



ÉVALUATION MÉDICO-ÉCONOMIQUE DES BIOTHÉRAPIES DANS LA PRISE EN CHARGE DE LA POLYARTHRITE RHUMATOÏDE EN FRANCE

MÉMOIRE TECHNIQUE

Pr Robert Launois

Soumaia Ben Hariz, MSc

Anastasiia Kabeshova, PhD

Préparé pour la HAS

Avril 2016

SOMMAIRE

<u>LISTE DES ACRONYMES</u>	<u>5</u>
<u>LISTE DES TABLEAUX</u>	<u>7</u>
<u>LISTE DES FIGURES</u>	<u>7</u>
<u>1 CONTEXTE</u>	<u>8</u>
1.1 LA PATHOLOGIE	8
1.2 LES OPTIONS THERAPEUTIQUES	8
1.3 LES MODALITES DE PRISE EN CHARGE	11
1.3.1 PHASE I : PRISE EN CHARGE INITIALE	11
1.3.2 PHASE II : PRISE EN CHARGE APRES REPONSE INADEQUATE AU PREMIER TRAITEMENT	11
1.3.3 PHASE III ; PRISE EN CHARGE APRES UN PREMIER ECHEC D'UN ANTI TNF	12
1.4 OBJECTIFS DE LA RECHERCHE	13
<u>2 REVUE SYSTEMATIQUE DE LA LITTERATURE</u>	<u>13</u>
2.1 PROTOCOLE DE LA RECHERCHE DOCUMENTAIRE	13
2.2 LES CRITERES A RETENIR POUR COMPARER LES TRAITEMENTS	15
2.2.1 LE CRITERE D'EFFICACITE	15
2.2.2 ESTIMATION DES ABANDONS PRECOCES	16
2.2.3 ESTIMATION DES EFFETS INDESIRABLES	16
2.2.4 APPRECIATION DE L'EFFICIENCE	16
2.3 STRATEGIE DE RECHERCHE	16
2.3.1 SELECTION DES ARTICLES	16
2.3.2 DESCRIPTION DES ETAPES DU PROCESSUS DE SELECTION	17
2.3.3 EXEMPLE DE GRILLE D'EXTRACTION SOUS EXCEL A DOCUMENTER	19
2.3.4 VERIFICATION DE L'HOMOGENEITE DES ARTICLES RETENUS	19
<u>3 META ANALYSE</u>	<u>21</u>
3.1 META-ANALYSE PAR PAIRES	21
3.1.1 MODELE A EFFETS FIXES	21
3.1.2 MODELE A EFFETS ALEATOIRES	22
3.2 META-ANALYSE BAYESIENNE	23
3.2.1 INFERENCE BAYESIENNE A POSTERIORI	23
3.2.2 META-ANALYSE MULTITRAITEMENT	24
3.3 MESURE DE LA QUANTITE D'EFFET	24
3.3.1 META-ANALYSE PAR PAIRE	24
3.3.2 META-ANALYSE MULTI-TRAITEMENT	25
3.4 METHODES DE SIMULATION MCMC	26
3.4.1 DIFFICULTE DE CALCUL DES INTEGRALES ET RE-ECHANTILLONNAGE	26
3.4.2 CONTROLE DE LA CONVERGENCE	26
3.5 INDICES THERAPEUTIQUES ESTIMES	27
3.6 ANALYSE DE NON-INFERIORITE	28
3.6.1 DEFINITIONS	28
3.6.2 CADRE CLASSIQUE	29

3.6.3	CADRE BAYESIEN	31
3.7	UN EXEMPLE DE MTC DANS LA POLYARTHRITE RHUMATOÏDE	34
4	<u>MAPPING DE L'ECHELLE PSYCHOMETRIQUE HAQ VERS LE QUESTIONNAIRE UTILITE PREFERENCE EQ-5D</u>	35
4.1	PRINCIPE DU « MAPPING »	36
4.2	SPECIFICATION DU « MAPPING »	37
4.2.1	« MAPPING » DIRECT SUR LE SCORE GLOBAL DE L'EQ-5D-3L	37
4.2.2	« MAPPING » INDIRECT SUR LES MODALITES DES CINQ DIMENSIONS DE L'EQ-5D-3L	37
4.3	MODELISATION	37
4.3.1	MAPPING SUR LE SCORE GLOBAL DE L'EQ-5D-3L	38
4.3.2	MAPPING SUR LES CINQ DIMENSIONS DE L'EQ-5D-3L	40
4.4	VALIDATION	41
5	<u>VALORISATION DES MOYENS MIS EN OEUVRE</u>	42
5.1	POINT DE VUE RETENU	42
5.2	RESSOURCES CONSOMMEES	43
5.2.1	CONSOMMATIONS MEDICALES DIRECTEMENT LIEES AUX SOINS	43
5.2.2	CONSOMMATION FAMILIALE DIRECTEMENT ASSOCIEE AUX AIDES INFORMELLES	44
5.3	VALORISATION DES COUTS	44
5.4	EXPLOITATION DES BANQUES DE DONNEES MEDICO ADMINISTRATIVES (BDMA)	44
5.5	ANALYSE STATISTIQUE DES COUTS	45
6	<u>MODELISATION MEDICO-ECONOMIQUE</u>	45
6.1	LES DIFFERENTS TYPES DE MODELISATIONS ENVISAGEABLES	46
6.1.1	ARBRES DE DECISION	46
6.1.2	MODELES DE MARKOV	47
6.1.3	MODELES INDIVIDUS CENTRES	48
6.2	HYPOTHESE DE TRAVAIL SUR LE TYPE DE MODELE A RETENIR	50
6.3	REALISATION PRATIQUE DU MODELE	51
6.3.1	APPLICATION DE LA SIMULATION INDIVIDUELLE	51
6.3.2	APPLICATION DE LA SIMULATION DES EVENEMENTS DISCRETS	54
6.4	LES DONNEES A INTRODUIRE DANS LE MODELE	55
6.4.1	POPULATION D'ANALYSE	55
6.4.2	CHOIX DE L'HORIZON TEMPOREL	55
6.4.3	CHOIX DU TAUX D'ACTUALISATION	55
7	<u>IDENTIFICATION ET QUANTIFICATION DES INCERTITUDES</u>	56
7.1	SOURCES DES INCERTITUDES	56
7.1.1	INCERTITUDES STOCHASTIQUES	56
7.1.2	INCERTITUDES PARAMETRIQUES	56
7.1.3	INCERTITUDES STRUCTURELLES	56
7.1.4	INCERTITUDES PAR HETEROGENEITE	57
7.2	REPRESENTATION DES INCERTITUDES PARAMETRIQUES	57
7.2.1	LA DISTRIBUTION NORMALE	57
7.2.2	LA DISTRIBUTION BETA	58

7.2.3	LA DISTRIBUTION DE DIRICHLET	58
7.2.4	LA DISTRIBUTION GAMMA	58
7.2.5	METHODES D'ESTIMATION DES PARAMETRES D'UNE LOI	59
7.3	TRAITEMENTS DES INCERTITUDES PARAMETRIQUES	59
7.3.1	ANALYSE DE SENSIBILITE DETERMINISTE (ASD)	59
7.3.2	LIMITES DE L'ANALYSE DE SENSIBILITE DETERMINISTE	60
7.3.3	ANALYSE DE SENSIBILITE PROBABILISTE (ASP)	60
7.4	PROPAGATION DES INCERTITUDES	61
7.4.1	SIMULATION DE MONTE-CARLO D'ORDRE 1	61
7.4.2	SIMULATION DE MONTE-CARLO D'ORDRE 2	61
8	<u>CARACTERISATION DE L'INCERTITUDE FINALE</u>	62
8.1	COURBE RETRAÇANT LES PROPORTIONS DE CAS GAGNANTS (CPCG)	62
8.2	FRONTIERE DES CHOIX QUI MAXIMISENT LE BENEFICE DE SANTE NET (FdCMax)	62
9	<u>EXPERIENCES ET COMPETENCES POUR L'EXECUTION DU PROJET</u>	63
10	<u>CALENDRIER PREVISIONNEL</u>	67
11	<u>DEROULEMENT ET ORGANISATION DE LA MISSION</u>	69
11.1	ORGANISATION	69
11.2	DEMARCHE QUALITE CHEZ REES FRANCE	69
12	<u>REFERENCES BIBLIOGRAPHIQUES</u>	71
<u>ANNEXE 1 : REALISATION D'ANALYSE DE SENSIBILITE PROBABILISTE SOUS VBA</u>		74

Liste des acronymes

ABA	L'abatacept, (ORENCIA®)
ACR	American College of Rheumatology
ADA	L'adalimumab (HUMIRA®)
AMM	Autorisations de Mise sur le Marché
ANA	L'anakinra, (KINERET®)
ASD	Analyse de sensibilité déterministe
ASP	Analyse de sensibilité probabiliste
BDNS	Bénéfice Différentiel de Santé Net
BNS	Bénéfice de Santé Net
CCAM	Classification Commune des Actes Médicaux
cDMARD	Conventional Disease Modifying Antirhumatic Drugs
CPCG	Courbe retraçant les proportions de cas gagnants
CTL	Le certolizumab (CIMZIA®)
DAS 28	Disease Activity Score
DMARD	Disease Modifying Antirhumatic Drugs
EFSA	Effort Financier Socialement Acceptable
EMEA	EMEA
EQ-5D	Questionnaire de qualité de vie utilité préférence (EuroQol)
ETA	L'étanercept (ENBREL®)
EULAR	European League Against Rheumatism
FDA	Food and Drug Administration
FdCMax	Frontière des Choix MaximisAnt le bénéfice de santé net
GHS	Groupe Homogène de Séjour
GOL	Le golimumab (SIMPONI®)
HAQ	Health Assessment Questionnaire
HAS	Haute Autorité de Santé
INF	L'infliximab (REMICADE®)
IPS	Individual Patient Simulation
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
Les anti-TNF	Les inhibiteurs du facteur de nécrose tumorale
MCMC	Monte Carlo par Chaîne de Markov
MCO	Régression par les Moindres Carrés Ordinaires
MTC	Mix Treatment Analysis
MTX	Méthotrexate
NABM	Nomenclature des Actes de Biologie Médicale
NGAP	Nomenclature Générale des Actes Professionnels
OR	Odds Ratio
PICOS	Population-Intervention-Comparateurs-Outcome-Schémas d'étude
PMSI	Programme de Médicalisation des Systèmes d'Information
PR	Polyarthrite Rhumatoïde
QALY	Années de Vie Ajustées sur la Qualité
RDCR	Ratio Différentiel Coût Résultat
RTX	Le rituximab (MABTHERA®)
SED	Simulation d'Événements Discrets

T2A

Tarification A l'Activité

TCL

Le tocilizumab, (ROACTEMRA®)

Liste des tableaux

Tableau 1 : Médicaments bénéficiant d'une AMM dans le traitement de la PR en France.....	10
Tableau 2 : Règle d'interprétation d'un odds ratio en fonction de son intervalle de confiance	25
Tableau 3 : Interprétation de la position de l'intervalle de confiance	30
Tableau 4 : Force de la preuve en fonction de la valeur de la cote d'une des hypothèses par rapport à l'autre	33
Tableau 5 : Transposition de l'échelle de Spiegelhalter en termes de probabilité a posteriori.....	33
Tableau 6 : Taux de répondeurs au critère ACR20 dans les 16 essais.....	34
Tableau 7 : Résultats du modèle MTC en termes de pourcentage de répondeurs ACR20 et de rapports de cotes .	35
Tableau 8 : Robustesse des différents modèles	41

Liste des figures

Figure 1 : Algorithme de prise en charge de la PR en 2014 selon les recommandationsde la SFR, PHASE I	11
Figure 2 : Algorithme de prise en charge de la PR en 2014 selon les recommandationsde la SFR, PHASE II	12
Figure 3 : Algorithme de prise en charge de la PR en 2014 selon les recommandationsde la SFR, PHASE III.....	13
Figure 4 : Équivalence et non-infériorité.....	29
Figure 5 : Non-infériorité dans le cadre bayésien.....	31
Figure 6. Exemple de définition d'un nouveau compteurs sur un arbre de décision.	52
Figure 7. Définition d'un nouveau nœud logique.....	52
Figure 8. Exemmmple de la définition de la distribution.	53
Figure 9. Résultats de la simulation Monte Carlo.	54
Figure 10. Extraction Excel de la matrice globale.....	54

1 Contexte

1.1 La pathologie

La prévalence de la polyarthrite rhumatoïde (PR) en France a été estimée à 0,31 % de la population âgée de plus de 18 ans en 2001 (0,51% pour les femmes, avec un sexe ratio de 5,7) (Guillemin et al., 2005). Cela correspondrait en décembre 2013, sur la base des données de l'INSEE, à 158 000 patients atteints. La maladie serait plus fréquente dans le Sud (Roux et al., 2007). Elle a une forte prédominance féminine, et son incidence annuelle est estimée à 90 cas pour 10 millions. Le pic d'incidence se situe entre 25 et 55 ans (Guillemin et al., 1994).

Au cours de la PR apparaît une inflammation de la membrane synoviale qui tapisse l'articulation, ce qui entraîne un épaissement et une prolifération de celle-ci. En l'absence de traitement, les articulations se détruisent, entraînant un handicap majeur.

La polyarthrite rhumatoïde se manifeste par le gonflement bilatéral et symétrique de nombreuses articulations. Elle touche surtout les poignets, les articulations digitales, métacarpo-phalangiennes ou interphalangiennes proximales, les genoux et les pieds. Les douleurs entraînent volontiers une insomnie. Le matin, les articulations sont raides et peuvent le rester pendant plusieurs heures. Le début est habituellement progressif.

Des symptômes extra-articulaires sont parfois présents comme une fatigue générale avec de la fièvre, une perte de poids, des nodules sous-cutanés, des atteintes cardiaques, pulmonaires ou une sécheresse de la bouche ou des yeux.

L'objectif du traitement de la polyarthrite rhumatoïde est d'induire et de maintenir une rémission clinique et de limiter la progression de la dégradation articulaire, donc le handicap ultérieur. La prise en charge a considérablement évolué au cours des dernières années du fait de l'utilisation de nouveaux traitements de fond, notamment les biothérapies (Anti TNF et autres) et l'arrivée des bio-similaires en 2014-2015.

Dans ce contexte s'inscrit ce présent projet qui a pour objectif d'identifier, par le biais d'une évaluation médico-économique, les stratégies de prise en charge les plus efficaces parmi l'ensemble des biothérapies, pour l'ensemble des indications dans le traitement de la PR.

1.2 Les options thérapeutiques

Selon les recommandations de la société Française de Rhumatologie (SFR) (Gaujoux-Viala et al., 2014) en se basant sur les recommandations de l'EULAR (Smolen et al., 2013) la prise en charge de la PR peut être illustrée dans trois phases décrite dans la Figure 3.

Les options thérapeutiques sont classées en trois catégories :

➤ LES TRAITEMENTS DE FOND SYNTHÉTIQUES :

Les traitements de fond ont pour but d'arrêter ou freiner l'évolution de la maladie, ce qui leur vaut le nom de Conventional Disease Modifying Antirhumatic Drugs (cDMARD) et sont fréquemment utilisés en première intention.

Parmi les traitements classiques de la PR, le **méthotrexate (MTX)** est le traitement de référence. À dire d'experts, 45 à 60 % des patients présentant une PR sont traités par MTX (HAS (Haute Autorité de Santé), 2006a) et 18 % d'entre eux échappent au traitement (HAS (Haute Autorité de Santé), 2005).

Traitements conventionnels (csDMARD)

- Le méthotrexate
- Le léflunomide
- La sulfasalazine

Traitement synthétique ciblé (tsDMARD)

- Le tofacitinib (Xeljanz®)

➤ LES TRAITEMENTS DE FOND BIOLOGIQUES (bDMARD)

En sus des traitements classiques, les biothérapies (aussi appelées « biological DMARD ») ont également leur place dans le traitement de la PR. Ces agents biologiques comprennent en cinq classes de biothérapies ciblant le processus inflammatoire.

Les Anti-TNF (Tumor Necrosis Factor/les inhibiteurs du facteur de nécrose tumorale): il s'agit de cinq molécules proposées en cas de réponse inadéquate ou d'intolérance aux traitements de fond conventionnels comprenant le méthotrexate. Ils peuvent aussi être employés en première intention (patients naïfs de MTX) dans certaines formes actives et sévères de PR.

- **ETA : L'étanercept** (ENBREL®), un récepteur soluble du TNF
- **ADA : L'adalimumab** (HUMIRA®), anticorps monoclonaux
- **INF : L'infliximab** (REMICADE®), anticorps monoclonaux
- **CTZ : Le certolizumab** (CIMZIA®),
- **GOL : Le golimumab** (SIMPONI®)

Dans les PR réfractaires au traitement par MTX, les anti-TNF sont efficaces dans les deux tiers des cas (Mariette, 2004). L'administration d'un autre anti-TNF en seconde intention en cas d'échec d'un premier anti-TNF n'est efficace que dans un cas sur deux (Hansen et al., 2004).

