats concernant les risques relatifs de progression du handicap et de survenue d'une poussée de chaque traitement par rapport au placebo n'ont pu être retrouvés dans le rapport complet des comparaisons indirectes.

Finalement, selon la CEESP, l'ensemble de ces éléments indique que la méthode utilisée est adaptée et de bonne qualité, ce qui permet d'avoir confiance en l'analyse d'efficacité fondée sur les données d'efficacité obtenues avec cette méta-analyse en réseau. On peut rappeler à cette occasion que la CT de son côté a également reconnu la qualité de cette méta-analyse en réseau, mais n'a pas valorisé les conclusions qui en étaient ressorties.

1.4 Résumé et discussion

Notre étude a pris en compte les avis d'efficacité rendus par la CEESP depuis la mise en application du décret du 2 octobre 2012 et clos à la date du 15 décembre 2014. De cette manière, 14 avis d'efficacité ont été relevés sur la période considérée (octobre 2013 à décembre 2014, soit environ 1 an et 2 mois). Sur ces 14 avis, seulement 5 avis évoquaient l'utilisation d'une comparaison indirecte, soit 36% des avis d'efficacité de l'étude.

A partir de ce constat, nous avons souhaité, dans un premier temps, saisir les éléments qui peuvent dispenser un laboratoire de recourir à une comparaison indirecte pour prendre en compte tous les comparateurs existants. Ainsi, deux types de justifications ont

été mentionnées dans les 9 avis sans comparaisons indirectes : il s'agissait soit d'une hétérogénéité entre les études trop importante pour pouvoir réaliser des comparaisons indirectes (5 avis), soit d'un faible nombre de comparateurs pertinents doublé de données de comparaisons directes disponibles (4 avis). L'exemple des avis ROTAREQ, ROTARIX, TIVICAY, VECTIBIX et XOFIGO (soit 5 avis sur les 9 avis sans comparaisons indirecte) nous a néanmoins démontré que la comparaison au placebo seul ou à une sélection non exhaustive de stratégies comparées, même accompagnée de justifications de la part du laboratoire, satisfait difficilement la CEESP quand il existe d'autres comparateurs pertinents car ce choix limite les conclusions d'efficacité. Celle-ci peut aller jusqu'à émettre une réserve de niveau « important » en rapport avec cette critique. En outre, l'argument sur les parts de marché n'est en soi pas suffisant pour la CEESP qui considère que ce n'est pas parce qu'une stratégie est moins utilisée en pratique qu'elle n'est pas sur la frontière d'efficacité (avis VECTIBIX). Certains laboratoires font alors le choix, afin de répondre aux recommandations de la HAS, de passer outre une trop grande hétérogénéité et d'utiliser une comparaison naïve des bras actifs de différents essais (cas des avis SOVALDI et OLYSIO). Dans tous les cas, notre étude a mis en exergue l'importance d'utiliser des données issues d'essais cliniques de bonne qualité, et de justifier tous les choix effectués dans l'analyse d'efficacité. La comparaison au placebo seul ou aux « soins courants » n'est acceptée que si strictement aucun autre comparateur n'existe (cas de l'avis ZOSTAVAX et DEFITELIO soit 14% des avis de l'étude).

Dans un deuxième temps, nous nous sommes préoccupés des 5 avis d'efficacité qui évoquaient l'utilisation d'une méthode de comparaison indirecte. Il s'est avéré que les indications thérapeutiques des médicaments impliqués concernaient des maladies où il y a beaucoup de stratégies comparées (oncologie, sclérose en plaques) et les maladies rares. Les méthodes employées ont été la méthode de Bucher et la méta-analyse en réseau, en proportion équivalente; ces techniques répondent aux recommandations de la HAS. Dans un des avis le terme comparaison indirecte a été utilisé de manière impropre (avis OPSUMIT). Nous avons ensuite désiré connaître les critiques méthodologiques et éventuellement les réserves qui ont été émises à l'égard de ces comparaisons indirectes présentées par les industriels. Dans l'ensemble, toutes les comparaisons indirectes présentées ont été jugées acceptables par la CEESP puisqu'aucune réserve majeure invalidant l'analyse n'a été émise. L'analyse de l'avis LEMTRADA met toutefois en évidence la nécessité de décrire

rigoureusement la façon dont la méta-analyse en réseau a été conduite afin de pouvoir en juger la qualité et la robustesse, sous peine de se voir attribuer une réserve importante. Cette condition a été plus fidèlement respectée dans l'avis TECFIDERA où la méta-analyse en réseau bien décrite a permis de répondre aux recommandations de la HAS concernant les stratégies comparées.