Autres traitements biologiques

- **RTX : Le rituximab** (Edwards and Cambridge, 2001; Leandro et al., 2002), (MABTHERA®), un anticorps monoclonal spécifique du **CD20**, un marqueur des lymphocytes B. Il est indiqué dans le cas d'intolérance ou réponse inadéquate à au moins un anti-TNF. Sa population cible comprend le tiers de patients traités par anti-TNF qui seront dans ce cas (HAS (Haute Autorité de Santé), 2006b). Il a la AMM qu'en cas d'échec à au moins un anti-TNF.
- **ABA : L'abatacept**, (ORENCIA®), un anti-**CD 28**, une protéine soluble, est un immunosuppresseur sélectif qui freine l'activation des lymphocytes T (Kremer et al., 2003). Ce traitement à la AMM chez les patients en échec des traitements de fond conventionnels dont le MTX, et en échec d'au moins un anti-TNF.
- **TCZ : Le tocilizumab**, (ROACTEMRA®), un anti-interleukine 6 (**IL6**). Ce traitement à la AMM chez les patients en échec des traitements de fond conventionnels, dont le MTX, et en échec d'au moins un anti-TNF.
- **ANA : L'anakinra**, (KINERET®), un anti-interleukine 1 (**IL1**), a une efficacité inférieure à celles des autres biothérapies.

➤ LES BIO-SIMILAIRES (bsDMARD)

En 2014 et 2015, deux autorisations de mise sur le marché (AMM) ont été délivrées pour des biosimilaires de l'infliximab (NFLECTRA® et REMSIMA®).

Le Tableau 1 présente les principaux traitements disponibles en France.

Tableau 1 : Médicaments bénéficiant d'une AMM dans le traitement de la PR en France

DCI	Nom commercial	Laboratoire	Indication(s) (Libellés d'AMM)
Méthotrexate	Méthotrexate Bellon®	Sanofi Aventis	Formes sévères, actives de PR de l'adulte
Méthotrexate	Métoject®	Nordic Pharma	PR active et sévère de l'adulte
Méthotrexate	Novatrex®	Wyett	PR
Léflunomide	Arava®	Sanofi Aventis	PR active – risque d'effets indésirables graves si traitement récent/concomitant avec MTX
Sulfasalazine	Salazopyrine®	Pfizer	PR
Infliximab	Remicade®	Schering Plough	PR, en association avec le MTX, chez les patients (1) ayant une maladie active lorsque la réponse aux traitements de fond, dont le MTX, a été inappropriée ; (2) ayant une maladie active, sévère et évolutive, non traitée auparavant par le MTX ni les autres DMARDs.
Etanercept	Enbrel®	Wyett	Monothérapie en cas d'intolérance au MTX ou lorsque sa poursuite est inappropriée. Traitement de la PR sévère, active et évolutive de l'adulte non précédemment traité par le MTX.
Adalimumab	Humira®	Abott	En association avec le MTX, dans le traitement de la PR (1) modérément à sévèrement active de l'adulte lorsque la réponse aux traitements de fond, y compris le MTX, est inadéquate ; (2) sévère, active et évolutive chez les adultes non précédemment traités par le MTX.
Hydroxychloroquine	Plaquénil®	Sanofi Aventis	Traitement symptomatique d'action lente de la PR.
Tiopronine	Acadione®	Sanofi Aventis	Traitement de fond de la PR.
D-pénicillamine	Trolovol®	Dexo	Traitement de fond de la PR.
Sels d'or	Allochrysin®	Genopharm	Traitement symptomatique d'action lente des rhumatismes inflammatoires chroniques, notamment de la PR.
Sels d'or	Ridauran®	Pierre Fabre	Traitement de fond de la PR de l'adulte
Ciclosporine	Néoral® / Sandimmun®	Novartis	Formes actives et sévères de PR en cas d'inefficacité, d'intolérance ou de contre-indications aux traitements classiques, y compris le MTX.
Azathioprine	Imurel®	GSK	PR sévère, chez les patients intolérants aux corticoïdes ou corticodépendants, ou dont la réponse thérapeutique est insuffisante en dépit de fortes doses de corticoïdes.
Cyclophosphamide	Endoxan®	Baxter	À faible dose, traitement des PR.
Anakinra	Kineret®	Amgen	Traitement des signes et symptômes de la PR en association avec le MTX, chez les patients dont la réponse au MTX seul n'est pas satisfaisante.
Rituximab	Mabthera®	WYETH	PR active, sévère, chez les patients adultes qui ont présenté une réponse inadéquate ou une intolérance aux traitements de fond, dont au moins un anti-TNF.
Abatacept	Orencia®	BMS	Traitement, en association au MTX, de la PR active modérée à sévère chez les patients adultes ayant eu une réponse insuffisante ou une intolérance à d'autres traitements de fond incluant au moins un anti-TNF.

DMARD : disease modifying antirhumatic drug – EIG : événement indésirable grave – MTX : méthotrexate – PR : polyarthrite rhumatoïde

1.3 Les modalités de prise en charge

1.3.1 Phase I : prise en charge initiale

La Société française de rhumatologie (SFR) a émis des recommandations en 2014 pour décrire les modalités recommandées de la prise en charge de la PR. Selon celles-ci, le médicament de référence pour une prise en charge initiale reste le MTX, en monothérapie (PR non sévère) ou en association (PR sévère).

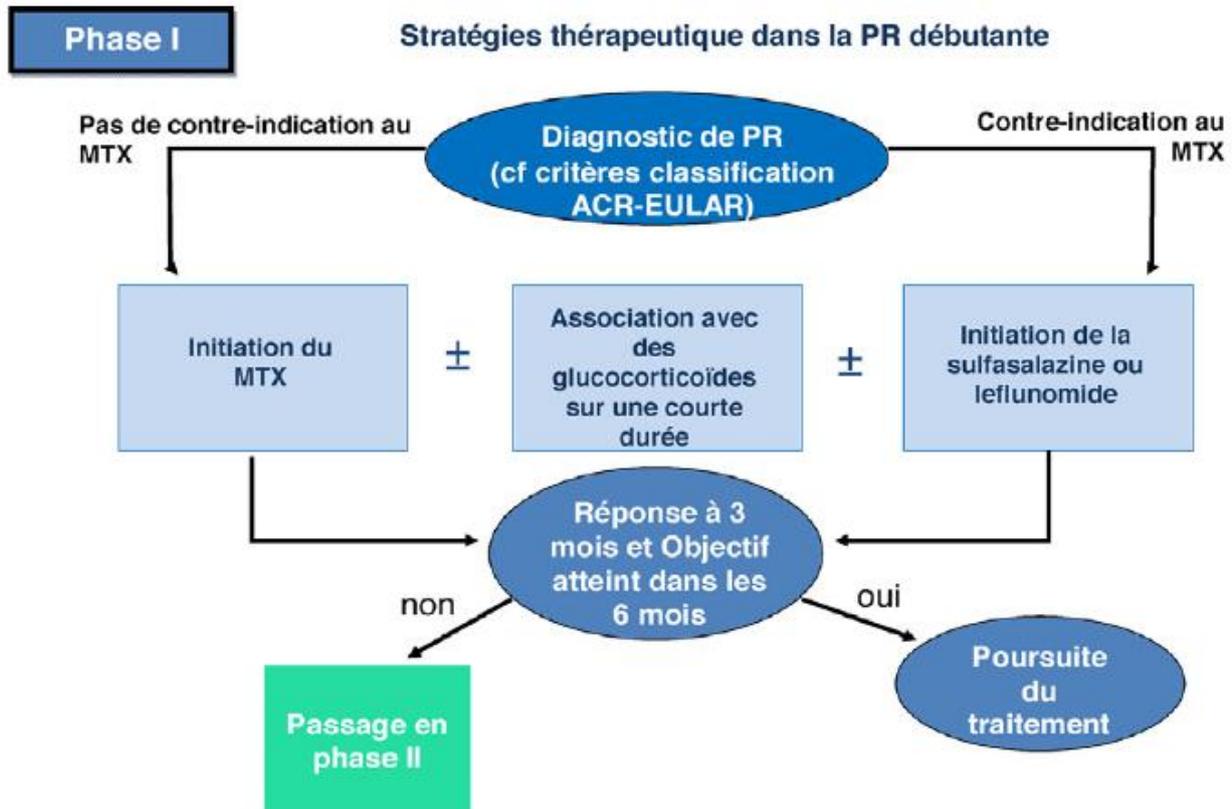


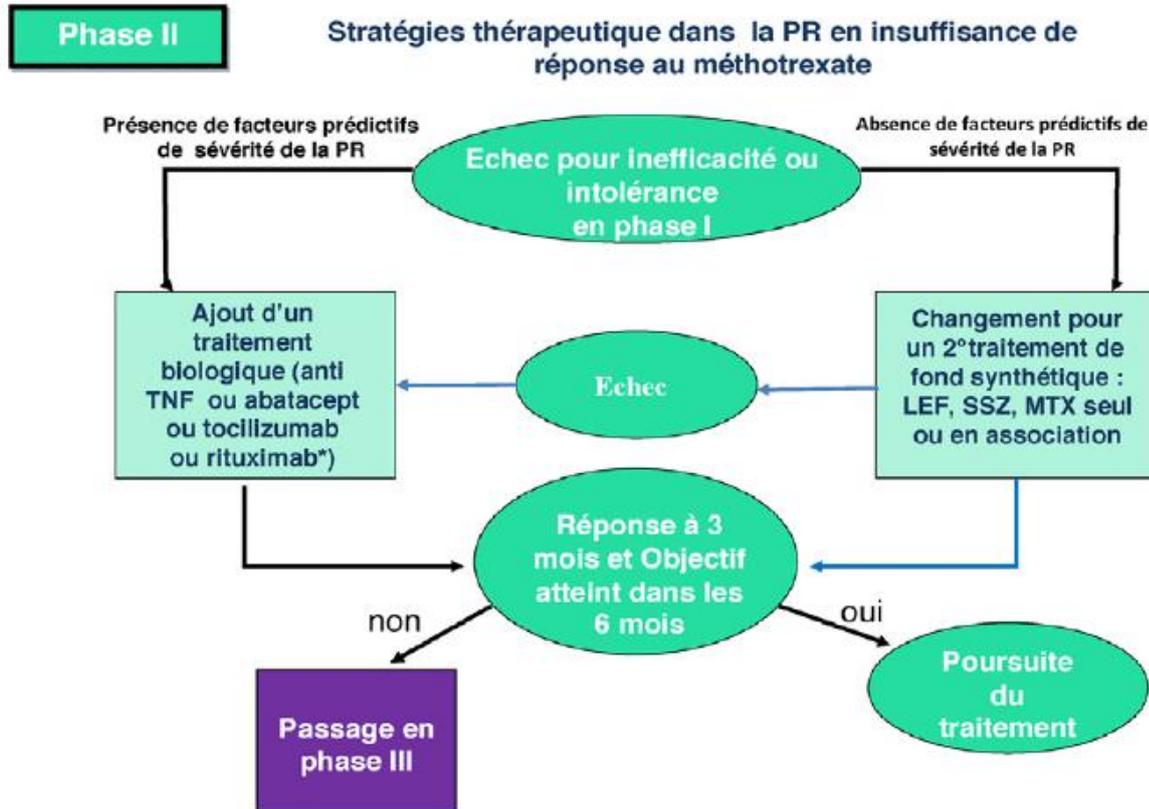
Figure 1 : Algorithme de prise en charge de la PR en 2014 selon les recommandations de la SFR, PHASE I

1.3.2 Phase II : prise en charge après réponse inadéquate au premier traitement

Plusieurs causes peuvent aboutir à un changement de traitement : la survenue d'événements indésirables, et l'insuffisance d'efficacité du traitement. L'insuffisance d'efficacité peut être le fait d'une absence de réponse au traitement (définie à partir de l'évolution du score DAS 28), d'une réponse modérée, d'un échappement thérapeutique (réponse puis aggravation), d'une progression de lésions, ou de l'impossibilité à réduire une corticothérapie sans réactiver la maladie (corticodépendance).

Le choix du traitement de seconde intention va alors dépendre du niveau d'activité de la PR.

Dans le cas d'une PR peu active, on peut commencer par envisager un traitement de fond classique (léflunomide, sulfasalazine) en monothérapie, les anti-TNF alpha n'étant envisagés qu'en cas d'échec de celui-ci. Dans le cas d'une PR active, le traitement de référence est l'association MTX/Anti-TNF α .



Facteurs prédictifs de sévérité : *atteinte ou progression structurale, activité clinique et /ou biologique élevée, taux élevé des auto-anticorps FR/ACPA.*

Rituximab* : *antécédent personnel de lymphome, tuberculose latente avec impossibilité de chimioprophylaxie, risque élevé de tuberculose, antécédent personnel de sclérose en plaque, antécédent personnel de cancer de moins de 5 ans*

Figure 2 : Algorithme de prise en charge de la PR en 2014 selon les recommandations de la SFR, PHASE II

1.3.3 Phase III : prise en charge après un premier échec d'un anti TNF

Les patients en échec d'un premier biomédicament peuvent recevoir un deuxième anti TNF ou un médicament reposant sur un autre mode d'action. Selon la SFR il n'existe pas actuellement d'argument pour privilégier un type de traitement plutôt qu'un autre. Tout dépend des antécédents du patient et des caractéristiques de la molécule. Certains résultats préliminaires suggèrent qu'un médicament avec un autre mode d'action pourrait être plus efficace que l'usage d'un nouveau traitement appartenant à la même classe thérapeutique. C'est un point important que nous nous proposons d'éclaircir dans le cadre de la méta analyse en réseau.

Phase III

Stratégies thérapeutique dans la PR en insuffisance de réponse à une biothérapie

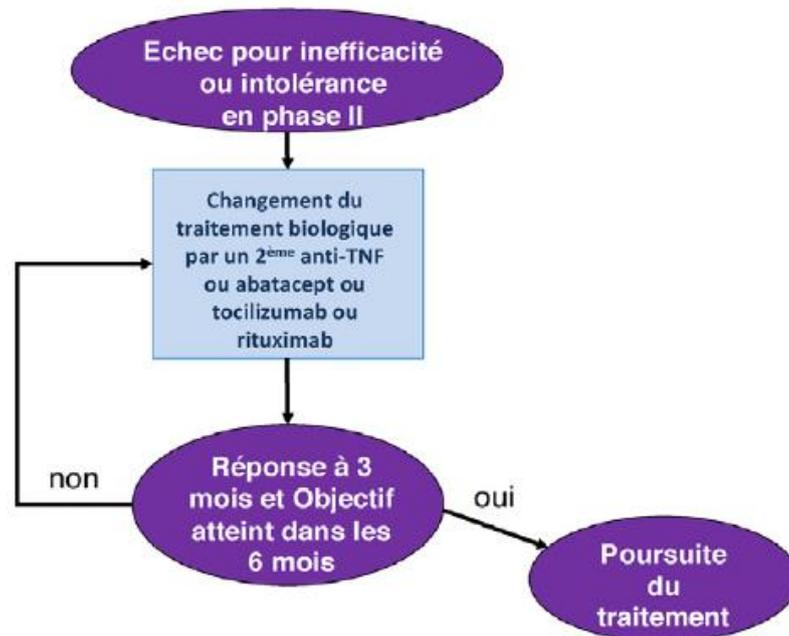


Figure 3 : Algorithme de prise en charge de la PR en 2014 selon les recommandations de la SFR, PHASE III

1.4 Objectifs de la recherche

L'évaluation du bénéfice thérapeutique attendu d'un traitement se fera en utilisant deux outils : la revue systématique de la littérature et la synthèse quantitative des données probantes. La première méthode permettra de recueillir, d'organiser, d'évaluer et de synthétiser les données de manière qualitative. La seconde permettra d'identifier de manière quantitative les séquences de traitements utilisées, les facteurs qui ont été à l'origine de leur mise en oeuvre et les résultats obtenus. Cette approche se réfère explicitement aux méta-analyses et aux méthodes de mélange de comparaison de traitements et de schémas d'étude. À la lumière des informations qui auront ainsi été colligées à partir de multiples sources : essais randomisés, cohortes prospectives, bases des données médico-administratives rétrospectives, registres prospectifs, les séquences thérapeutiques mises en oeuvre en situation réelle pourront alors être documentées en fonction des signes et symptômes qui auront été identifiés et quantifiés. Le choix du modèle à développer sera alors arrêté de façon définitive en fonction de quantité de données disponibles et mobilisables dont la pertinence sera évaluée par le groupe de travail. Son développement devra nécessairement reposer sur une vision dynamique de l'évolution du marché. L'arrivée des biosimilaires et la généralisation des modes d'administration sous-cutanée entraîneront de forts changements dans les parts de marché des différents produits qu'il conviendra d'anticiper.