3 Conclusion

Depuis le décret du 2 octobre 2012, mis en application à partir d'octobre 2013, l'évaluation médico-économique fait partie du processus d'accès au remboursement des médicaments (via la fixation du prix) pour les médicaments revendiquant une ASMR de niveau I à III et susceptibles d'avoir un impact significatif sur les dépenses de l'assurance maladie. La HAS, responsable de cette évaluation des analyses médico-économiques déposées par les laboratoires, recommande la prise en compte de tous les comparateurs pertinents. La proportion modérée de dossiers ayant eu recours à une méthodologie de comparaison indirecte pour répondre à cette problématique (5 avis sur les 14 avis d'efficience rendus depuis octobre 2013, soit 36%), peut s'expliquer par deux éléments. D'une part, les médicaments qui revendiquent un ASMR de I à III peuvent être des médicaments innovants pour lesquels il n'existe pas encore de comparateurs et pour lesquels la comparaison au placebo seul convient (2 avis sur les 14 avis de l'étude, soit 14%). D'autre part, bien que plusieurs comparateurs existent, une trop forte hétérogénéité entre les essais de comparaison directe freine le laboratoire à l'utilisation d'une comparaison indirecte (5 avis sur les 14 avis de l'étude, soit 36%). Malgré cela, la CEESP accepte difficilement que des comparateurs ne soient pas inclus dans l'étude, puisque 5 des 9 dossiers (soit 56%) n'ayant pas utilisé de comparaison indirecte en l'absence de données de comparaison directe se sont vu infliger une réserve mineure ou importante pour ce motif. Il semble alors préférable de considérer tous les comparateurs dans l'analyse, même si cela doit passer par l'utilisation d'une comparaison naïve des bras actifs lorsque l'hétérogénéité est trop forte entre les études (avis SOVALDI et OLYSIO). Dans la mesure où les analyses de sensibilité appliquées au RDCR permettent de prendre en compte l'incertitude sur les résultats d'efficacité, cette pratique semble tolérée par la CEESP (pas de réserve majeure en

rapport avec l'utilisation d'une comparaisons naïve dans les avis SOVALDI et OLYSIO), bien qu'elle ne réponde pas aux recommandations éditées en 2011.

Par ailleurs, nous avons pu constater que toutes les comparaisons indirectes présentées dans l'étude ont été jugées acceptables par la CEESP puisqu'aucune réserve majeure invalidant l'analyse n'a été émise. L'examen de l'avis d'efficience TECFIDERA montre que la CEESP valide volontiers une analyse reposant sur des données issues d'une méta-analyse en réseau bien conduite et suffisamment décrite pour pouvoir juger de la qualité et de la robustesse.

Conclusion

Le choix du médicament le plus efficace, le mieux toléré en vue de l'accès au remboursement, et/ou le plus efficient, afin de fixer le prix le plus juste, nécessite une évaluation comparative entre les technologies de santé. Les essais contrôlés randomisés sont une source importante d'éléments de preuve pour ces comparaisons. Cependant, en l'absence de ces comparaisons directes, rares eu égard au champ immense des comparaisons à réaliser, ou afin d'obtenir des comparaisons multiples, les comparaisons indirectes peuvent être alors la seule source d'estimation de l'efficacité comparative. Si les deux types de comparaisons, indirectes et directes, sont disponibles, les comparaisons indirectes peuvent apporter des preuves supplémentaires.

Deux études rétrospectives ont montré que la Commission de la transparence (CT) utilise, sous réserve de leur qualité méthodologique, les résultats des comparaisons indirectes comme preuves complémentaires et contributives, et non pas comme éléments déterminants sur lesquels se fonde l'Amélioration du Service Médical Rendu (ASMR). La Commission d'évaluation économique et de santé publique (CEESP) valide fréquemment les résultats des comparaisons indirectes intégrés dans les modèles médico-économiques, avec les mêmes réserves méthodologiques.

Cette différence apparente dans la manière de considérer et d'utiliser les comparaisons indirectes par ces deux commissions de la Haute Autorité de Santé peut s'expliquer par les objectifs différents qui sont les leurs.

Au travers de l'ASMR, la CT doit positionner les médicaments les uns par rapport aux autres afin de pouvoir apprécier l'apport du médicament, le prix se fondant entre autre sur cette appréciation. On perçoit alors pourquoi l'ASMR ne peut se fonder sur le seul résultat d'une comparaison indirecte, même bien conduite, mais davantage sur un faisceau d'arguments. L'ASMR est établi vis-à-vis de comparateurs choisis comme pertinents, et non en perspective « tous payeurs », ce qui rejoint la logique clinique généralement utilisée lors de la conception des essais cliniques du développement des produits de santé.

La CEESP n'utilise pas les résultats de méta-analyses en réseau pour conclure à une supériorité d'un traitement par rapport à un autre, mais plutôt pour construire le modèle médico-économique qui permettra de calculer le ratio différentiel coût résultats (RCDR). Les

résultats de ces méta-analyses sont utilisés pour renseigner les paramètres cliniques de chaque bras du modèle sur la base de données prenant en compte un effet traitement relatif de sorte à biaiser le moins possible les résultats. Le modèle comprend les incertitudes autour de cette estimation ponctuelle considérant l'intervalle de confiance ou d'incertitude des paramètres dans les analyses de sensibilité. Le résultat issu de ces analyses de sensibilité, notamment probabilistes, rend compte de l'incertitude des estimations de la méta-analyse en réseau. L'approche « tous payeurs » utilisée dans l'évaluation médico-économique permet d'identifier si le nouveau produit de santé modifiera la stratégie actuelle de prise en charge. La condition de complétude des comparateurs expose à l'indisponibilité de l'intégralité des comparaisons directes et donc au recours plus fréquent aux données de comparaisons indirectes.