2 Revue systématique de la littérature

2.1 Protocole de la recherche documentaire

De nombreux articles et recommandations ont spécifié le protocole qu'il convient de respecter pour faire une revue systématique de la littérature. La mise en oeuvre de ces techniques obéit à des règles strictes pour rechercher, sélectionner, évaluer et critiquer l'ensemble des études, publiées ou non, qui rendent le résultat de la recherche transparent et crédible (Systematic reviews : CRD's Guidance for undertaking

reviews in health care, 2009), (Cochrane Handbook for systematic review, 2008). Ces règles comportent sept étapes.

1. La stratégie de recherche qui sera adoptée aura pour but d'identifier l'ensemble des essais randomisés et les études observationnelles (cohortes, registres) réalisés entre 2005 et 2016 sur des patients souffrant de polyarthrite rhumatoïde. En principe ce travail sera fait en interne par la SEESP, mais nous demandons à y être associés. Les objectifs explicites de la recherche documentaire seront identifiés et formulés en fonction des critères PICOS. Un acronyme dont les composantes désignent: les caractéristiques de la population cible à laquelle on s'intéressera (P), ce qui suppose d'analyser les indices thérapeutiques par ligne de traitement pour étudier quelle peut être la place de chacun de cDMARD ou bDMARD dans la séquence thérapeutique; la nature de l'intervention « I » ; les comparateurs choisis « C » ; les « outcomes » (O) retenus comme critères de jugement autrement dit la mesure quantitative mise en œuvre pour estimer l'efficacité ; les schémas d'étude (S) qui sont jugés adéquats pour fournir un niveau de preuves robustes.

L'interrogation sera stratifiée en fonction des sous-groupes suivants tant sur le plan clinique qu'en ce qui concerne les études cliniques que les études médico-économiques.

- Traitements de fond prescrits chez des patients naïfs MTX en monothérapie vs triple association de traitements synthétiques avec ou sans corticothérapie
- cDMARD utilisés seuls en remplacement du MTX, ou en association vs MTX chez les patients insuffisamment répondeurs à un traitement de fond synthétique et en l'absence de facteurs de mauvais pronostic
- bDMARD+ cDMARD vs cDMARD chez les patients insuffisamment répondeurs aux traitements à un ou plusieurs traitement(s) de fond synthétiques et présentant des facteurs de mauvais pronostic,
- bDMARD utilisé en remplacement vs autres bDMARD (4 sc : adalimumab, certolizumab-pegol, etanercept, golimumab, infliximab iv chez les patients insuffisamment répondeurs aux traitements à un ou plusieurs traitement(s) anti-TNF
- bDMARD ayant un autre mode d'action vs Bdmard (IL1 Anakinra, IL6 tolicixumab, anti CD 20, anti CD 28) chez les patients insuffisamment répondeurs aux traitement(s) anti-TNF,

..

2. Formulation ex ante des critères d'éligibilité et de non-éligibilité des études au regard des critères choisis.

3. Choix des bases documentaires à interroger (deux au moins) et délimitation de la période d'analyse. Seront interrogées en première intention les bases de données électroniques MEDLINE, EMBASE et The Cochrane Central Register of Controlled Trials.

4. Identification des descripteurs correspondants (CISMef, EMTREE, MESH ou autres) et écriture des équations de recherche

5. Sélection sous END Notes, des articles qui répondent aux critères d'éligibilité, fusion, élimination des doublons, sélection sur titres et abstracts ; double lecture intégrale des articles présélectionnés pour vérification des critères d'éligibilité ; les références publiées dans les articles sélectionnés seront mobilisées pour compléter l'analyse bibliographique. Construction du diagramme de flux

6. Evaluation de la qualité des études (validité interne-externe) à la lumière des critères Grades, Srobo pour les essais randomisés et les études observationnelles et des critères de DRUMMOND (Drummond 1996) et de PHILLIPS (PHILLIPS 2004) pour les études médico-économiques et les modélisations.

7. Extraction des données sous EXCEL selon une grille standardisée définie a priori

2.2 Les critères à retenir pour comparer les traitements

2.2.1 Le critère d'efficacité

L'évolution clinique de la polyarthrite rhumatoïde commence habituellement avec des symptômes non spécifiques tels que l'épuisement, l'anorexie et la perte de poids, suivie par des symptômes spécifiques (par exemple, la douleur des articulations). En règle générale, l'affection des articulations se produit de manière symétrique et affecte principalement les articulations des mains et des genoux. Le cours de la maladie est caractérisé par des phases périodiques d'activité de la maladie, qui peuvent être mesurées par le *Disease Activity Score* (DAS) (Anderson et al., 2012). Les états de la maladie peuvent être classés en utilisant le système de notation publié par *American College of Rheumatology* (ACR) (Aletaha et al., 2010; Arnett et al., 1988). À long terme, l'inflammation chronique conduit à l'endommagement des articulations et à la destruction du cartilage. L'incapacité fonctionnelle des patients atteints de PR augmente et est souvent mesurée par l'indice d'invalidité de *Health Assessment Questionnaire* (HAQ). Le HAQ est l'un des instruments standard de l'évaluation de la qualité de vie dans la rhumatologie qui donne une impression détaillée des déficiences fonctionnelles spécifiques. Une mortalité prématurée de la PR est principalement causée par la maladie, mais peut également être influencée par des événements indésirables des médicaments antirhumatismaux (Riise et al., 2001).

Il ressort de l'étude de Durieux et al. (Durieux S, Saraux A, Guillemin F, 2003) que les trois critères les plus utilisés pour mesurer la sévérité de la PR sont : la classe fonctionnelle ACR, le score HAQ et le score DAS 28.

Les critères de l'American College of Rheumatology (ACR) (Felson et al., 1993) sont surtout cliniques et la présence des quatre premiers critères permet de diagnostiquer une polyarthrite rhumatoïde :

- Raideur des articulations le matin au réveil d'une durée supérieure à 1h depuis au moins 6 semaines.
- Inflammation articulaire d'au moins trois articulations depuis au moins 6 semaines.
- Inflammation articulaire d'au moins une des régions suivantes : poignets, articulations métacarpo-phalangiennes ou interphalangiennes proximales depuis au moins 6 semaines.
- Inflammation articulaire simultanée, bilatérale et symétrique depuis au moins 6 semaines.
- Signes radiologiques aux mains ou aux poignets (érosions ou décalcifications osseuses non équivoques).
- Présence de nodules sous-cutanés.
- Présence de facteurs rhumatoïdes dans le sang.

Le score HAQ (Health Assessment Questionnaire) (Fries et al., 1980; Guillemin et al., 1991) permet, quant à lui, d'évaluer les capacités fonctionnelles des patients et de définir leurs aptitudes dans la réalisation des actes de la vie quotidienne. Ce score évalue 8 domaines ou catégories d'activités dans la vie quotidienne sur 20 questions.

La sévérité de la polyarthrite rhumatoïde peut également être évaluée par le critère DAS 28 (Disease Activity Score) (van der Heijde et al., 1990) qui est couramment utilisé dans les études cliniques. Il s'agit d'un index linéaire qui se situe entre 0 et 10, qui tient compte du nombre d'articulations tuméfiées et du nombre d'articulations douloureuses à la pression parmi les 28 sites proposés par l'EULAR ainsi que la vitesse de sédimentation. Au final, si le DAS est inférieur à 2, le patient est en rémission ; s'il est compris

entre 2 et 3,2 l'activité de la maladie est faible ; s'il est compris entre 3,2 et 5,1 l'activité de la maladie est modérée ; et si le DAS est supérieur à 5,1, l'activité de la maladie est élevée.

Selon la disponibilité des données qui auront été relevées dans la revue de littérature, les critères ci-dessus seront étudiés. Si les essais sélectionnés mentionnent régulièrement d'autres critères d'efficacité (par exemple les données radiologiques), il pourra être envisagé de les étudier également.

2.2.2 Estimation des abandons précoces

Les abandons précoces, définis comme étant les abandons pour manque d'efficacité du traitement sont un critère important. En effet, un patient qui abandonne prématurément son traitement ne va pas consommer autant de ressources qu'un patient qui reçoit son traitement pendant toute la durée de l'étude.

2.2.3 Estimation des effets indésirables

Certains effets indésirables peuvent être sévères et donner lieu à une hospitalisation ou à un abandon du traitement. Il est important de les analyser. L'objet de notre projet est d'estimer de façon rigoureuse la fréquence de survenue des effets indésirables qui sont associés à l'usage des produits concurrents par le biais d'une revue systématique et quantitative des essais qui leur sont consacrés afin d'estimer leurs répercussions financières sur la base de données scientifiques.

2.2.4 Appréciation de l'efficacité

2.3 Stratégie de recherche

2.3.1 Sélection des articles

Les études seront sélectionnées en adoptant les critères suivants :

- Population éligible : Patients âgés de plus de 18 ans présentant une polyarthrite rhumatoïde, diagnostiqués sur la base des critères de l'American College of Rheumatology, en situation d'échec à un traitement par DMARD conventionnel (dont le méthotrexate).
- Comparateurs : cDMARD traditionnels, biothérapies, biosimilaires
- Critères de jugement : ACR ? HAQ ? DAS 28 ?
- Schémas d'étude : Essais randomisés, en double-aveugle et avec groupe contrôle. Les études observationnelles avec groupe contrôle, les revues systématiques, les méta-analyses...
- Horizon temporel : Résultats publiés entre 2000 et octobre 2016

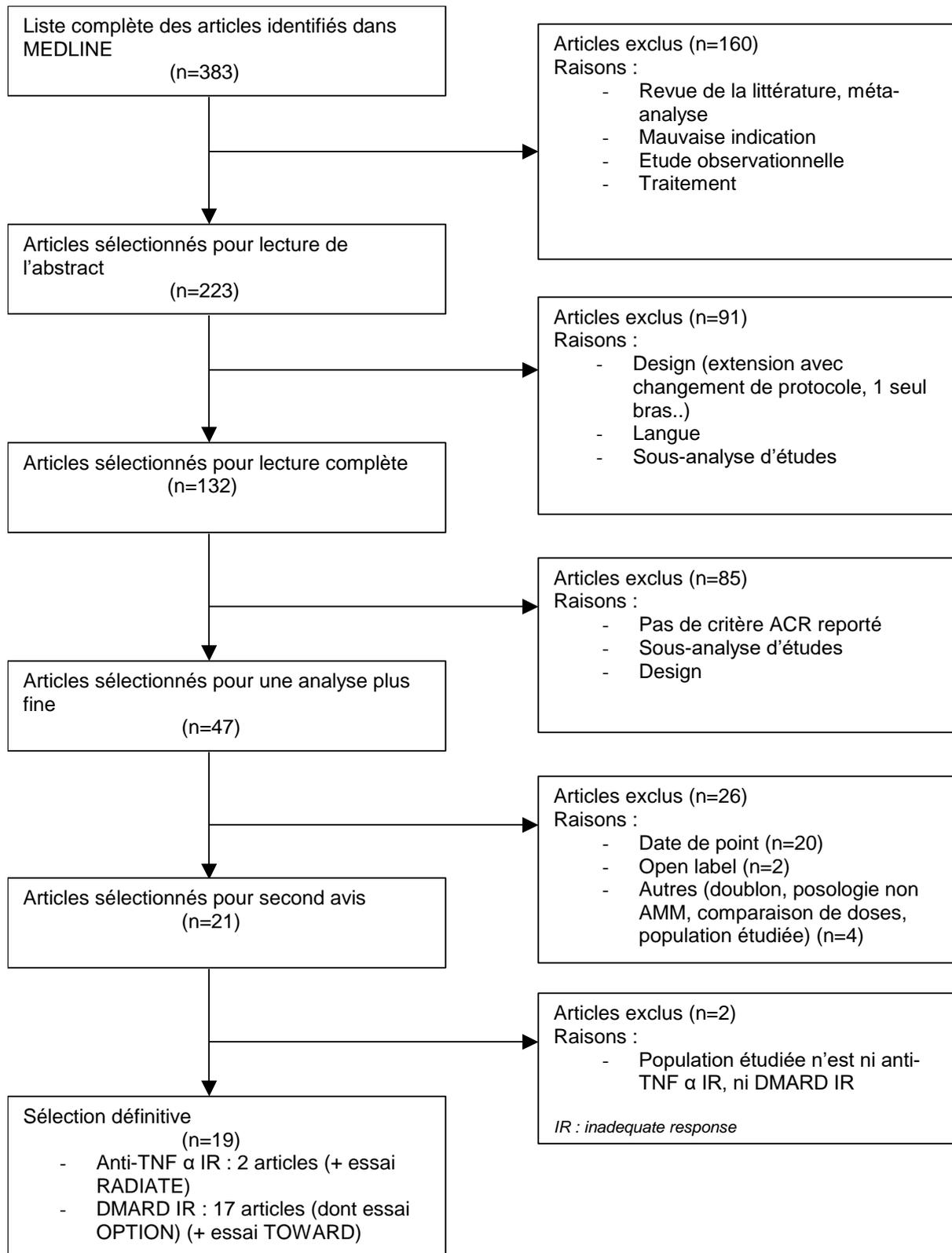
Les essais exclus à ce stade de la recherche, le seront pour une ou plusieurs des raisons suivantes :

- Population non éligible :
 - Patients naïfs de traitement de fond classique à l'inclusion, majorité de patients naïfs dans l'un des groupes de l'essai,
 - Patients atteints d'une autre pathologie que la polyarthrite rhumatoïde.
- Traitements écartés : traitements anti-TNF utilisés à des dosages différents de ceux recommandés par l'AMM.
- Design des essais : seront exclues les études open-label, les études cross-over ;

La gestion des articles qui répondront aux critères d'éligibilité sera faite sous Endnote.

2.3.2 Description des étapes du processus de sélection

Après élimination des doublons, la sélection sera opérée sur titres et abstracts et après double lecture intégrale des articles restants. Un diagramme de flux du type de celui présenté ci-après sera construit.



La qualité des études sera évaluée à l'aide de l'échelle de Jadad. Les données cliniques seront extraites selon une grille standardisée définie a priori et relevées sous EXCEL

2.3.3 Exemple de grille d'extraction sous Excel à documenter

Une première grille de lecture peut être proposée. Les informations recueillies pourront être les suivantes:

- Le nom de l'auteur, du journal et l'année de publication
- Les objectifs de l'étude
- Le type d'essai mis en place
- Les protocoles médicaux utilisés
- Les critères principaux et secondaires de jugement
- La taille de l'échantillon
- Le lieu de recrutement
- Les méthodes
 - Les critères d'inclusion
 - Les critères d'efficacité recherchés par l'étude
- La population recrutée
 - L'âge moyen ou médian
 - La proportion d'hommes
 - La dose moyenne ou médiane
 - La durée de l'étude et la période d'évaluation
- Le traitement
 - Le type de traitement
 - Les changements de dose possibles
 - Les traitements précédents
- Les résultats
 - Le nombre de patients suivis
 - Le critère principal
 - Les critères secondaires
- Les toxicités
 - Le nombre de patients suivis
 - La tolérance et la sûreté des produits évalués
- La conclusion

2.3.4 Vérification de l'homogénéité des articles retenus

La recherche de l'hétérogénéité constitue un préalable à toute méta-analyse. On cherche à détecter, parmi les essais qui respectent les critères d'inclusions choisis a priori, les spécificités qui peuvent expliquer le caractère atypique de certains des résultats observés. Ces facteurs d'hétérogénéité peuvent être d'origine

statistique, clinique, ou méthodologique sans qu'aucune de ces catégories ne soit exclusive l'une de l'autre.

Au niveau statistique, les résultats obtenus dans certains essais peuvent s'écarter très sensiblement de ceux qui sont en général relevés dans la même ligne de traitement sur des populations semblables, il convient alors de s'interroger sur le bien-fondé de leur maintien dans la synthèse quantitative que l'on se propose de réaliser.

Sur le plan clinique, il pourra s'agir d'essais présentant des caractéristiques particulières au regard des populations recrutées, des dosages administrés ou des séquences thérapeutiques étudiées. Enfin d'un point de vue méthodologique, il pourra s'agir d'études faites à des moments différents du temps avec des niveaux d'exigences évolutifs dans l'appréciation de la gravité des symptômes observés ou du nombre de réponses véritablement obtenues.

La façon la plus intuitive de détecter une hétérogénéité est de visualiser la représentation graphique du résultat des essais, notamment de leurs odds ratio et de leurs intervalles de confiance sous forme de forest plot. En complément, deux indicateurs seront utilisés pour évaluer numériquement l'hétérogénéité : le test du Q de Cochrane et l'I² de Higgins. Le test du Q de Cochrane est un test d'hypothèse classique qui évalue l'hypothèse H0 « il existe de l'hétérogénéité entre les essais ». Si la p-value obtenue est inférieure à 0.05, alors on conclura à la présence d'hétérogénéité, sinon, on rejettera cette hypothèse. L'I² de Higgins permet une interprétation plus détaillée de l'hétérogénéité. Cet indicateur mesure le pourcentage d'hétérogénéité entre les essais qui ne peut s'expliquer par des fluctuations d'échantillonnage. Un I² inférieur ou égal à 25% indique une hétérogénéité faible ; un I² autour de 50% montre une hétérogénéité modérée ; enfin, un I² supérieur ou égal à 75% indique une hétérogénéité forte. La portée de ces tests doit cependant être relativisée, ces derniers ne prenant de véritable valeur qu'en présence d'un nombre important d'essais.

L'hétérogénéité clinique ou méthodologique entraîne inévitablement une hétérogénéité statistique qu'il convient de neutraliser en mettant en œuvre les techniques de la méta-régression avant de passer à l'usage systématique de modèles à effets aléatoires.

Cette technique possède les mêmes propriétés générales qu'une régression linéaire simple : elle vise à identifier des associations entre des caractéristiques liées à l'étude ou aux individus et le paramètre estimé, ce qui n'aurait pas été possible au sein d'une seule étude.

Pour rendre cette recherche d'association possible, il faut au préalable extraire les caractéristiques potentiellement liées au paramètre des publications retenues pour la méta-analyse. Dans le cas de la polyarthrite rhumatoïde, la date de réalisation de l'étude et la dose initiale de traitement sont des paramètres qui peuvent influencer les réponses. Plus la date de réalisation de l'étude est ancienne, plus le pourcentage de patients répondeurs au traitement semble être important. De la même manière, selon la dose initiale du traitement, la réponse peut être potentiellement modifiée.

Si l'on note la date de réalisation de l'étude i par x_i , le modèle de méta-régression s'écrit comme suit :

$$P_i \sim N(\delta_i, \sigma_i^2) \text{ avec } \delta_i = a_i + \beta x_i \text{ et } \theta_i \sim N(md, \tau^2), i = 1, \dots, k$$

$$md \sim [-, -], \tau^2 \sim [-, -] \text{ et } \beta \sim [-, -]$$

β est le coefficient de régression permettant d'évaluer l'impact de la date de réalisation de l'étude sur l'effet du traitement. Une distribution *a priori* doit être spécifiée pour ce coefficient, en plus des distributions sur μ et τ^2 déjà nécessaires pour un modèle à effets aléatoires bayésien.

Malheureusement, il risque d'être impossible d'utiliser cette technique dans le cadre de notre travail compte tenu du nombre trop faible d'essais disponibles. Pour la mettre en œuvre il faudrait fallu disposer pour chaque traitement d'au moins 15 essais (Higgins and Thompson, 2002).

Après avoir vérifié l'homogénéité des essais disponibles, la réalisation d'une méta analyse sera décidée. Divers modèles (à effets fixes ou à effets aléatoires) et plusieurs paradigmes (fréquentiste ou bayésien) pourront être envisagés pour la réaliser. Nous présenterons successivement les différents modèles d'effets envisageables.

3 Méta analyse

3.1 Méta-analyse par paires

Dans une méta-analyse classique, nous disposons d'un certain nombre d'essais cliniques comparant le traitement T à un comparateur C commun. Il existe donc plusieurs comparaisons directes entre T et C, qu'il est possible de synthétiser. Dans le cas d'un critère de jugement binaire (e.g. critère ACR 20 rempli), on peut raisonner en rapport de cote ou en log de rapport de cotes.

Pour chaque essai, nous partirons du tableau de contingence type suivant :

	ACR 20 non rempli	ACR 20 rempli
Traitement contrôle C	a	b
Traitement testé T	c	d

La probabilité d'atteinte de l'ACR 20, critère le plus souvent rapporté dans les essais, est donnée pour le groupe traité par $P_T = \frac{d}{d+c}$

L'odds ratio (OR), mesure d'efficacité relative, se mesure quant à lui par $OR_{T/C} = \frac{\frac{P_T}{1-P_T}}{\frac{P_C}{1-P_C}} = \frac{ad}{bc}$

3.1.1 Modèle à effets fixes

Dans un modèle à effets fixes, on postule qu'il existe un effet du traitement d unique et commun à tous les essais méta-analysés. L'effet du traitement dans l'essai i , dans les modèles les plus simples, est présumé distribué selon une loi normale de variance V et de moyenne d .

Un modèle plus complexe, mais plus approprié dans le cas où l'on dispose de peu d'information est basé sur le modèle logistique. On connaît alors r_{Ti} et r_{Ci} , le nombre de succès (patients remplissant le critère ACR 20, par exemple) pour les traitements T et C dans l'essai i . Ce nombre de succès suit une loi binomiale fonction des effectifs dans chacun des bras de l'essai et de la probabilité de succès pour chacun des traitements de l'essai.

À l'échelle logarithmique, on peut écrire, pour le groupe contrôle de l'essai i :

$$\text{logit}(p_{Ci}) = \log\left(\frac{p_{Ci}}{1-p_{Ci}}\right) = \mu_i$$

Pour le groupe traité, on a :

$$\text{logit}(p_{Ti}) = \mu_i + d$$

Où d est l'effet du traitement (ici, mesuré en tant que logarithme d'OR), commun à tous les essais.

3.1.2 Modèle à effets aléatoires

Lorsque l'on dispose de plusieurs essais cliniques évaluant le même traitement, il est courant d'observer que l'effet estimé du traitement varie d'un essai à l'autre. Le modèle à effets fixes postule que cette variation entre les essais n'est due qu'à la fluctuation d'échantillonnage : il y a *homogénéité* des effets estimés du traitement. Il est toutefois possible que la variation entre les essais soit également due à des différences systématiques entre les essais. Dans ce cas, la variation entre les essais dépasse celle que l'on attendrait du simple fait de la fluctuation d'échantillonnage. Il n'y a plus un seul effet unique du traitement, mais un effet qui varie d'un essai à l'autre en raison de différences systématiques entre les essais. On parle alors d'*hétérogénéité* des effets estimés du traitement. Il existe, dans le cas de la méta-analyse classique, lorsque le nombre d'essais disponibles est suffisant, des tests statistiques d'hétérogénéité. Certains auteurs défendent toutefois qu'il est plus approprié de se passer de test et d'étudier directement l'amplitude de la variation entre les essais.

Alors que le modèle à effets fixes postule un effet constant du traitement pour tous les essais, le modèle à effets aléatoires permet de prendre en compte l'hétérogénéité entre les essais. Deux essais cliniques peuvent en effet différer entre eux dans leur design (critères d'inclusion, modalités d'administration du traitement, etc.). Ces différences se traduisent au niveau statistique par une variance de l'effet du traitement plus importante que si les essais étaient homogènes. En d'autres termes, en présence d'hétérogénéité, les intervalles de confiance issus d'un modèle à effets fixes seront artificiellement trop peu étendus.

Le modèle à effets aléatoires définit un effet du traitement intra-essai, δ_i et une distribution inter-essais de l'effet du traitement, le plus souvent selon une loi normale d'espérance d . La variance de cette loi normale mesure la variabilité inter-essais. Lorsque l'on dispose de peu d'essais (moins de 10), il devient difficile d'estimer cette variance. Dans le cadre bayésien, le choix de la distribution a priori de celle-ci risque d'avoir une influence importante sur l'estimation a posteriori.

Si l'on introduit des effets aléatoires, le modèle décrit précédemment devient :

$$\text{logit}(p_{Ti}) = \mu_i + \delta_i \text{ avec } \delta_i \sim \text{Normal}(d, \sigma^2).$$

La mise en œuvre informatique des méta-analyses classiques a été programmée à l'aide du logiciel statistique R.

3.2 Méta-analyse bayésienne

3.2.1 Inférence bayésienne a posteriori

La description ci-dessus des modèles à effets fixes et à effets aléatoires se place dans un contexte classique, fréquentiste, c'est-à-dire que les paramètres seront supposés fixes et inconnus (d et σ par exemple). Dans un paradigme bayésien, au contraire, les paramètres sont supposés aléatoires, ce qui permet d'associer une distribution de probabilité sur chacun d'entre eux. Alors que l'école classique fonde le jugement sur une succession d'échantillons virtuels dont la loi statistique est conditionnelle à des paramètres fixes, mais inconnus, l'école bayésienne considère que les paramètres sont des variables aléatoires au même titre que les grandeurs observables. Il y a donc un sens à considérer leur distribution probabiliste et à les revoir à la lumière des nouvelles informations. Cette propriété ouvre des possibilités supplémentaires en termes de modélisation. Il devient possible d'assigner des distributions a priori pour chaque paramètre inconnu, fondées sur des connaissances antérieures. En effet, avant de conduire une étude, un chercheur peut avoir formé une intuition a priori concernant l'option thérapeutique étudiée. Cette intuition peut être fondée sur sa propre expérience ou sur des études antérieures menées sur le même thème. Les modèles bayésiens permettent de combiner ces intuitions ou connaissances a priori avec les données issues de l'étude menée (grâce au théorème de Bayes). Les résultats reflètent alors l'actualisation de l'état des connaissances a priori sur l'effet de la stratégie considérée en une connaissance a posteriori sur la base des observations introduites à travers la fonction de vraisemblance.

Lorsque le critère de jugement est un pourcentage, la modélisation est simple. Par exemple, notons respectivement r_{Ti} et r_{Ci} le nombre de patients répondeurs dans les bras avec le traitement testé et le traitement de référence de l'étude i et n_{Ti} et n_{Ci} le nombre total de patients dans le bras testé et le bras de contrôle pour cette même étude i . Alors, nous pouvons supposer que r_{Ti} et r_{Ci} suivent respectivement une loi binomiale de paramètres (p_{Ti}, n_{Ti}) et (p_{Ci}, n_{Ci}) où p_{Ti} et p_{Ci} représentent les pourcentages recherchés de répondeurs dans chacun des bras. À partir de la fonction logistique de p_{Ti} et p_{Ci} , il est ensuite possible de calculer un rapport de cotes (odds ratio). C'est ce nouveau critère qui sera « poolé », car plus facilement modélisable.

Si l'on reprend les notations précédentes du modèle à effets aléatoires, le modèle bayésien se représentera ainsi :

$$r_{Ti} \sim \text{Bin}(p_{Ti}, n_{Ti})$$

$$\text{logit}(p_{Ti}) = \mu_i + \delta_i, \text{ avec } \delta_i \sim \text{Normal}(d, \sigma^2), i = 1, \dots, k$$

$$\text{avec } d \sim [-, -] \text{ et } \sigma^2 \sim [-, -]$$

Nous retrouvons exactement le même modèle que précédemment, à cette différence près que l'analyste transforme ses connaissances a priori sous la forme de deux paramètres d et σ^2 indépendants des données observables. Ce sont des opinions qui sont formulées avant toute considération des résultats des plans d'expérience mis en place. Plusieurs types de lois a priori peuvent être utilisées. Lorsqu'aucune intuition a priori sur les paramètres n'est possible, des lois a priori dites non informatives sont employées.

3.2.2 Méta-analyse multitraitement

Supposons que l'on ne dispose non plus d'un traitement et d'un comparateur, mais de plusieurs traitements $T_1, T_2, T_3, T_4, \dots$. Nous avons dans ce cas non plus seulement d'un ensemble d'essais comparant T_1 et T_2 , mais également d'essais comparant T_1 et T_3, T_2 et $T_4 \dots$

Pour gérer l'ensemble des comparaisons, nous allons avoir recours à un modèle de méta analyse multitraitement également appelé modèle de méta-analyse en réseau ou modèle de mélange de comparaisons de traitements, noté MTC (Mixed Treatment Comparisons). Les MTC combinent des sources d'information indirectes (la combinaison d'essais T_1 vs. T_2 et T_2 vs. T_3 pour estimer l'effet de T_3 par rapport à T_1 , par exemple) et des sources d'information directes (les essais comparant directement T_1 à T_3).

Les MTC supposent de raisonner en termes d'effet relatif des traitements.

Le modèle d'une MTC s'écrit de manière similaire à celui d'une méta-analyse classique. Soit P_{ik} l'efficacité absolue (la probabilité pour le patient de remplir le critère ACR 20, par exemple) du traitement k dans l'essai i . Soit b le traitement *baseline* (celui servant de comparateur). Si l'on a affaire au traitement b , alors $\text{logit}(P_{ib}) = \mu_{ib}$. Si on a affaire à un traitement k différent de b , alors $\text{logit}(P_{ik}) = \mu_{ib} + \delta_{ibk}$. δ_{ibk} est donc le log-OR du traitement k par rapport au traitement b dans l'essai i . Dans le cas d'un modèle à effets aléatoires, les δ_{ibk} sont distribués selon une loi normale d'espérance d_{bk} et de variance σ^2 . Pour simplifier le modèle, en effet, on va considérer que la variance inter-essais est constante, quels que soient les traitements comparés (hypothèse d'homogénéité des variances).

Dans le cas d'essais comparant plus de deux traitements, il va falloir également prendre en compte la covariance entre les différents bras comparés au traitement de référence. Dans le cas de l'hypothèse d'homogénéité des variances et d'essais à trois bras, cette covariance est de $\sigma/2$.

3.3 Mesure de la quantité d'effet

3.3.1 Meta-analyse par paire

La réalisation de la méta-analyse classique permet d'estimer les rapports de cotes poolés. Une cote est définie comme le rapport entre le risque de perdre et la chance de gagner. Dans le domaine de la PR, ce sera par exemple le rapport entre la probabilité de répondre à l'ACR20 et son complément le risque de ne pas répondre. Le rapport des cotes (OR) est égal au quotient de la cote mesurée dans le groupe testé et de celle relevée dans le groupe contrôle. Le rapport des cotes est un nombre sans dimension qui permet de comparer la probabilité d'un résultat entre plusieurs groupes. Si l'on compare la fréquence de survenue d'une réponse dans le groupe testé par rapport à un groupe témoin par exemple et que la valeur du rapport de cote est égale à l'unité, cela signifie que la survenue d'une réponse est équiprobable pour les deux groupes. Si le nombre est supérieur à 1 les chances d'être répondeurs dans le groupe du médicament testé sont plus grandes qu'elles ne l'étaient dans le groupe traitement de référence. Si le nombre est inférieur à 1 c'est l'inverse. Ces mesures de quantité d'effet sont ponctuelles. Le calcul de l'intervalle de confiance permet de préciser la marge d'erreur inhérente à leur estimation. Traditionnellement c'est l'intervalle de confiance à 95% qui est utilisé. Celui-ci fournit une fourchette dans laquelle, 95 fois sur 100, le résultat se trouvera inclus si l'expérience est répétée un grand nombre de fois. En pratique la précision d'une

mesure est appréciée par l'étroitesse de son intervalle de confiance. Lorsque l'intervalle chevauche les valeurs 0 pour le bénéfice absolu ou 1 pour le risque relatif ou le rapport de cotes, il n'existe pas de différence significative entre les groupes étudiés.

Les odds ratios s'interprètent donc en fonction de la position de leurs intervalles de confiance par rapport à 1. On applique les règles d'interprétation suivantes :

Tableau 2 : Règle d'interprétation d'un odds ratio en fonction de son intervalle de confiance

Intervalle de confiance	Interprétation
Les bornes supérieures et inférieures de l'intervalle de confiance de l'odds ratio sont supérieures à 1.	Le produit comparé est significativement supérieur au produit de référence.
La borne supérieure de l'intervalle de confiance de l'odds ratio est supérieure à 1, la borne inférieure est inférieure à 1.	Aucune différence significative n'est détectée, le produit comparé est équivalent au produit de référence.
Les bornes supérieures et inférieures de l'intervalle de confiance sont supérieures à 1.	Le produit testé est significativement supérieur au produit de référence.

Pour une interprétation plus rapide des odds ratio, on utilise une « représentation graphique en grille » communément désignée sous le nom de « forest plot ». Les intervalles de confiance des odds ratio y sont représentés par des segments de droite. Lorsqu'une de ces droites coupe l'axe 1, alors les rapports de cotes ne sont significativement différents ; dans le cas contraire ils le sont.

3.3.2 Méta-analyse multi-traitement

La méta-analyse multitraitement permet d'estimer l'efficacité brute (le taux de réponse ACR) du traitement, l'efficacité relative du traitement par rapport à ses comparateurs (sous forme d'odds ratio et de log odds ratio), et la probabilité d'être l'un des trois meilleurs traitements. L'efficacité brute d'un traitement correspond au pourcentage de patients qui vont atteindre le critère de jugement, l'ACR 50 ou subsidiairement l'ACR 20. L'efficacité relative est exprimée sous forme d'odds ratio ou de log odds ratio. L'interprétation des représentations graphiques de l'odds ratio est similaire à celle décrite précédemment pour la méta-analyse classique (position par rapport à 1). Le log odds ratio correspond à un odds ratio exprimé sur une échelle logarithmique, son interprétation est donc la même que celle de l'odds ratio, mais son seuil de non-efficacité est matérialisé par rapport à 0.

L'analyse bayésienne fournit des résultats plus directs que ceux de l'analyse classique. L'exemple le plus flagrant est la définition de l'intervalle de confiance d'un paramètre d'un modèle. Pour les bayésiens qui préfèrent parler d'intervalle de crédibilité (IcR), il s'agit de l'intervalle qui contient le paramètre avec une crédibilité donnée. Dans l'inférence fréquentiste, une telle interprétation ne pouvait être possible puisque la valeur du paramètre inconnu était supposée être une grandeur fixe. L'interprétation correcte de l'intervalle de confiance par rapport à l'intervalle de crédibilité est donc totalement différente. Il faut imaginer que sur l'ensemble des échantillons qui pouvaient être obtenus à partir d'un modèle de population, 95% des intervalles de confiance calculés contiennent la vraie valeur du paramètre à estimer.