On peut admettre que les comparaisons indirectes soient exploitées différemment par la CT et par la CEESP, puisque leur finalité diffère. Il peut toutefois être regrettable que ces deux commissions, qui officient au sein de la même autorité, analysent des mêmes données de comparaisons indirectes de manière indépendante sans concertation, ce qui peut exposer à des risques de discordances d'appréciation. Sur ce point, on peut tout à fait imaginer, à l'occasion de la mission en cours qui vise à définir de nouveaux critères d'évaluation des médicaments¹⁰³, une évolution vers la prise en compte de l'évaluation médico-économique y compris dans les critères d'admission au remboursement, et donc une meilleure coordination, voir une fusion, des deux commissions, CT et CEESP⁹⁰, pour la part dévolue à l'évaluation des médicaments.

S'agissant de la problématique de la complétude des comparateurs, le choix du comparateur dans les essais cliniques étant actuellement libre, on pourrait tendre vers une attitude plus prescriptive, notamment pour la construction des modèles médico-économiques⁹⁰ dans le cadre des rencontres précoces qui ont lieu entre les développeurs et les agents de l'évaluation des technologies de santé. Cependant, cela implique d'anticiper sur les molécules en développement et une certaine harmonisation entre les pays.

Des recommandations nationales et internationales sur la réalisation de méta-analyses en réseau pour les dossiers de demande de remboursement sont disponibles¹⁰⁴ (Australie, Belgique, Canada, Allemagne, Ecosse, Espagne, Afrique du Sud, Royaume-Uni,

EUnetHTA pour l'Europe). Le réseau européen pour l'évaluation des technologies de santé, l'EUnetHTA, a en effet élaboré en 2013 des recommandations¹⁰⁵ auxquelles les pays peuvent se référer. La plupart des pays appellent à contrôler la cohérence entre les preuves directes et indirectes, suggérant fortement que les preuves indirectes devraient être considérées comme un complément de la preuve directe. Aucune des recommandations nationales n'ont explicitement indiqué que les différences entre les preuves directes et indirectes rendaient les preuves indirectes irrecevables. L'ISPOR (« *International Society For Pharmacoeconomics and Outcomes Research* ») a publié de nombreux documents sur les bonnes pratiques pour conduire des comparaisons indirectes^{62,106,107}. Cela correspond à un besoin croissant d'informer les décideurs de santé et praticiens sur ces nouvelles méthodes statistiques parfois complexes, mais néanmoins très utiles et qui mériteraient d'être davantage exploitées.

Annexes

Annexe 1: Rankogram en valeurs absolues (Figure A) et en valeurs cumulatives (Figure B) pour présenter les probabilités et les classements d'une méta-analyse en réseau.

Figure A: Rankogram en valeurs absolues.

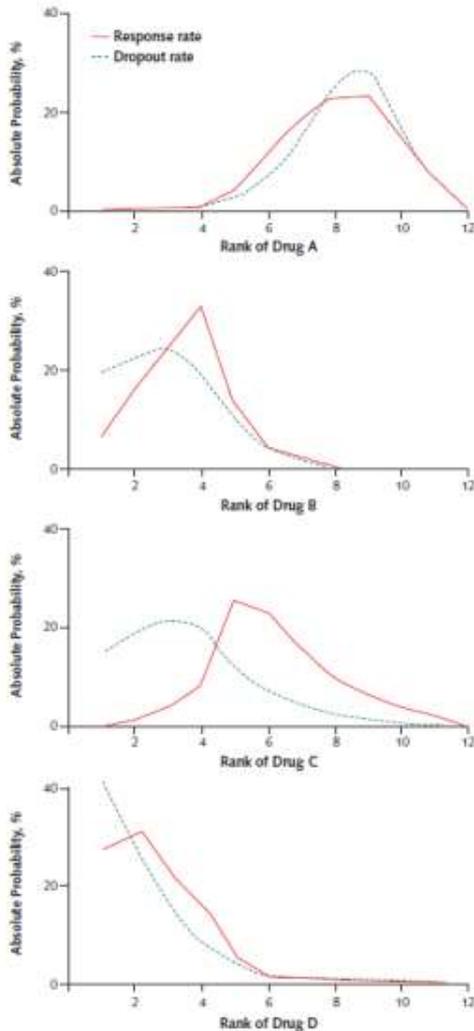
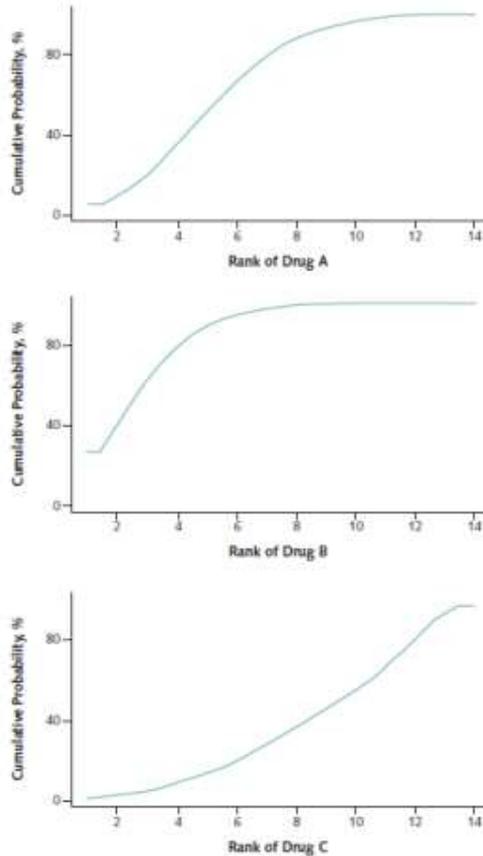


Figure B: Rankogram en valeurs cumulatives.