3.4 Méthodes de simulation MCMC

3.4.1 Difficulté de calcul des intégrales et ré-échantillonnage

Le calcul analytique de l'intégrale, souvent multidimensionnelle, du dénominateur de la formule de Bayes est infaisable dans la plupart des cas. La découverte et la possibilité de mettre en œuvre des algorithmes de simulation capables d'obtenir des tirages aléatoires de la loi *a posteriori* des paramètres a permis de résoudre le problème. L'expression mathématique de la loi reste inconnue, mais avec le tirage d'un nombre significatif d'échantillons de cette loi on peut calculer empiriquement, la moyenne, la variance et toutes les grandeurs statistiques qui la décrivent. La procédure de simulation qui sera utilisée est la méthode dite de Monte Carlo par Chaîne de Markov (méthode MCMC) utilise l'échantillonneur de *Gibbs*. Cette méthode permet d'obtenir les tirages d'une loi jointe de plusieurs paramètres à partir des lois de chacune de ses composantes conditionnellement à celles de toutes les autres. À chaque itération i , il y a k tirages successifs des lois conditionnelles et la procédure est répétée N fois pour obtenir une distribution stationnaire. Autrement dit la simulation d'une variable aléatoire de dimension N est réduite à la simulation de N variables unidimensionnelles. Une fois vérifié que les valeurs obtenues sont indépendantes des valeurs initiales choisies (diagnostic de convergence de la chaîne de Markov construite), alors chaque paramètre de la loi multidimensionnelle peut être considéré comme extrait de la loi *a posteriori* conjointe des paramètres approximée par simulation.

Sur le plan technique, nos simulations MCMC mises en œuvre dans les méta-analyses multitraitements seront réalisées sous le logiciel bayésien WinBUGS v1.4.3.

3.4.2 Contrôle de la convergence

Le problème pratique est de déterminer à partir de combien d'itérations on peut considérer que la « convergence est atteinte ». La méthode du diagnostic multi-chaînes est aujourd'hui la plus répandue. Son esprit est très simple : pour chaque itération on vérifie que les dernières valeurs obtenues de chaque chaîne constituent des échantillons de la même loi.

Pour chaque méta-analyse multitraitements, seront réalisées les simulations de Monte-Carlo sur 3 chaînes de Markov. Dans un premier temps, 40 000 simulations seront réalisées au titre du « burn-in » (période de chauffe destinée à entraîner la convergence des paramètres vers leurs valeurs *a posteriori*). 40 000 nouvelles simulations permettront ensuite de calculer les distributions *a posteriori* des paramètres.

Une fois ces simulations effectuées, un diagnostic de convergence (test de Gelman-Rubin) et une estimation des autocorrélations seront effectués sur les paramètres de bases. La méthode de Gelman Rubin se base sur la comparaison entre la variance inter-chaînes (entre les valeurs de chaînes différentes) et la variance intra-chaînes (entre les valeurs d'une même chaîne). Dans un premier temps, les chaînes subissent l'influence des points initiaux très dispersés et la variance inter-chaînes est supérieure à la variance intra-chaîne. À partir du moment où le rapport de ces deux grandeurs tend vers l'unité, on peut considérer que la convergence est établie. Si tel n'était pas le cas, 40 000 simulations supplémentaires seront réalisées ; en présence d'autocorrélations, les distributions *a posteriori* des paramètres seront réévaluées en ne tenant compte que d'une simulation sur trois.

3.5 Indices thérapeutiques estimés

Dans un essai individuel, l'analyse de l'effet d'un traitement peut se faire par régression logistique : on pose alors $\text{logit}(P) = \alpha + \beta * \text{Traitement}$ où Traitement vaut 0 si l'on a affaire au groupe contrôle, et 1 si l'on a affaire au traitement évalué. β , le coefficient associé à la variable traitement, mesure son effet. Il s'agit d'un log-OR.

Dans une MTC, l'effet du traitement est estimé également par un log-OR. Les log-OR des traitements comparés par rapport à un traitement choisi comme référence sont donc les paramètres de base des MTC.

À l'issue d'une MTC bayésienne, nous disposerons d'une distribution a posteriori de chaque log-OR, ce qui nous permettra d'obtenir :

- Une estimation de la probabilité a posteriori d'obtenir une réponse positive au traitement sur le critère de jugement choisi ACR20 ou ACR50 à 24s \pm 10 appelées ci-après Taux Absolu de Réponse (TAR) ou efficacité absolue (quel que soit l'appellation il s'agit toujours du SMR) que l'on désignera dans la programmation WindBUGS par la lettre (T).

La probabilité a posteriori de l'efficacité absolue d'un traitement (T) est calculée sans briser les randomisations en additionnant :

- o l'effet placebo, mesuré en termes de cotes sur l'ensemble des bras contrôle de tous les essais pour estimer le risque moyen de son occurrence,
- o et la plus-value thérapeutique spécifique du traitement évalué, mesurée par le logarithme de l'odds ratio du critère de jugement dans le bras évalué par rapport au bras de référence.

Les estimations du taux absolu de réponse (TAR) d'un traitement intègrent l'incertitude relative à la probabilité d'efficacité absolue du comparateur, et elles ne peuvent de ce fait être directement utilisées, à travers leurs intervalles de crédibilité, pour juger de la supériorité ou de l'équivalence entre deux traitements. Il sera donc préférable de vérifier si l'intervalle de crédibilité du log-OR comparant deux traitements contient la valeur 0 ou si l'intervalle de crédibilité de l'OR correspondant contient la valeur 1.

- Une estimation de la valeur ajoutée thérapeutique relative (VAR) du traitement A par rapport au Traitement B mesurée sur l'ACR20 ou ACR50 à 24s \pm 10 et exprimée sous la forme d'un indice thérapeutique : le logarithme de rapport de cotes moyen ou log odds ratio de A par rapport à B noté $\text{Log-OR}_{[BA]}$, dont la valeur est égale à l'espérance de l'exponentielle de la distribution a posteriori des log-OR, par rapport au traitement choisi comme comparateur. Pour repasser de l'échelle logarithmique à l'échelle métrique il suffit de prendre l'exponentielle du log-OR ce qui permet de présenter les résultats de façon plus classique en rapport de cotes moyen ou Odds ratio de A par rapport à B (noté $\text{OR}_{[BA]}$). Lors du passage de l'espérance du log-OR, $E[\text{log-OR}]$, à celle de l'OR, il faut noter que si la distribution a posteriori du log-OR est très asymétrique, on n'a pas forcément $E[\exp(\text{log-OR})]$ égale à $\exp(E[\text{log-OR}])$. En cas de conflit, il faut raisonner préférentiellement sur le log-OR, qui est le paramètre de base estimé par le modèle.
- Une estimation de la valeur ajoutée thérapeutique différentielle (VAD) du traitement A par rapport au traitement pivot choisi comme traitement de référence mesurée sur l'ACR20 ou ACR50 à 24s \pm 10 et exprimée sous la forme d'une différence de log de rapports de cotes par rapport à un comparateur de référence commun (dans le cas d'espèce les DMARD conventionnels)

- Une estimation de la probabilité a posteriori pour qu'un traitement donné soit supérieur à l'ensemble des autres traitements au regard du critère de jugement adopté : taux de réponse sur l'ACR20 ou ACR50 à 24s ± 10 , qui n'est rien d'autre que le nombre de fois où le traitement considéré est arrivé en tête du classement au cours des différentes simulations MCMC, du fait d'une probabilité a posteriori d'efficacité absolue supérieure à celle de ses comparateurs (un indicateur ordinal de rang et non de quantité d'effet),
- Une estimation de la probabilité a posteriori pour qu'un traitement donné soit supérieur en termes d'efficacité absolue à un autre traitement spécifique (et non plus à l'ensemble des traitements), calculée comme la fréquence, au cours des simulations, avec laquelle la probabilité d'efficacité du premier traitement est supérieure à celle du traitement auquel il est comparé.
- Une estimation de la probabilité a posteriori pour qu'un traitement particulier figure parmi les trois meilleurs traitements, calculée comme la fréquence, au cours des simulations, avec laquelle la probabilité d'efficacité de ce traitement se trouve placée soit en première, soit en deuxième, soit troisième position en termes de rang.

Bien que le log-OR soit le paramètre central d'une MTC, et celui à partir duquel découlent toutes les autres estimations, nous présenterons préférentiellement les probabilités absolues d'efficacité (puisque c'est ainsi que sont présentés les résultats des essais cliniques) et les OR (grandeur plus classique que leur logarithme).

3.6 Analyse de non-infériorité

3.6.1 Définitions

Nous pouvons regrouper les essais cliniques en trois catégories (Piaggio et al., 2006):

- Les essais de supériorité, les plus classiques, ont pour objectif de démontrer qu'une intervention est plus efficace qu'un comparateur ;
- Les essais d'équivalence cherchent eux à démontrer qu'une intervention soit thérapeutiquement similaire à une autre ;
- Les essais de non-infériorité cherchent enfin à démontrer qu'une intervention n'est pas plus mauvaise qu'un comparateur.

Il est impossible de montrer que deux interventions ont exactement la même efficacité. En pratique, les essais d'équivalence vont chercher à montrer que les interventions ne diffèrent pas au-delà d'une certaine quantité B . On définit ainsi une marge d'équivalence $[-B ; +B]$ à l'intérieur de laquelle est censé se situer le différentiel d'efficacité entre deux interventions, si l'on veut démontrer l'équivalence. Dans le cadre d'un essai de non-infériorité, par contre, il s'agit de montrer qu'une intervention n'est pas plus mauvaise qu'une autre de plus que la marge d'équivalence. À l'inverse, tout effet différentiel en faveur de l'intervention évaluée ne contredit pas la non-infériorité (Snapinn, 2000).

Bien entendu, un essai de non-infériorité (ou d'équivalence) n'est justifié que dans la mesure où la supériorité du traitement comparateur par rapport à placebo ou au traitement de référence a été démontrée au préalable. Même dans ce cadre, la recherche de non-infériorité d'une nouvelle intervention par rapport à une ancienne ayant fait la preuve de son efficacité ne se justifie que si, pour une efficacité comparable, la nouvelle intervention apporte des éléments d'amélioration : moindre coût, administration plus aisée ou moins invasive, plus grande tolérance...

3.6.2 Cadre classique

4.6.2.1 Règle de la non-infériorité

Dans le cadre classique, on raisonnera le plus souvent à partir de l'intervalle de confiance (généralement à 95 %) du différentiel d'efficacité. La figure 1 illustre plusieurs cas de figure et le tableau 1 donne leur interprétation.

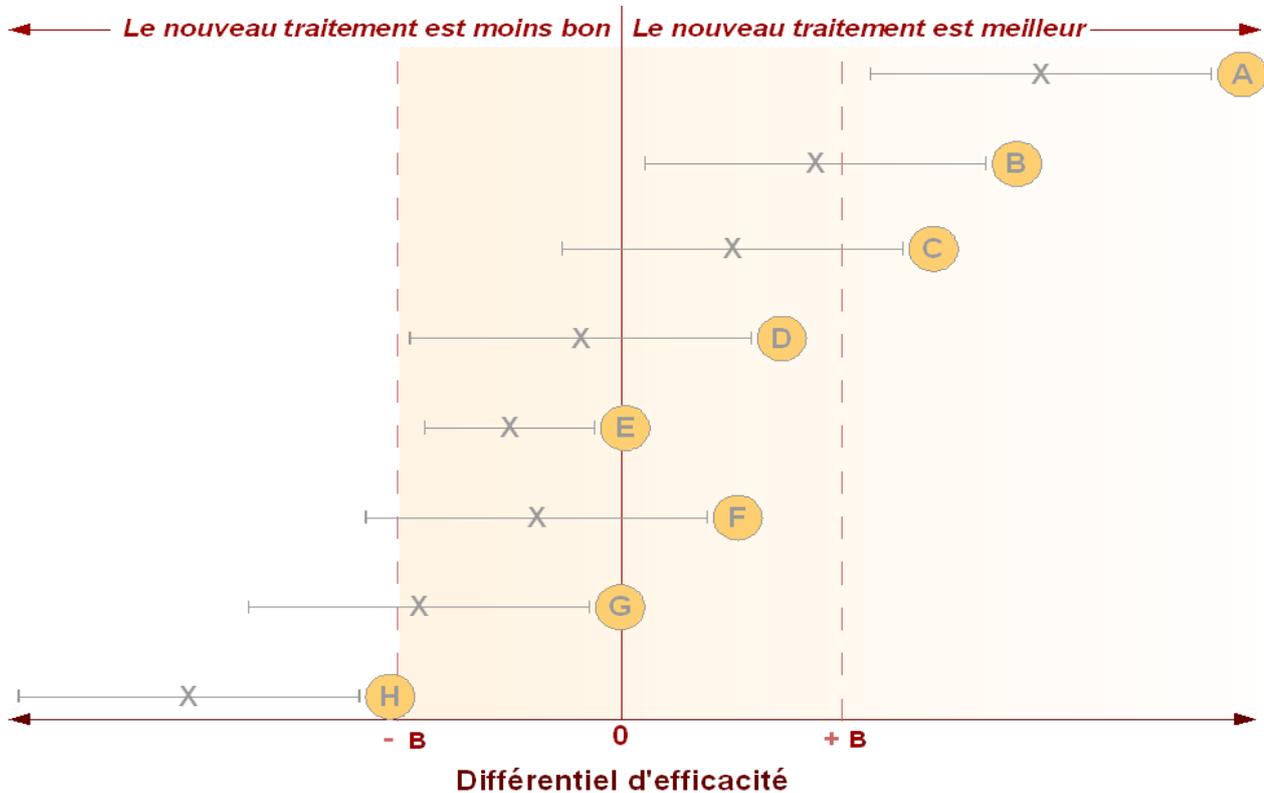


Figure 4 : Équivalence et non-infériorité

Dans le cas d'un essai d'équivalence, la totalité de la marge d'équivalence doit être prise en compte pour les interprétations. Dans celui d'un essai de non-infériorité, seule compte la borne inférieure $-B$, à partir de laquelle on considère qu'il y a infériorité. De ce fait, le risque d'erreur de première espèce est de 2,5 %.

Dans le cas A, l'intervalle de confiance ne contient ni la valeur 0 (parfaite équivalence) ni la marge de supériorité. On conclut dans les deux cas à une supériorité du traitement évalué.

Dans le cas B, l'intervalle de confiance ne contient pas la valeur 0, mais contient la marge de supériorité. On ne peut pas conclure dans le cas d'un essai d'équivalence, mais on conclut toujours à la supériorité du traitement évalué dans le cas d'une analyse de non-infériorité.

Dans le cas C, l'intervalle de confiance contient à la fois la valeur 0 et la marge de supériorité. On ne peut toujours pas conclure dans le cas d'un essai d'équivalence, on ne peut plus conclure à la supériorité dans le cas d'un essai de non-infériorité, mais on a par contre non-infériorité du traitement évalué.

Dans le cas D, l'intervalle de confiance contient la valeur 0, mais se situe dans la marge d'équivalence. On conclut à l'équivalence dans le cas d'un essai d'équivalence, et à la non-infériorité dans le cas d'un essai de non-infériorité.

Dans le cas E, l'intervalle de confiance ne contient plus la valeur 0, mais se situe dans la marge d'équivalence. On conclut à l'équivalence dans le cas d'un essai d'équivalence, et à la non-infériorité dans le cas d'un essai de non-infériorité.

le cas d'un essai de non-infériorité. Dans le cas d'un essai classique, on aurait toutefois conclu que le traitement évalué est moins bon que le traitement comparateur (différentiel d'efficacité significativement inférieur à 0).

Dans les cas F à H, l'interprétation dépend du fait que l'intervalle de confiance recoupe ou non la borne d'infériorité, et les conclusions sont donc les mêmes que l'on soit en essai d'équivalence ou en essai de non-infériorité. Ces conclusions sont toutefois différentes de celles d'un essai de supériorité dans le cas G.

Tableau 3 : Interprétation de la position de l'intervalle de confiance

Cas	Équivalence	Non-Infériorité
A	Nouveau traitement supérieur	Nouveau traitement supérieur
B	On ne peut conclure	
C		Équivalence
D		
E		
F	On ne peut conclure	
G		
H	Nouveau traitement inférieur	

Il est acceptable, dans le cas d'un essai de non-infériorité où la non-infériorité a été démontrée, de poursuivre par un test de supériorité. Il n'est pas acceptable, dans le cas d'un essai de supériorité ou la supériorité n'a pas été démontrée, de poursuivre par un test de non-infériorité : les marges d'équivalence se définissent a priori.

4.6.2.2. Choix de la marge d'équivalence

Il n'existe pas de méthode unique pour déterminer la valeur de B. Cependant, il est bien clair que celle-ci n'est jamais choisie arbitrairement. De nombreux facteurs peuvent être pris en considération (Wiens, 2002):

- Définition d'une efficacité « cliniquement significative ». Celle-ci est le plus souvent très subjective, et représente le différentiel d'efficacité minimum tel qu'une différence dans le statut du patient soit cliniquement perceptible. Une telle approche n'est justifiée que dans le cas où l'efficacité du comparateur est suffisamment importante : si le différentiel d'efficacité du comparateur par rapport à placebo est de l'ordre de 5, et que le différentiel d'efficacité considéré comme cliniquement significatif est de l'ordre de 10, l'analyse perd son sens : un traitement nocif par rapport à placebo pourrait être considéré comme non-inférieur à un traitement bénéfique par rapport à placebo.
- Définition de B par rapport au différentiel d'efficacité observé pour le traitement comparateur évalué contre placebo. Il faudra toutefois procéder à des ajustements pour prendre en compte la variabilité inter-essais. Certaines méthodes plus calculatoires s'inspirent également de cette idée (Rothmann and Tsou, 2003).

L'Agence Européenne du médicament (EMA) a émis des recommandations sur le choix de la marge de non-infériorité (EMA, 2005). Elle demande que cette marge soit justifiée à la fois sur le plan statistique et sur le plan clinique, qu'elle permette de montrer que l'intervention a une efficacité non nulle. Elle ne recommande pas d'adopter cette marge sur la base d'une proportion de l'efficacité différentielle entre le

comparateur et placebo, d'une taille d'effet (efficacité différentielle divisée par l'écart type), sur la base d'une puissance ou d'une taille d'échantillon.

3.6.3 Cadre bayésien

4.6.2.3. Distribution a posteriori et raisonnement d'équivalence

L'analyse bayésienne permettra de disposer directement de la distribution a posteriori de l'efficacité différentielle. Il est alors possible de construire la probabilité de non-infériorité. Le cadre bayésien, en outre, s'affranchit des limites du raisonnement fréquentiste puisque la symétrie au niveau des bornes d'infériorité et de supériorité n'est plus imposée.

La figure suivante présente le raisonnement d'équivalence dans un cadre bayésien. La courbe grise correspond à la distribution a posteriori de l'efficacité différentielle (OR ou log OR). Les deux lignes pointillées correspondent aux bornes inférieures et supérieures de la marge d'équivalence (respectivement B_i et B_s). L'aire sous la courbe permet de donner les probabilités d'équivalence. L'aire située à droite de la borne supérieure donne la probabilité de supériorité ; l'aire située entre les bornes supérieures et inférieures donne la probabilité d'équivalence ; enfin, l'aire située à gauche de la borne inférieure donne la probabilité d'infériorité. La probabilité de non-infériorité est ainsi égale à 1 – la probabilité d'infériorité.

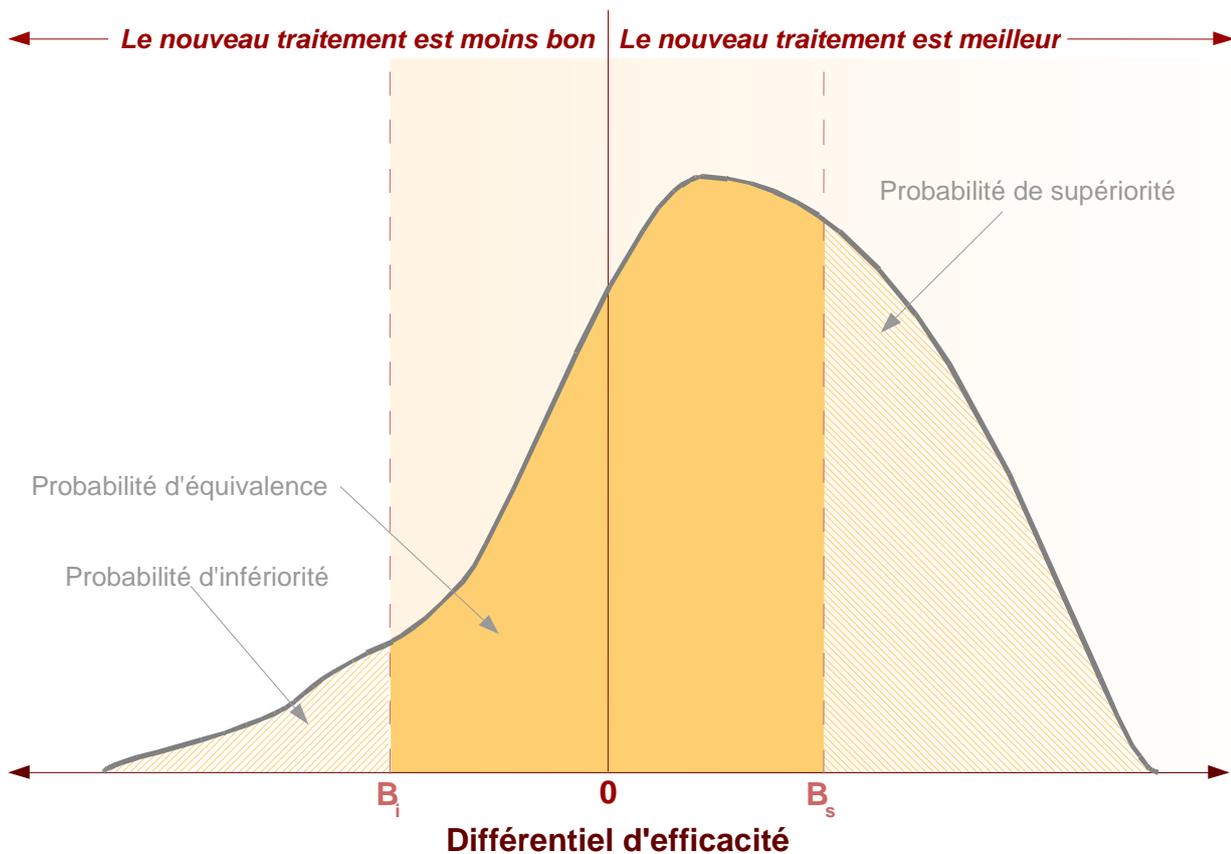


Figure 5 : Non-infériorité dans le cadre bayésien

La marge d'équivalence bayésienne constitue le point clé de l'analyse de non-infériorité. La méta-analyse multitraitements raisonne en termes de log odds-ratios, cependant il est plus confortable pour un spécialiste de raisonner en termes de différentiel d'efficacité absolue (*différentiel de taux absolu de réponse* sur

l'ACR 20 ou l'ACR 50). La marge d'équivalence sera définie selon ce principe et sera fixée dans un premier temps à 15%.

La façon la plus simple de gérer le passage de la probabilité d'efficacité, et du différentiel absolu d'efficacité, au log-OR est de considérer la probabilité d'efficacité du comparateur comme fixe.

Soit T_{MT} , la moyenne des efficacités absolues des autres anti-TNF. Appelons DEA, le différentiel d'efficacité absolue qui peut être considéré comme cliniquement significatif. À partir de ces deux éléments, la Borne Inférieure de la Marge d'Équivalence (BIME) peut être calculée :

$$BIME = T_{MT} - DEA$$

Puisque la méta-analyse multitraitement sous WinBUGS sera conduite en log odds-ratio, la borne inférieure de la marge d'équivalence (BIME) par rapport à la moyenne de l'efficacité des traitements mesurée en ACR20 devra être exprimée dans les mêmes unités c.-à-d. en log de rapport de cotes :

$$\text{Log}(OR_{[MT,BIME]}) = \log\left(\frac{BIME}{1 - BIME}\right) - \log\left(\frac{T_{MT}}{1 - T_{MT}}\right)$$

4.6.2.4. Ratio de Bayes et décision statistique

La statistique fréquentiste formule pour tout test un couple d'hypothèses : une hypothèse nulle, spécifiant la nullité d'un paramètre, et une hypothèse alternative. Cette formulation n'est pas naturelle dans le raisonnement bayésien, mais il est néanmoins possible de la transposer en terme bayésien. Dans ce cadre, l'alternative aux tests d'hypothèses classiques a été introduite par Jeffrey en 1961 et repose sur 'utilisation du facteur de Bayes. Kass et Raftery (Kass and Raftery, 1995) définissent celui-ci comme la « cote a posteriori d'une hypothèse, lorsque les probabilités a priori des deux hypothèses comparées sont identiques, c'est à dire :

$$\frac{p(H_0/Données)}{p(H_1/Données)} = \left(\frac{p(Données/H_0)}{p(Données/H_1)}\right) \times 1$$

Odds a posteriori Facteur de Bayes Odds a priori

On teste l'hypothèse H_0 : « Le traitement X est non-inférieur à son comparateur », contre l'hypothèse H_1 : « X est inférieur à son comparateur ».

Le rapport $p(\text{données} / H_0) / p(\text{données} / H_1)$ nous renseigne sur les chances de gagner le pari « l'hypothèse H_0 de non-infériorité du traitement X par rapport à la borne inférieure de la marge d'équivalence de son comparateur est plausible » par rapport à la plausibilité du pari contraire H_1 . Il varie entre 0 (on juge comme certain que l'hypothèse H_1 de l'infériorité est vraie) et l'infini (on juge comme certain que l'hypothèse H_0 de non-infériorité est vraie).

En 2004, Spiegelhalter (Spiegelhalter DJ, et al., 2004) a proposé d'interpréter le facteur de Bayes en fonction de la force de la preuve sur la base de l'échelle de cotation suivante :

Tableau 4 : Force de la preuve en fonction de la valeur de la cote d'une des hypothèses par rapport à l'autre

Cote de l'hypothèse H0 (X non inférieur est vraie) par rapport à son complémentaire H1	Force de la preuve
> 100	Preuves indiscutables
32 à 100	Très fortes preuves
10 à 32	Fortes preuves
3,2 à 10	Preuves substantielles
1 à 3,2	Preuves faibles

Cote de l'hypothèse H1 (X non inférieur est faux) par rapport à son complémentaire H0	Force de la preuve
de 1 à 1/3,2	Preuves faibles
de 1/3,2 à 1/10	Preuves substantielles
de 1/10 à 1/32	Fortes preuves
1/32 à 1/100	Très fortes preuves
<1/100	Preuves indiscutables

Si ce rapport, ou si cette cote, est égale est égale à un réel x alors la probabilité (p) de gagner le pari, est égale $x/x+1$ Les correspondances entre les différents seuils exprimés en cotes et les probabilités seuils correspondantes sont alors les suivantes :

Tableau 5 : Transposition de l'échelle de Spiegelhalter en termes de probabilité a posteriori

Cote de l'hypothèse H0 (GOL non-inférieur est vraie) par rapport à son complémentaire H1	Probabilité a posteriori	Force de la preuve en faveur de H0
> 100	$P > 99\%$	Preuves indiscutables
32 à 100	$97\% < P < 99\%$	Très fortes preuves
10 à 32	$91\% < P < 97\%$	Fortes preuves
3,2 à 10	$76\% < P < 91\%$	Preuves substantielles
1 à 3,2	$50\% < P < 76\%$	Preuves faibles

Cote de l'hypothèse H1 (GOL non inférieur est faux) par rapport à son complémentaire H0	Probabilité a posteriori	Force de la preuve en défaveur de H0 et en faveur de H1
--	-----------------------------	--

de 1 à 1/3,2	24% < P < 50%	Preuves faibles
de 1/3,2 à 1/10	9% < P < 24%	Preuves substantielles
de 1/10 à 1/32	3% < P < 9%	Fortes preuves
1/32 à 1/100	1% < P < 3%	Très fortes preuves
<1/100	P < 1%	Preuves indiscutables

En présence de distributions à priori « vagues », ce seront donc les fréquences de la simulation qui détermineront les chances d’être inférieures ou non inférieures à la borne basse de la marge d’équivalence définie sur le taux de réponse de ses concurrents

3.7 Un exemple de MTC dans la polyarthrite rhumatoïde

L’équipe de REES France a déjà mis en place la méthode de mélange de comparaison de traitement dans la polyarthrite rhumatoïde qui a déjà été publiée dans *join bone spine* en 2008. Le but alors était d’évaluer évaluer non pas les toxicités relatives, mais l’efficacité relative des traitements sur le critère ACR 20.

Tableau 6 : Taux de répondeurs au critère ACR20 dans les 16 essais

Traitement	MTX+PLC	MTX+RTX ₁₀₀₀	MTX	MTX+ETA ₂₅	ETA ₂₅	MTX+ADA ₄₀	PLC	ADA ₄₀	MTX+INF _{3/8}
Cohen 2005	18 %	51 %							
Emery 2006	28 %	54 %							
Edwards 2004		73 %	38 %						
Weinblatt 1999	27 %			71 %					
Bathon 2000			65 %		72 %				
Moreland 1999					59 %		11 %		
Klareskog 2004			75 %	85 %	76 %				
Kristensen (TC1)					63 %			62 %	
Kristensen (TC2)					59 %			52 %	
Keystone 2004	29,5 %					63 %			
van de Putte 2003							10 %	57 %	
Weinblatt 2003	14,5 %					67,2 %			
van de Putte 2004							19,1 %	46 %	
Maini 1999	20 %								50 %
St Clair 2004	53,6 %								62,4 %
Geborek 2002				60 %					43,3 %

Le modèle de MTC donne les taux de répondeurs au critère ACR20 présentés dans le tableau qui suit, de même que les rapports de cotes (notés OR) entre les différents traitements. La médiane est utilisée comme indicateur de position, car elle est plus stable que la moyenne. Les percentiles à 2,5 % et 97,5 % forment un intervalle de confiance à 95 % (méthode des percentiles).

Tableau 7 : Résultats du modèle MTC en termes de pourcentage de répondeurs ACR20 et de rapports de cotes

Traitement	Médiane	Percentile 2,5 %	Percentile 97,5 %
MTX+PLC	26,8 %	23,2 %	30,5 %
MTX+RTX ₁₀₀₀	62,5 %	42,9 %	78,3 %
MTX+ETA ₂₅	61,1 %	39,2 %	79,6 %
ETA ₂₅	44,0 %	18,5 %	72,3 %
MTX+ADA ₄₀	69,8 %	49,3 %	85,0 %
ADA ₄₀	38,1 %	12,8 %	70,5 %
MTX+INF _{3/8}	44,2 %	28,1 %	62,7 %
Monothérapies			
OR _{ETA/ADA}	0,78	0,37	1,61
Associations MTX			
OR _{RTX/ETA}	0,94	0,33	2,84
OR _{RTX/ADA}	1,39	0,44	4,79
OR _{RTX/INF}	0,48	0,17	1,39
OR _{ETA/ADA}	1,48	0,41	5,42
OR _{ETA/INF}	0,51	0,20	1,30
OR _{ADA/INF}	0,34	0,11	1,10

4 Mapping de l'échelle psychométrique HAQ vers le questionnaire utilité préférence EQ-5D

Nous avons besoin d'une commune mesure pour comparer les effets des choix les plus en amont du système de santé ; ceux qui portent sur des techniques et des équipements utilisables par des populations. Le critère de jugement principal retenu dans les études cliniques ne permet pas de répondre à cette problématique. L'estimation du coût par taux de réponse ACR n'éclaire en rien le débat démocratique puisqu'il n'est rapportable à aucun étalon de mesure à l'aune duquel il pourrait juger. Pour disposer d'un critère au regard duquel l'impact du traitement ou les mérites respectifs de deux traitements alternatifs peuvent être comparés, **le résultat obtenu évalué au regard du bénéfice qu'en retire le malade dans sa vie de tous les jours**, ce qui est en réalité le critère ultime en fonction duquel il fait ses choix. C'est le principe des études de qualité de vie. On distingue, en ce domaine, deux types d'instruments : les échelles psychométriques et l'approche utilité préférence. Nous proposons de les mettre en œuvre conjointement en intégrant à la fois le HAQ et EQ-5D dans le Ccadre d'une étude sur coupe transversale pour dégager les corrélations existantes entre les scores de l'échelle psychométrique HAQ et les coefficients de qualité de vie de l'EQ-5D-3L

Le travail servira à intégrer le ressenti des patients français par rapport aux conséquences fonctionnelles des traitements de la polyarthrite rhumatoïde dont ils ont bénéficié. Ce ressenti doit être appréhendé au travers du prisme que constituent les préférences individuelles des patients. Les questionnaires

habituellement utilisés dans les essais cliniques ne permettent pas en général d'évaluer directement l'utilité que le patient accorde à son état de santé. Les données recueillies dans ce contexte, sous forme d'échelles psychométriques comme le HAQ, peuvent être alors mobilisées pour obtenir un indicateur synthétique qui permet d'estimer l'utilité qui s'y rattache sous la forme d'un score synthétique global qui est fondamental pour l'analyse économique.