Les méta-analyses en réseau permettent une estimation de la probabilité que chaque intervention soit la meilleure pour chaque résultat. Les probabilités pour chaque traitement, en prenant en compte chaque rang possible, peuvent être tracées dans des rankograms en valeurs absolues ou des rankograms en valeurs cumulatives.

Figure A : Les rankograms en valeurs absolues (modifiés à partir de la référence²⁸) classent les traitements à partir du taux de réponse (l'efficacité est la ligne rouge continue) et du taux d'abandon (la tolérance est la ligne bleue pointillée). Le classement indique la probabilité d'être le meilleur traitement, le deuxième meilleur, le troisième meilleur, et ainsi de suite, dans le groupe des 12 antidépresseurs étudiés. Par exemple, le médicament A a une probabilité plus élevée d'être parmi les pires antidépresseurs en termes d'efficacité et d'acceptabilité, alors que le médicament B est bon.

Figure B : Les rankograms en valeurs cumulatives (modifiés à partir de la référence²⁸) montrent la distribution des probabilités pour chaque traitement d'être classé à chacune des 14 positions possibles au sein du groupe de médicaments anti-maniaques qui font l'objet de l'étude. Plus la surface sous la courbe de classement cumulatif (généralement appelée « SUCRA ») est importante, plus il est probable que le médicament soit dans les rangs les plus bas (c'est à dire le traitement le plus efficace ou le mieux toléré). Par exemple, les médicaments A et B ont une probabilité plus élevée d'être parmi les meilleurs médicaments anti-maniaques (le médicament B est meilleur que le médicament A), alors que le médicament C est probablement parmi les pires. La SUCRA peut être mesurée et enregistrée dans des tableaux pour montrer ses valeurs moyennes ainsi que les IC 95%.

Annexe 2 : Guide de lecture critique utilisé pour le relevé des données des études rétrospectives.

Pertinence globale pour le décideur				Analyse de la recherche et essais inclus		Homogénéité et similarité		Méthode statistique			
Choix de la population	Choix des traitements	Effet traitements	Contexte	Exhaustivité (biais de publication)	Qualité des études incluses	Hétérogénéité	Similarité (modificateurs d'effet de traitement?)	Comparaison naïve?	Méthode (analyse de sensibilité?)	Cohérence	Présentation des résultats

Choix de la population : Les populations des ECR qui constituent la base de la méta-analyse en réseau correspondent-elles à la population d'intérêt pour le décideur ? (Maladie d'intérêt, stade, gravité, antécédents de traitements, âge, sexe etc.) Choix des traitements : Les médicaments inclus dans la comparaison indirecte (méta-analyse en réseau) correspondent-ils à ceux d'intérêt pour le décideur ? Tous les comparateurs pertinents ont été considérés ? Effets traitements : Les critères d'évaluation sont-ils pertinents ? (Clinical endpoint, intermediate endpoint : ex : mortalité cardiovasculaire, taux de cholestérol etc). Contexte : La définition de la pathologie, les populations de patients, les modalités générales de prise en charge sont-elles susceptibles d'avoir évolué entre les moments où ont été réalisées les différentes études incluses dans la méta-analyse en réseau ? Exhaustivité (biais de publication) : Un processus de recherche exhaustive et reproductible des essais de type méta-analyse a-t-il été suivi ? S'agit-il d'essais sélectionnés pour donner les résultats de comparaisons indirectes souhaités ou s'agit-il de tous les essais réalisés dans le domaine ? Qualité des études incluses : Le rapport doit fournir les renseignements concernant les caractéristiques clés de chaque ECR, comme la méthode de randomisation, les perdus de vue etc. Hétérogénéité : Les études rassemblées dans les méta-analyses doivent être homogènes (quantification de l'hétérogénéité grâce à l'indice de Higgins I²). Utilisation d'un modèle à effets aléatoires pour prendre en compte une certaine variabilité aléatoire de l'effet traitement d'un essai à un autre, ou bien d'un modèle à effet fixe si le véritable effet du traitement est commun à tous les essais. Similarité : Dans les conditions de l'essai de B, le traitement A aurait donné les mêmes résultats que ceux obtenus dans l'essai A, et vis versa. Cela suppose un équilibre dans les caractéristiques des patients ou de l'étude en terme de modificateur de l'effet relatif de traitement. Cette hypothèse se vérifie à l'aide des analyses en sous-groupe et du test d'interaction. Comparaison naïve : La méthode utilisée pour réaliser la comparaison indirecte est une comparaison naïve donc inappropriée. Méthode : La méthode statistique employée est suffisamment décrite pour en juger de l'acceptabilité (aperçu global des méthodes statistiques, justification du choix des critères d'évaluation, des estimations des effets relatifs, du choix d'un modèle à effets fixes ou aléatoires, du choix de la distribution *a priori* pour les paramètres du modèle si les analyses ont été effectuées dans un cadre bayésien, description des différentes analyses de sensibilités). Cohérence : L'hypothèse de cohérence est vérifiée s'il n'y a pas de différence entre les résultats des comparaisons directes et ceux des comparaisons indirectes, ou si le résultat de la comparaison indirecte ne varie pas en fonction du « chemin » emprunté. Présentation des résultats : présence d'une représentation graphique du réseau de preuves et d'un tableau récapitulatif avec les études incluses, les interventions, et les résultats observés avec chaque intervention dans chaque étude.