L'EQ-5D comporte 5 dimensions intégrant 3 niveaux de sévérité (pas de problème, problème modéré, incapacité). Les 5 dimensions sont les suivantes : Mobilité, Soins personnels, Activités usuelles regroupant le travail, les études, les travaux ménagers et les activités sociales, Douleur/Inconfort et Anxiété/Dépression. Les états de santé disponibles provenant de la combinaison des différents niveaux des différentes dimensions sont au nombre de 243 (3^E 5^E). Si on rajoute les états "Mort" et "Inconscient" qui ne font pas partie des combinaisons disponibles, mais dont la description complète les états de santé indispensables pour décrire correctement l'état de santé d'une population, le nombre total d'états de santé décrits par le système EuroQol est de 245. Pour la facilité de dénomination, les états de santé sont notés sous la forme d'un numéro à 5 chiffres : par exemple l'état 12223 est le suivant :

1. Je n'ai aucun problème pour me déplacer à pied
2. J'ai des problèmes pour me laver ou m'habiller tout(e) seul(e)
2. J'ai des problèmes pour accomplir mes activités courantes
2. J'ai des douleurs ou une gêne modérée(s)
3. Je suis extrêmement anxieux (se), ou déprimé(e)

4.1 Principe du « mapping »

L'estimation d'un score global d'utilité peut être obtenue en mettant en œuvre les techniques dites de « mapping » qui consistent à développer et à utiliser de modèles ou des algorithmes pour prédire la valeur de l'utilité liée à la santé à partir des résultats disponibles en termes d'ACR20 puis d'HAQ et finalement d'EQ-5D (Longworth and Rowen, 2011). Il s'agit donc de trouver une fonction qui relie les autres données des patients, en provenance d'un questionnaire ou d'une mesure « source » avec l'utilité fondée sur l'approche des préférences. De sorte que : $mesure\ cible / utilité = f(mesure\ source)$

Pour réaliser le « mapping », il convient d'identifier les éléments suivants :

Les valeurs cibles : il s'agit de l'utilité qu'on cherche à prédire dont la valorisation s'appuie usuellement sur un instrument de mesure basé sur l'approche utilité préférence, le plus fréquemment l'EQ-5D, et en particulier la version EQ-5D-3L. L'extrapolation peut se faire aussi bien sur score synthétique de l'instrument que sur les modalités de ses différentes dimensions. Un score de qualité de vie globale obtenue pour la France dans le cadre d'une étude faite en population générale par Julie Chevalier (Chevalier, 2010).

Les données sources : il s'agit des données recueillies sur des échelles psychométriques (HAQ).

L'application des données sources vers la valeur cible : il s'agit de spécifier un modèle permettant d'expliquer l'utilité (cible) par les échelles psychométriques de l'instrument source susceptibles de l'influencer et éventuellement des caractéristiques des patients. Cette spécification doit être justifiée et sa robustesse vérifiée.

La conception de la fonction de « mapping » implique l'estimation d'une relation statistique entre la mesure cible et la mesure source par des données. Il est impératif de disposer d'un échantillon de patients ayant été soumis à la fois aux questionnaires de l'instrument source et de l'instrument cible. Cet

échantillon est primordial puisqu'il doit avoir des caractéristiques aussi similaires que possible à ceux des individus formant les données sur lesquelles sera appliqué l'algorithme de « mapping ».

Le processus de mapping peut être décomposé en trois étapes successives :

1. La spécification des éléments qui vont intervenir ;
2. Le choix du type de modélisation à mettre en œuvre ;
3. La validation du modèle et des résultats obtenus.

4.2 Spécification du « mapping »

Au commencement d'un mapping, il est d'abord nécessaire de considérer les éléments que l'on veut modéliser et les éléments dont on dispose pour atteindre l'objectif. L'étape de spécification va permettre de caractériser les données EQ-5D disponibles, de déterminer si l'on va privilégier une approche directe (qui s'attache à donner une estimation du score global de l'EQ-5D) ou indirecte (qui cherche à prédire les modalités des 5 cinq dimensions, et convertir les réponses ainsi obtenues en un score global), quelles sont les mesures psychométriques disponibles à l'issue de l'étude clinique et/ou de la revue de la littérature associée, ainsi que d'éventuelles covariables cliniques ou démographiques que l'on souhaite faire intervenir dans l'analyse.

L'analyse des données EQ-5D dont on dispose, ainsi que le choix de l'approche en fonction de l'élément que l'on souhaite modéliser et estimer en sont les points principaux. Ils permettent d'éclairer le statisticien sur le type de modèle à utiliser par la suite.

4.2.1 « Mapping » direct sur le score global de l'EQ-5D-3L

L'approche directe du mapping, appliquée à l'EQ-5D, permet de prendre en compte la distribution multimodale du score global d'utilité observé dans le questionnaire EQ-5D-3L. l'objectif de cette approche est de trouver une fonction qui relie les données patients du questionnaire psychométrique HAQ directement avec le score global de l'EQ-5D-3L.

4.2.2 « Mapping » indirect sur les modalités des cinq dimensions de l'EQ-5D-3L

Une méthode alternative à considérer dans le cadre d'un mapping consiste à adopter une approche indirecte, ne se focalisant plus sur le score d'utilité, mais sur les réponses qui permettent de calculer ce score. Cette approche est parfois appelée « modèle de réponse ». Dans le cas de l'EQ-5D-3L, l'objectif sera de construire une fonction statistique permettant de prédire à partir de l'échelle psychométrique, pour chacune des 5 dimensions du questionnaire, laquelle des trois modalités de réponses sera choisie par le patient.

4.3 Modélisation

Dans la suite :

Y : La mesure cible, score ou réponses du questionnaire EQ-5D-3L

X : Le vecteur de l'échelle psychométrique source, les items de toutes les questions ou les scores standardisés des dimensions.

j : Le niveau 1, 2 ou 3 de la mesure cible.

i : Le patient

4.3.1 Mapping sur le score global de l'EQ-5D-3L

4.3.1.1 Régression linéaire par la méthode des moindres carrés ordinaires

L'une des méthodes les plus couramment utilisées dans la littérature (Young et al., 2015) consiste à réaliser une régression linéaire simple par la méthode des moindres carrés ordinaires (MCO) directement sur le score global d'utilité de l'EQ-5D-3L.

Le modèle de régression linéaire est le suivant :

$$Y_i = X_i\beta + \varepsilon_i$$

Ce modèle permet de déterminer une relation linéaire entre la variable dépendante et les variables indépendantes. Le terme ε_i représente le terme d'erreur. Le modèle linéaire repose sur différentes hypothèses :

- La distribution de l'erreur ε est indépendante de X (homogénéité).
- L'erreur est centrée et de variance constante (homoscédasticité) :

$$\forall i = 1, \dots, n \quad E(\varepsilon_i) = \mathbf{0} \quad \text{et} \quad V(\varepsilon_i) = \sigma^2$$

- Les erreurs ne sont pas autocorrélées : $\text{cov}(\varepsilon_i, \varepsilon_j) = \mathbf{0} \quad \forall i \neq j$
- L'erreur a une distribution normale : $\varepsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2)$.

L'estimation des $\hat{\beta}$ se fait par la méthode des moindres carrés ordinaires c'est-à-dire en minimisant la somme des carrés des écarts entre les observations et le modèle.

Une fois que les coefficients sont estimés, la valeur prédite du score global se calcule comme suit :

$$\hat{Y}_i = X_i\hat{\beta}$$

4.3.1.2 Modèle Tobit

Par rapport à la régression linéaire simple, le modèle Tobit présente l'avantage de prendre en compte la limite supérieure du score d'utilité de l'EQ-5D qui est bornée à 1. Elle tient également compte du fait qu'une importante proportion des observations est généralement égale à 1.

Le modèle définit une variable latente Y_i^* suivant une loi normale conditionnelle et tronquée à 1. Cela permettra en outre au modèle de modéliser une forte proportion des scores d'utilité à la valeur 1. Le modèle TOBIT se présentera de la façon suivante (Hernández Alava et al., 2012):

$$Y_i^* = \beta x_i + \varepsilon_i$$

$$Y_i = \begin{cases} Y_i^* & \text{si } Y_i^* < 1 \\ 1 & \text{si } Y_i^* \geq 1 \end{cases}$$

Avec Y_i le score d'utilité de l'individu i (valeur cible), X_i le vecteur des variables indépendantes permettant d'estimer le score d'utilité (données sources et éventuelles covariables), β un vecteur de coefficients inconnus et ε le terme d'erreur ($\varepsilon \sim N(0, \sigma^2)$).

Plusieurs variantes de ce modèle ont été développées, d'ordre 2 ou prenant en compte les interactions entre les différentes variables indépendantes (Rowen et al., 2009) (interaction entre les données sources).

4.3.1.3 Modèle CLAD (Censored Least Absolute Deviations)

Le modèle CLAD représente une alternative à l'estimation par le modèle Tobit. L'estimateur proposé est, contrairement aux méthodes d'estimation fondées sur l'hypothèse de normalité de la distribution du terme d'erreur (Tobit, MCO...), convergent et asymptotiquement normal, et est également robuste à l'hétéroscédasticité.

De par de la nature du score EQ-5D (limité dans l'intervalle $[-0.529, 1]$), on se trouve dans le cadre de la forme de la censure la plus simple est celle de la censure par intervalle. Le modèle CLAD tel qu'il est présenté par Powell (1984) vient remédier à cette problématique.

4.3.1.4 Modèle à deux parties

Le modèle à deux parties repose sur la combinaison de deux modèles, une régression logistique et une régression linéaire par méthode des moindres carrés ordinaires tronquée. La régression logistique permet dans un premier temps d'estimer la probabilité qu'un patient se situe dans un état de santé parfait, i.e. présente un score d'utilité de 1. Le modèle MCO tronqué est ensuite utilisé pour modéliser les patients ne se trouvant pas dans un état de santé parfait.

Les résultats de ces deux modèles sont ensuite combinés pour calculer l'espérance du score d'utilité global de l'EQ-5D :

$$E(EQ - 5D) = (p(\text{santé parfaite}) \times (\text{utilité} | \text{santé parfaite})) + (p(\text{santé imparfaite}) \times (\text{utilité} | \text{santé imparfaite}))$$

$$Y_i = (p(\textit{santé parfaite}) \times 1) + (p(\textit{santé imparfaite}) \times (X_i\beta + \varepsilon_i))$$

4.3.2 Mapping sur les cinq dimensions de l'EQ-5D-3L

4.3.2.1 Régression logistique multinomiale

Une approche du mapping se concentre sur les réponses données par les patients pour chacune des 5 dimensions du questionnaire EQ-5D-3L. Un modèle de régression logistique multinomial est ajusté sur chacune des dimensions et permet de classer les patients en fonction des trois niveaux de réponse disponibles. Les valeurs moyennes sont ensuite calculées à l'aide de simulation de Monte-Carlo.

La variable Y est la variable dépendante avec $K(3)$ modalités non ordonnées. L'objectif est d'estimer les probabilités $P(Y = j | x)$ qui est une combinaison linéaire de $x\beta_j$. De plus afin de calculer les différentes probabilités pour chaque modalité de la variable Y une modalité de référence est choisie (souvent la modalité $K(1)$) et chaque probabilité est exprimée en fonction de cette référence. Cependant, le modèle ne dépend pas de la modalité de référence choisie.

La probabilité associée au choix d'une modalité est la suivante :

$$P(Y = j | x_i) = \frac{\exp(x_i\beta_j)}{\sum_{k=1}^K \exp(x_i\beta_k)}$$

Le modèle n'est pas identifiable dans la mesure où trop de paramètres sont estimés par rapport au nombre d'équations. En effet, le modèle logistique multinomial est un modèle à multiples équations, $K - 1$ où chacune d'entre elles est une équation d'un modèle logistique binaire. Ainsi, une contrainte d'identification est indispensable tel que l'un des β_k soit égal à 0. Par exemple si $\beta_1 = \mathbf{0}$, les probabilités de la variable dépendante sont :

$$P(Y = 1 | x_i) = \frac{1}{1 + \sum_{k=2}^K \exp(x_i\beta_k)}$$

et pour $j > 1$,

$$P(Y = j | x_i) = \frac{\exp(x_i\beta_j)}{1 + \sum_{k=2}^K \exp(x_i\beta_k)}$$

Les valeurs prédites pour chaque dimension et chaque niveau de l'EQ-5D-3L font l'objet d'une simulation de Monte-Carlo. Cette procédure assure une valeur prédite non biaisée et la plus probable, mais permet également de s'adapter à la distribution de chaque dimension de l'EQ-5D-3L. 250 simulations de Monte-Carlo sont effectuées. De plus un nombre aléatoire est tiré afin de le comparer à la valeur prédite. Cette comparaison permet de déterminer le niveau de réponse. Une fois que le profil de chaque patient est prédit, le score global d'utilité est calculé.

4.4 Validation

Pour chacun des modèles précédents, une procédure d'acceptation-validation ou la validation croisée est effectuée. Le principe de l'acceptation-validation consiste à diviser l'échantillon en n parties, à estimer le modèle sur les $(n-1)/n$ parts de l'échantillon (échantillon d'apprentissage) puis à tester ce modèle sur la partie exclue (échantillon test). Pour une partie fixée, on effectue le modèle sur les $(n-1)$ parties restantes en conservant la somme des résidus (RMSE). La somme totale des $(n-1)$ RMSE est appelée CVPress et sert d'indicateur pour valider le modèle (Proskorovsky et al., 2014). Plus cet indicateur est faible plus le modèle est valide.

Cette procédure sera donc effectuée sur chaque spécification retenue de la sélection pas à pas pour tester à la fois la sensibilité et la validité du modèle en fonction des échantillons. Dans cette optique, une simulation de Monté Carlo peut être faite surtout pour des petits échantillons afin d'éviter des prédictions biaisées (Gray et al., 2006).

À partir de l'échantillon d'apprentissage, le modèle est développé et les coefficients de régression estimés. De plus, dans le but d'obtenir le meilleur modèle prédictif possible, la méthode de sélection des variables BACKWARD est appliquée lors de la construction du modèle. En partant du modèle complet, la méthode de sélection de variables BACKWARD consiste à supprimer au fur et à mesure la variable qui permet d'obtenir un modèle plus performant en utilisant notamment le critère AIC. Ce critère juge de la qualité d'un modèle et plus la valeur de ce critère est petite, meilleur sera le modèle.

D'autres critères peuvent être pris en compte pour juger de la qualité du modèle par exemple le R^2 ou le test de Wald. Le R^2 permet également de juger de la qualité d'un modèle, mais il est plus sensible aux nombres de variables présentes. Le test de Wald, quant à lui, permet de tester le modèle complet contre le modèle comprenant uniquement la constante. Cela permet de savoir s'il est pertinent de laisser les variables.

Il est indispensable de vérifier que les deux échantillons créés ont les mêmes caractéristiques pour ne pas qu'il y ait un déséquilibre et un biais lors des estimations.

Pour valider le modèle, le modèle construit est testé sur l'échantillon test en calculant l'erreur quadratique moyenne (MSE) et l'erreur absolue moyenne (MAE). L'erreur quadratique moyenne est la moyenne arithmétique des carrés des écarts entre les prévisions et les observations et l'erreur absolue moyenne est la moyenne arithmétique des valeurs absolues des écarts entre les prévisions et les observations.

Plus la valeur de ces indicateurs est faible plus le modèle est valide. Ces indicateurs nous permettront de comparer les modèles entre eux. De même, le modèle peut être validé en confrontant les valeurs observées et les valeurs prédites.

Le choix du modèle final sera fait en se basant sur l'ensemble des critères présentés dans le tableau ci-dessous.

Tableau 8 : Robustesse des différents modèles

Performance du modèle	« Mapping » sur les réponses	« Mapping » sur score global	Choix final
Sélection	Pas à pas (<i>Backward</i>)		

Qualité de l'ajustement	Pseudo R ²	R ²	Meilleurs indicateurs
	BIC	R ² ajusté	
	AIC	AIC	
LogL	BIC		
Pouvoir prédictif	RMSE Prédit vs observé		
Validation	Validation croisée (VPress)		

Le mapping sur les réponses des dimensions de l'EQ-5D-3L permettra ainsi de prédire le profil de chaque patient et de lui associer son score d'utilité. Le score obtenu sera alors comparé au score global obtenu par la régression linéaire.

La méthode que nous proposons présente trois avantages. Tout d'abord elle permet de disposer d'une **mesure exhaustive combinant en un score unique**, à la fois les symptômes de la maladie, ses diverses répercussions physiques, sensorielles, socio-émotionnelles, cognitives et les effets indésirables du traitement. Elle permet de calculer une **espérance de vie pondérée en fonction de la qualité de vie**, ce qui n'est pas le cas des profils qui se contentent d'étudier le retentissement multiforme de la maladie de façon intemporelle. **Le rapprochement des résultats et des coûts est possible, dans la mesure où ces deux paramètres ont une dimension cardinale faible.**

5 Valorisation des moyens mis en oeuvre

5.1 Point de vue retenu

Le cadrage des coûts à prendre en compte dans toute évaluation économique doit correspondre aux préoccupations budgétaires des décideurs auxquels elle est destinée. Quatre types d'interlocuteurs sont *a priori* identifiables :

- les producteurs de biens et de soins médicaux (en premier lieu les hôpitaux, mais aussi les médecins de ville),
- les acheteurs en charge du financement des dépenses reconnues et remboursées (caisses d'assurance maladie),
- les régulateurs du système de soins qui surveillent l'évolution de la consommation de bien
- et services médicaux (la DSS) et la collectivité appréhendée en tant qu'entité politique constituée.