Annexe 3 : Résumé de l'analyse de référence HAS⁹².

L'analyse de référence distingue les références « à respecter », auxquelles l'auteur de l'évaluation doit s'astreindre, et les références « à privilégier » qui autorisent le recours à une méthode différente dès lors qu'elle est clairement justifiée.

	Analyse de référence	Statut
La méthode d'évaluation économique	Analyse coût-utilité ou coût-efficacité en fonction de la nature des effets des interventions sur la santé	Référence à respecter
	<ul style="list-style-type: none"> - Si la qualité de vie est une conséquence importante, l'ACU est privilégiée. - Si la qualité de vie n'est pas une conséquence importante, l'ACE est privilégiée. 	Références à privilégier
La perspective <ul style="list-style-type: none"> - Sur les coûts - Sur les résultats 	Perspective collective <ul style="list-style-type: none"> - Tous financeurs - Population dont la santé est affectée (identification et mesure) et population générale (scores de préférence) 	Référence à respecter
La population d'analyse	Ensemble des individus concernés directement ou de manière induite	Référence à privilégier
Les interventions à comparer	Toutes les interventions en concurrence avec l'intervention étudiée sont identifiées. La sélection des interventions comparées est de la responsabilité de l'auteur, qui argumente son choix.	Référence à respecter
L'horizon temporel	Horizon temporel suffisamment long pour intégrer l'ensemble des différentiels de coût et de résultat attendus	Référence à respecter
L'actualisation	L'actualisation adopte le taux d'actualisation public fixé à 4 % au moment de l'édition de ce guide et considère que le prix relatif du résultat de santé pour la collectivité est invariant au cours du temps. Le taux d'actualisation décroît après 30 ans jusqu'à 2%.	Référence à respecter
Synthèse des données	- Basée sur une revue systématique et critique des études cliniques et économiques	Référence à respecter
	- Tous types d'études, sous réserve de leur pertinence, de leur capacité à limiter les biais et à rendre compte de la réalité des pratiques	Référence à respecter
	- Données françaises	Référence à privilégier
Critère de résultat	- Le QALY dans les ACU	Référence à respecter
	- La durée de vie dans les ACE	Référence à privilégier
Critère de coût	Coûts de production	Référence à respecter
Conclusion de l'évaluation	- Calcul d'un ratio différentiel coût-résultat sur les interventions non dominées	Référence à respecter
	- Analyse des transferts de dépenses entre financeurs	Référence à privilégier
Analyse critique de l'évaluation	- Analyse de la variabilité et de l'incertitude, quelle qu'en soit la source	Référence à respecter
	- Discussion des conclusions et des limites de l'évaluation	Référence à respecter

Bibliographie

- 1 OCDE (2013), Panorama de la santé 2013 : Les indicateurs de l'OCDE, Éditions OCDE. http://dx.doi.org/10.1787/health_glance-2013-fr.
- 2 Direction de la Sécurité sociale. Les Comptes de la Sécurité Sociale - Rapport - Résultats 2013, Prévisions 2014 et 2015. 2014 <http://www.securite-sociale.fr/IMG/pdf/rapport-ccss-20141110-101730-392-97.pdf>.
- 3 Inspection Générale des Affaires Sociales. Rapport 2012 IGAS, L'hôpital. 2012 http://www.igas.gouv.fr/IMG/pdf/L_hopital_-_Rapport_IGAS_2012.pdf.
- 4 Haute Autorité de Santé. Rapport de synthèse. Les comparaisons indirectes : Méthodes et validités. Juillet 2009.
- 5 Wells GA. Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis. Canadian Agency for Drugs and Technologies in Health, 2009.
- 6 Jansen JP, Trikalinos T, Cappelleri JC, et al. Indirect Treatment Comparison/Network Meta-Analysis Study Questionnaire to Assess Relevance and Credibility to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report. *Value Health* 2014; 17: 157–73.
- 7 Glenny A-M, Altman DG, Song F, et al. Indirect comparisons of competing interventions. Gray Publishing, 2005 http://www.journalslibrary.nihr.ac.uk/___data/assets/pdf_file/0005/59963/ExecutiveSummary-hta9260.pdf (accessed Aug 23, 2014).
- 8 Song F. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003; 326: 472–472.
- 9 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: 683–91.
- 10 Higgins JP, Green S, others. Cochrane handbook for systematic reviews of interventions. Wiley Online Library, 2008 <http://onlinelibrary.wiley.com/doi/10.1002/9780470712184.fmatter/summary> (accessed Jan 17, 2015).
- 11 Song F, others. What is indirect comparison. *Hayward Med Commun* 2009. http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/What_is_ind_com_p.pdf (accessed Jan 17, 2015).
- 12 Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331: 897–900.

- 13 Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23: 3105–24.
- 14 Moayyedi P, Soo S, Deeks J, Delaney B, Innes M, Forman D. Pharmacological interventions for non-ulcer dyspepsia. *Cochrane Database Syst Rev* 2006; : CD001960.
- 15 Evelinda T, Devidas M, Assessment CCO for HT, Jean-François B. Selective Serotonin Reuptake Inhibitors (SSRIs) for Major Depression. Canadian Coordinating Office for Health Technology Assessment, 1997.
- 16 Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002; 21: 2313–24.
- 17 Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17: 279–301.
- 18 Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med* 2013; 159: 130–7.
- 19 Elliott WJ, Meyer PM. Incident diabetes in clinical trials of antihypertensive drugs: a network meta-analysis. *The Lancet* 2007; 369: 201–7.
- 20 Goodman SN. Toward evidence-based medical statistics. 1: The P value fallacy. *Ann Intern Med* 1999; 130: 995–1004.
- 21 Van der Valk R, Webers CAB, Lumley T, Hendrikse F, Prins MH, Schouten JSAG. A network meta-analysis combined direct and indirect comparisons between glaucoma drugs to rank effectiveness in lowering intraocular pressure. *J Clin Epidemiol* 2009; 62: 1279–83.
- 22 Psaty BM, Lumley T, Furberg CD, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents: A network meta-analysis. *JAMA* 2003; 289: 2534–44.
- 23 Jeffreys H. The theory of probability. Oxford University Press, 1998 <https://books.google.fr/books?hl=fr&lr=&id=vh9Act9rtzQC&oi=fnd&pg=PA1&dq=jeffreys+theory+of&ots=fdRxJSZ6j-&sig=d9GKmaqL-0p0nTcB3XWTdqAumrU> (accessed Feb 21, 2015).
- 24 Bernardo JM, Smith AFM. Bayesian Theory. John Wiley & Sons, 2009.
- 25 Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. John Wiley & Sons, 2004.
- 26 Sutton AJ, Abrams KR, Jones DR, Jones DR, Sheldon TA, Song F. Methods for meta-analysis in medical research. J. Wiley, 2000 <http://tocs.ulb.tu-darmstadt.de/95518479.pdf> (accessed Jan 17, 2015).

- 27 Abrams K, Myles J, Spiegelhalter D. Bayesian approaches to clinical trials and health-care evaluation. Wiley, Chichester, UK, 2004 <http://www.library.wisc.edu/selectedtocs/be209.pdf> (accessed Jan 17, 2015).
- 28 Salanti G, Ades AE, Ioannidis JPA. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; 64: 163–71.
- 29 Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: a multiple-treatments meta-analysis. *Lancet* 2011; 378: 1306–15.
- 30 Sterne JA, Egger M, Moher D. Addressing reporting biases. *Cochrane Handb Syst Rev Interv Cochrane Book Ser* 2008; : 297–333.
- 31 Aaron SD, Fergusson D, Marks GB, et al. Counting, analysing and reporting exacerbations of COPD in randomised controlled trials. *Thorax* 2008; 63: 122–8.
- 32 Mills EJ, Druyts E, Ghement I, Puhan MA. Pharmacotherapies for chronic obstructive pulmonary disease: a multiple treatment comparison meta-analysis. *Clin Epidemiol* 2011; 3: 107–29.
- 33 Eckert L. La méta-régression comme méthodologie de comparaisons indirectes : une réponse aux exigences contemporaines de l'évaluation du médicament. Paris 11, 2007 <http://www.theses.fr/2007PA11T029> (accessed Jan 17, 2015).
- 34 Thompson SG, Higgins J. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002; 21: 1559–73.
- 35 Schmid CH, Stark PC, Berlin JA, Landais P, Lau J. Meta-regression detected associations between heterogeneous treatment effects and study-level, but not patient-level, factors. *J Clin Epidemiol* 2004; 57: 683–97.
- 36 Cooper NJ, Sutton AJ, Morris D, Ades AE, Welton NJ. Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med* 2009; 28: 1861–81.
- 37 Salanti G, Marinho V, Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol* 2009; 62: 857–64.
- 38 Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *J R Stat Soc Ser A Stat Soc* 2010; 173: 613–29.
- 39 Coory M, Jordan S. Frequency of treatment-effect modification affecting indirect comparisons: a systematic review. *Pharmacoeconomics* 2010; 28: 723–32.

- 40 Mills EJ, Wu P, Chong G, et al. Efficacy and safety of statin treatment for cardiovascular disease: a network meta-analysis of 170,255 patients from 76 randomized trials. *QJM Mon J Assoc Physicians* 2011; 104: 109–24.
- 41 Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman HI, Anti-Lymphocyte Antibody Induction Therapy Study Group. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002; 21: 371–87.
- 42 Song F, Loke YK, Walsh T, Glenny A-M, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *Bmj* 2009; 338. <http://www.bmj.com/content/338/bmj.b1147> (accessed Aug 15, 2014).
- 43 Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557–60.
- 44 Cochrane handbook for systematic reviews of interventions/chapter_9/9_5_2_identifying_and_measuring_heterogeneity. http://handbook.cochrane.org/chapter_9/9_5_2_identifying_and_measuring_heterogeneity.htm.
- 45 Ades AE, Lu G, Higgins JPT. The Interpretation of Random-Effects Meta-Analysis in Decision Models. *Med Decis Making* 2005; 25: 646–54.
- 46 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: 455–63.
- 47 Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence Synthesis for Decision Making 2 A Generalized Linear Modeling Framework for Pairwise and Network Meta-analysis of Randomized Controlled Trials. *Med Decis Making* 2013; 33: 607–17.
- 48 Mills EJ, Ioannidis JA, Thorlund K, Schünemann HJ, Puhan MA, Guyatt GH. How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA* 2012; 308: 1246–53.
- 49 Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010; 29: 932–44.
- 50 Lu G, Ades AE. Assessing Evidence Inconsistency in Mixed Treatment Comparisons. *J Am Stat Assoc* 2006; 101: 447–59.
- 51 Song F, Xiong T, Parekh-Bhurke S, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. *BMJ* 2011; 343: d4909–d4909.

- 52 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: 261–82.
- 53 Khan A, Khan SR, Leventhal RM, Brown WA. Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: a replication analysis of the Food and Drug Administration Database. *Int J Neuropsychopharmacol Off Sci J Coll Int Neuropsychopharmacol CINP* 2001; 4: 113–8.
- 54 Kirsch I, Moore TJ, Scoboria A, Nicholls SS. The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prev Treat* 2002; 5. DOI:10.1037/1522-3736.5.1.523a.
- 55 Williams JW, Mulrow CD, Chiquette E, Noël PH, Aguilar C, Cornell J. A systematic review of newer pharmacotherapies for depression in adults: evidence report summary. *Ann Intern Med* 2000; 132: 743–56.
- 56 Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373: 746–58.
- 57 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; 358: 252–60.
- 58 Ioannidis JPA. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med PEHM* 2008; 3: 14.
- 59 Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med* 1996; 15: 2733–49.
- 60 Ioannidis JPA. Ranking antidepressants. *Lancet* 2009; 373: 1759–60; author reply 1761–2.
- 61 Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann Intern Med* 2011; 155: 772–85.
- 62 Jansen JP, Fleurence R, Devine B, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. *Value Health* 2011; 14: 417–28.
- 63 Kyrgiou M, Salanti G, Pavlidis N, Paraskevidis E, Ioannidis JPA. Survival benefits with diverse chemotherapy regimens for ovarian cancer: meta-analysis of multiple treatments. *J Natl Cancer Inst* 2006; 98: 1655–63.

- 64 Dwan K, Gamble C, Kolamunnage-Dona R, Mohammed S, Powell C, Williamson PR. Assessing the potential for outcome reporting bias in a review: a tutorial. *Trials* 2010; 11: 52.
- 65 Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009; 151: W – 65.
- 66 Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. *The Lancet* 2013; 382: 951–62.
- 67 Higgins J, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration’s tool for assessing risk of bias in randomised trials. *Bmj* 2011; 343. <http://www.bmj.com/content/343/bmj.d5928.full> (accessed Jan 17, 2015).
- 68 Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17: 1–12.
- 69 Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. *Ann Intern Med* 2008; 148: 544–53.
- 70 Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR. Bayesian methods in health technology assessment: a review. *Core Research*, 2000 <http://www.journalslibrary.nihr.ac.uk/hta/volume-4/issue-38> (accessed Feb 21, 2015).
- 71 Sung L, Hayden J, Greenberg ML, Koren G, Feldman BM, Tomlinson GA. Seven items were identified for inclusion when reporting a Bayesian analysis of a clinical study. *J Clin Epidemiol* 2005; 58: 261–8.
- 72 Pullenayegum EM, Guo Q, Hamilton SJH, Hopkins RB. Developing critical thinking about reporting of Bayesian analyses. *J Stat Educ* 2012; 20: n1.
- 73 Gartlehner G, Morgan LC, Thieda P, et al. Drug Class Review: Second Generation Antidepressants: Final Report Update 4. 2008. <http://europepmc.org/abstract/MED/21089243> (accessed Jan 17, 2015).
- 74 Mills EJ, Thorlund K, Bansback N. Multiple treatment meta-analysis in rheumatoid arthritis can provide weak inferences due to methodological approaches. 2012.
- 75 Bouvenot G, Vray M, Vicari F. ESSAIS cliniques—Théorie, pratique et critique. *Acta Endosc* 2006; 36: 375–375.
- 76 Sylvain AFFINITO. Intégration de l’évaluation médico-économique des médicaments en France au sein de la HAS : impact sur l’accès et le maintien des médicaments sur le marché français. 2014; published online Avril.
- 77 Haute Autorité de Santé. Rapport d’activité 2013. http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-06/rapport_acitivite_has_2013.pdf (accessed Jan 17, 2015).

- 78 Gilles Bouvenot. Le rôle de la Commission de la transparence de la HAS dans la pertinence de l'admission au remboursement des médicaments par l'Assurance maladie. Regards Numéro 44. 2013; published online Nov. <http://www.en3s.fr/articles-regards/44/Bouvenot.pdf> (accessed Jan 17, 2015).
- 79 Accord cadre du 5 décembre 2012 entre le Comité économique des produits de santé et les entreprises du médicament. 2012 http://www.sante.gouv.fr/IMG/pdf/accord_cadre_du_051212.pdf.
- 80 COUR DES COMPTES. LA SÉCURITÉ SOCIALE. SEPTEMBRE 2007. http://www.mutuelles-de-france.fr/autonomiz/uploads/_FMF/template2_84/rapport_securite_sociale_2007.pdf (accessed March 29, 2015).
- 81 LOI n° 2011-2012 du 29 décembre 2011 relative au renforcement de la sécurité sanitaire du médicament et des produits de santé. 2011.
- 82 Chapple CR, Khullar V, Gabriel Z, Muston D, Bitoun CE, Weinstein D. The effects of antimuscarinic treatments in overactive bladder: an update of a systematic review and meta-analysis. *Eur Urol* 2008; 54: 543–62.
- 83 Chapple C, Khullar V, Gabriel Z, Dooley JA. The effects of antimuscarinic treatments in overactive bladder: a systematic review and meta-analysis. *Eur Urol* 2005; 48: 5–26.
- 84 Scholz H, Trenkwalder C, Kohnen R, Riemann D, Kriston L, Hornyak M. Dopamine agonists for restless legs syndrome. *Cochrane Database Syst Rev* 2011; : CD006009.
- 85 Hornyak M, Trenkwalder C, Kohnen R, Scholz H. Efficacy and safety of dopamine agonists in restless legs syndrome. *Sleep Med* 2012; 13: 228–36.
- 86 HAS. Avis de la Commission de la transparence du 7 mai 2014 TECFIDERA, CT 13005. http://www.has-sante.fr/portail/jcms/c_1743058/fr/tecfidera-dimethyl-fumarate-traitement-de-fond-de-la-sclerose-en-plaques.
- 87 HAS. Avis de la Commission de la transparence TIVICAY du 28 mai 2014, CT13530. http://www.has-sante.fr/portail/jcms/c_1748396/en/tivicay.
- 88 Mickael F. Drummond, Bernie J. O'Brien, Greg L. Stoddart, George W. Torrance. Méthodes d'évaluation économique des programmes de santé. 1998.
- 89 Cost-effectiveness thresholds. World Health Organ. http://www.who.int/choice/costs/CER_thresholds/en/ (accessed March 29, 2015).
- 90 Jeantet Marine. Lopez Alain. Evaluation médico-économique en santé. Rapport IGAS (Inspection générale des affaires sociales). 2014 <http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/154000078/0000.pdf>.
- 91 Haute Autorité de Santé. Décision n°2013.0111/DC/SEESP du 18 septembre 2013 du collège de la Haute Autorité de santé relatif à l'impact significatif sur les dépenses de

- l'assurance maladie déclenchant l'évaluation médico-économique des produits revendiquant une ASMR ou une ASA de niveaux I, II ou III. http://www.has-sante.fr/portail/upload/docs/application/pdf/2013-09/c_2013_0111_definition_impact_significatif.pdf.
- 92 Haute Autorité de Santé. Choix méthodologiques pour l'évaluation économique à la HAS. 2011 http://www.has-sante.fr/portail/upload/docs/application/pdf/2011-11/guide_methodo_vf.pdf.
- 93 Haute Autorité de Santé. Valeurs de références pour l'évaluation économique en santé. Revue de la littérature. 2014 http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-12/valeurs_de_reference_vf.pdf.
- 94 APM News. La Ceesp favorable à plusieurs évolutions de l'évaluation médico-économique des produits de santé. 2014; published online décembre.
- 95 Haute Autorité de Santé. Dépistage de l'infection par le VIH en France. Stratégies et dispositifs de dépistage. 2009 http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-10/synthese_depistage_vih_volet_2_vfv_2009-10-21_16-48-3_460.pdf.
- 96 Goodacre S, Nicholl J, Dixon S, et al. Randomised controlled trial and economic evaluation of a chest pain observation unit compared with routine care. *BMJ* 2004; 328: 254.
- 97 Briggs AH, O'Brien BJ, Blackhouse G. Thinking outside the box: recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annu Rev Public Health* 2002; 23: 377–401.
- 98 HAS. Avis d'efficience KADCYLA du 11 mars 2014. 2014; published online March 11. http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-12/kadcyla_11032014_avis_efficience.pdf.
- 99 HAS. Projet d'avis d'efficience LEMTRADA du 15 avril 2014 (demande retirée). 2014; published online avril. http://www.has-sante.fr/portail/upload/docs/application/pdf/2015-01/lemtrada_15042014_retrait_ceesp.pdf.
- 100 HAS. Avis d'efficience SOVALDI du 15 avril 2014. http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-12/sovaldi_15042014_avis_efficience.pdf.
- 101 HAS. Avis d'efficience TIVICAY du 27 mai 2014. 2014; published online mai. http://www.has-sante.fr/portail/upload/docs/application/pdf/2014-12/tivicay_27052014_avis_efficience.pdf.
- 102 HAS. Avis d'efficience ADEMPAS du 14 octobre 2014. http://www.has-sante.fr/portail/upload/docs/application/pdf/2015-02/adempas_14102014_avis_efficience.pdf.

- 103 Le groupe de travail sur l'évaluation des médicaments devra rendre ses conclusions fin juin (EH1NM15EF). APM News 2015; published online March 30.
- 104 Laws A, Kendall R, Hawkins N. A comparison of national guidelines for network meta-analysis. *Value Health J Int Soc Pharmacoeconomics Outcomes Res* 2014; 17: 642–54.
- 105 EUnetHTA. Methodological guideline for REA of pharmaceuticals: Direct and indirect comparison. 2013
<http://www.eunetha.eu/sites/5026.fedimbo.belgium.be/files/Direct%20and%20indirect%20comparisons.pdf>.
- 106 Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med* 2013; 11: 159.
- 107 Hoaglin DC, Hawkins N, Jansen JP, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 2. *Value Health* 2011; 14: 429–37.