Il est généralement admis que la perspective sociétale doit être privilégiée puisqu'elle transcende les barrières budgétaires, tout en mettant en évidence les délestages de responsabilité qui peuvent se produire au détriment des familles que ce soit par l'élévation des tickets modérateurs de droit ou de fait ou par l'accroissement des charges de garde ou de soins des aidants familiaux. Trois champs sont analysables dans cette optique. Le premier se limite à l'évaluation des consommations du secteur sanitaire au sens où l'entendent les Comptes de la Santé. Les ressources prises en compte dans la valorisation correspondent alors aux postes de la consommation de soins et de biens médicaux – tickets modérateurs inclus – qui constituent l'essentiel de la consommation médicale finale. Le second plus extensif, prend en compte les coûts médico-sociaux laissés à la charge des familles parfois qualifiés de coûts non médicaux directs. La définition la plus large du coût de la santé pour la collectivité intègre la valeur des pertes de production

pour l'économie générale dues à la morbidité lorsque celle-ci s'accompagne d'une sortie provisoire ou définitive du monde du travail (ces coûts étaient autrefois appelés coûts indirects). **La feuille de route nous incite à étendre le périmètre de notre analyse à la troisième définition : les coûts sociétaux recensés dépasseront le seul décompte des coûts médicaux et non médicaux directs pour intégrer des pertes de production.**

Le périmètre retenu est donc plus large que celui défini par les dépenses reconnues et remboursées par l'assurance maladie puisqu'il intégrera les restes à charge et les coûts pour les aidants familiaux ; mais en même temps il sera plus étroit qu'il aurait pu l'être si l'ensemble des prestations en nature ou en espèce de l'assurance maladie était pris en compte. Sa spécificité par rapport à la perspective « système de soins » sera tout aussi réelle. A la différence des calculs qui sont habituellement faits dans cette optique, les coûts indirects : pertes de productivité liées à l'absentéisme, et au renouvellement du personnel figureront dans les décomptes.

5.2 Ressources consommées

Seront incluses dans l'analyse les variables suivantes :

5.2.1 Consommations médicales directement liées aux soins

- Hospitalisations

Les venues à l'hôpital aux modalités d'administration de certains des traitements et les ré-admissions hospitalières dues aux complications liées ces prises en charge feront l'objet d'un décompte systématique dans le cadre de la comparaison directe des événements morbides de chaque option thérapeutique.

- Prescriptions médicamenteuses

Afin d'alléger le cahier d'observation, seules les prescriptions médicamenteuses liées à l'utilisation aux différents protocoles étudiés seront recueillies.

- Actes para cliniques

Ces consommations recouvrent :

- Les différents actes radiologiques (radiographie, tomographie haute résolution),
- Les différents actes d'exploration fonctionnelle. (Échographie, ECG)

- Consultations et actes paramédicaux

Ces consommations recouvrent :

- Les consultations et les visites à domicile considérées comme urgentes (jours fériés, visites de nuit...), du médecin traitant,
- Les consultations chez les spécialistes de ville,
- Les consultations hospitalières : consultations programmées et consultations aux urgences,
- Les actes de kinésithérapie,
- Les actes infirmiers
- Les Actes paracliniques. Ces consommations recouvrent :
 - Les différents actes biologiques,
 - Les différents actes radiologiques,

- Les différents actes d'exploration fonctionnelle
 - Les transports sanitaires
 - Les aides à domicile

5.2.2 Consommation familiale directement associée aux aides informelles

- Transport non sanitaire entre le domicile et l'hôpital
- Temps consacré par la famille à la garde et à la surveillance de ses proches
- Hébergements dans des institutions non médicalisées : EHPAD
- Un cahier patient sera programmé pour recueillir l'ensemble des ressources consommées en milieu ambulatoire ou à domicile

5.3 Valorisation des coûts

A l'hôpital, la valorisation des ressources consommées sera réalisée sur la base des données médico-économiques du Programme de Médicalisation des Systèmes d'Information (PMSI). Chaque séjour hospitalier en court séjour sera classé via la CIM 10 dans un groupe homogène de séjour (GHS) selon les règles du PMSI T2A version v11 si celle-ci est disponible, en prenant en compte le diagnostic principal, les diagnostics associés et reliés (co-morbidité), l'âge et les actes classants. Pour l'ambulatoire, les consultations médicales, les actes paramédicaux seront valorisés sur la base des cotations de la Nomenclature Générale des Actes Professionnels (NGAP) dite « résiduelle » pour les soins non repris dans la CCAM. Les actes biologiques seront valorisés sur la base de la cotation de la Nomenclature des Actes de Biologie Médicale (NABM). Les actes de radiologie et d'exploration fonctionnelle seront valorisés sur la base de la dernière version de la Classification Commune des Actes Médicaux (CCAM) lorsqu'il s'agira d'actes techniques. Les actes des auxiliaires paramédicaux seront valorisés à partir de la valeur de la lettre clé affectée à leur profession pondérée par le coefficient correspondant à la technicité de l'acte réalisé. Les transports médicalisés seront valorisés sur la base des forfaits de prise en charge départementaux et des tarifs kilométriques spécifiques à chaque catégorie de transport mise en œuvre. La valeur du temps passé par les aidants familiaux au chevet de leurs proches sera estimée en fonction du salaire moyen de la catégorie professionnelle dont ils relèvent (31 410€ /an pour les cadres ; 23 548 € /an pour les professions intermédiaires ; 17729 € /an pour les employés ; Si la profession du chef de famille est inconnue, on prendra le salaire moyen toutes catégories soit 25 304€ /an Source INSEE). L'ensemble de ces coûts seront actualisés au taux de 4%.

5.4 Exploitation des banques de données médico administratives (BDMA)

Une nouvelle ère s'amorce aujourd'hui avec la recherche de l'effectivité relative des traitements en vie réelle par rapport à des comparateurs actifs. Les essais randomisés ont été construits pendant très longtemps sur des protocoles mille fois trop limités : trop courts dans leur durée, trop sélectifs dans leur recrutement, trop protocolés dans leurs modalités de prise en charge, si bien que la portée de leurs enseignements, pour le suivi en pratique courante des malades tout venant, était extrêmement réduite. Le comparateur choisi était, en outre, souvent un placebo, alors que le problème pour les cliniciens n'est pas de savoir s'il faut traiter de préférence à ne rien faire, mais de choisir, parmi les traitements actifs disponibles, celui qui est le plus adapté à son patient.

Faute de pouvoir généraliser les résultats des essais randomisés en pratique médicale quotidienne, les médecins de santé publique et les épidémiologistes ont recommandé d'exploiter les bases de données médico-administratives. Ces bases constituent autant de sources secondaires, au sens qu'elles ont été mises

en place pour atteindre des objectifs autres que ceux pour lesquels elles sont actuellement exploitées. La plupart du temps, il s'agit de bases de remboursement (SNIIRAM) ou de financement hospitalier (PMSI) auxquelles on peut faire jouer un rôle, en termes d'étiologie clinique. Ces bases sont facilement accessibles, mobilisables à très court terme, pour un coût modeste. Elles sont par ailleurs représentatives parce qu'exhaustives et leur contenu reflète la réalité des pratiques ce qui contribue à renforcer leur validité externe (généralisation) des résultats. Tout biais de mémorisation ou de non-réponse est par définition exclu. Les banques de données se caractérisent également par la possibilité de mesurer les résultats observés à la suite de décisions individuelles. On peut alors dresser le profil longitudinal de l'utilisation des traitements, ce qui permet de comparer les résultats obtenus sous différentes options thérapeutiques pendant la même période d'observation, tout comme dans un essai clinique. Les données sont recueillies rétrospectivement (une fois que les décisions thérapeutiques ont été prises et que le traitement a été administré) de sorte que ce recueil n'a aucune incidence sur le déroulement de la prise en charge clinique. En revanche, la démonstration du lien de cause à effet entre le traitement et les résultats obtenus est délicate compte tenu de la présence d'interférences parasites.

Comment arriver à reconstituer un parcours de soins à partir des informations du SNIIR-AM ?

Le plus simple est de prendre un exemple, fictif, mais plausible, de parcours de soins. Il s'agit ici du cas d'une personne qui souffre de PR. Elle consulte d'abord un médecin généraliste, qui lui prescrit des médicaments et une analyse médicale, avant de l'adresser au rhumatologue. Celui-ci va l'examiner, lui prescrit à nouveau des médicaments (les mêmes ou d'autres), un complément d'analyse et une radio. Lors d'une consultation suivante, il décidera ou non de changer de traitement au vu du résultat de ces examens. Selon le traitement prescrit, il pourra décider de faire hospitaliser son patient. L'épisode se conclura par un examen de sortie, une prescription d'arrêt de travail, de médicaments et d'une radio de contrôle. Le généraliste de ville prend alors le relais, pour un suivi après retour à domicile. La pathologie n'est pas disponible dans le SNIIRAM, néanmoins, les spécialistes impliqués dans les soins, les médicaments prescrits, les analyses médicales effectuées, considérés comme des traceurs de cette pathologie, autorisent une analyse comparée du coût et de l'efficacité des prises en charge.

5.5 Analyse statistique des coûts

La méthode la plus classique pour comparer les coûts moyens entre deux groupes de traitement est le test de Student. Ce test est toutefois fondé sur l'hypothèse d'indépendance des données et de normalité des distributions de coûts dans chacun des groupes comparés, ce qui va à l'encontre de la nature de la distribution des coûts, le plus souvent dispersée vers la droite. L'hypothèse de normalité a donc toutes les chances de ne pas être respectée. Le test de Mann-Whitney pourrait éventuellement être utilisé puisqu'il ne fait aucune hypothèse sur la distribution des données. Toutefois, il ne compare pas les coûts moyens, mais les coûts médians. De plus, les résultats sont sensibles à l'hétérogénéité des variances des deux groupes il serait enfin possible d'obtenir un autre estimateur de la moyenne en supposant que les données sont log-normales. Toutefois, la comparaison des coûts associés par cette méthode n'est justifiée que si la loi log-normale est proche de la vraie distribution des coûts. Une alternative à ces méthodes paramétriques et non-paramétriques de comparaison des coûts entre deux groupes de traitement est le **bootstrap non paramétrique** ; c'est celle qui sera mise en œuvre.

6 Modélisation médico-économique

6.1 Les différents types de modélisations envisageables

Plusieurs types de modélisation sont envisageables. Les économistes de la santé qui effectuent des analyses coût-efficacité utilisent les modèles de décision pour estimer les coûts à long terme et les résultats d'efficacité (QALY) associés aux différentes options de traitement pour les groupes de patients définis (Briggs et al., 2006). Le plus souvent ceux-ci prennent la forme des modèles de cohorte, tels que les arbres de décision ou les modèles de Markov multi états . Comme tous les patients se déplacent en « meutes », les coûts et les QALY peuvent être facilement calculés pour chaque cycle et s'accumulent sur l'horizon temporel retenu. Ce type de modèle est généralement beaucoup plus rapide à exécuter qu'un modèle de microsimulation, mais il a du mal à rendre compte de la progression complexe d'une maladie dont l'évolution est liée aux caractéristiques individuelles du patient (Karnon, 2003). Deux approches du travail de modélisation demandé peuvent être envisagées :

- ✦ **Une approche à l'échelle d'une population** : qui correspond à la majorité des modèles utilisés dans les évaluations économiques.
 - C'est alors la **distribution de l'espérance des paramètres cliniques** qui nous intéressera. [E[T] puisque ce sera la valeur centrale d'une distribution de probabilité qui est une représentation de l'incertitude de la moyenne dans la population, alors que la moyenne caractérise un échantillon]
 - Loi normale : $E[T] \sim \text{Norm}(\mu, \sigma^2)$,
 - $\sigma^2 = s^2/n$ est une représentation de l'incertitude sur la variabilité de la moyenne dans la population

- ✦ **Une approche « individus centrés »** :
 - C'est la **distribution des paramètres** T (la taille par ex) **dans l'échantillon** qui nous intéressera pour pouvoir les utiliser comme estimateurs des paramètres liés à la population.
 - Loi Normale : $T \sim \text{Norm}(X, s^2)$, avec :
 - X = moyenne de l'échantillon
 - s^2 = variance estimée à partir de l'échantillon

Récemment une revue systématique a été réalisée sur les études économiques disponibles (Scholz and Mittendorf, 2014). Les forces et faiblesses des modélisations utilisant l'une ou l'autre de ces deux approches y ont été examinées.

6.1.1 Arbres de décision

Le type de modèle le plus simple à utiliser en économie de la santé est l'arbre de décision. Le problème thérapeutique est décomposé spécifiquement à l'aide d'un arbre de décision qui comporte trois types de nœuds : les nœuds de choix, les nœuds aléatoires et les nœuds terminaux. Les embranchements issus des premiers correspondent aux différentes possibilités d'action thérapeutique. Les branches émanant des nœuds aléatoires représentent les divers événements qui peuvent survenir à l'issue des initiatives prises. Enfin, les nœuds terminaux enregistrent à la fois les résultats obtenus et les coûts.

Pour chaque modalité de traitement envisagé, il est calculé une espérance globale d'efficacité et une espérance globale de coût. Pour ce faire, on part des nœuds terminaux où sont inscrits conjointement le coût du traitement et les résultats obtenus. On calcule pour chaque nœud aléatoire la valeur des branches qui en sont issues en multipliant la valeur du nœud terminal par la probabilité de la branche en question et en sommant les produits correspondants. La valeur de chaque nœud aléatoire est donc égale à la somme des valeurs de chacune des branches (procédure du calcul de la moyenne). La remontée de l'arbre s'effectue en descendant des plus hautes branches vers les racines, chaque nœud aléatoire distribuant à son tour sa valeur à la branche à laquelle il est attaché (procédure du chaînage arrière).

Ces arbres peuvent être construits pour suivre le cheminement de patients individuels ou des cohortes sans qu'aucune interaction entre les individus qui la compose ne soit possible. Néanmoins, ils ne peuvent être utilisés que pour des périodes de temps déterminées, ce qui rend la simulation de l'évolution des maladies sur des périodes plus ou moins longues très compliquée. En outre, pour répondre à l'hétérogénéité entre les groupes de patients, les paramètres d'entrée doivent être changés et le modèle doit être relancé séparément pour chaque sous-groupe.

6.1.2 Modèles de Markov

Les modèles de Markov appartiennent à la famille des modèles dynamiques et stochastiques. Un processus de Markov dont les probabilités de transition varient en fonction du temps (« time-dependent Markov chain ») relève des techniques de modélisation dynamiques dont l'évolution est soumise ou non à la survenue d'événements aléatoires. Ces modèles sont très utilisés dans le domaine de l'évaluation médicale et médico-économique (Briggs et al., 2006) notamment parce qu'ils permettent de « mimer » l'évolution dans le temps d'une cohorte et ceci sur une longue période.

Dans un modèle de Markov, les individus se trouvent à chaque période dans un état S_i appartenant à un ensemble fini S d'états possibles. Les états de Markov sont définis de manière exhaustive (à la période t , un individu se trouve nécessairement dans l'un des états qui a été prédéfini) et exclusive (à la période t , un individu ne peut se trouver que dans un seul état).

Dans le cas d'un modèle à temps discret, il importe de définir la durée des cycles. Celle-ci est choisie en fonction des critères cliniques d'évolution de la pathologie. La dynamique du modèle est déterminée par les probabilités de transition. D'une période à l'autre, les individus changent d'état selon en fonction des valeurs prises par celles-ci. Les probabilités de transition déterminent le nombre de patients qui restent dans un état de santé à la fin d'un cycle et le nombre de sujets qui se déplacent à vers les autres états de santé. P_{ij} désigne la probabilité conditionnelle pour un groupe d'individu d'être dans l'état j à la période t alors qu'il était dans l'état i à la période précédente. L'ensemble des probabilités de transition P_{ij} constitue la matrice de transition P du modèle. Selon la propriété fondamentale d'un modèle de Markov (propriété dite « markovienne »), l'état dans lequel se trouve un patient à un instant t ne dépend que de l'état dans lequel il se trouvait à la période $t - 1$. L'état du système aux périodes antérieures ($t - 2, t - 3$, etc.) n'intervient pas et le processus est dit « sans mémoire ». Cette propriété confère au modèle toute sa puissance et sa simplicité. À l'usage, elle s'avère être peu restrictive.

Un système markovien et sa dynamique sont donc complètement décrits par :

- L'ensemble S des états de Markov,
- La matrice de transition P comportant les probabilités de transition d'un état vers l'autre y compris les probabilités de transition d'un état vers lui-même pour la fraction de la population dont la maladie reste stable d'une période à l'autre,
- La répartition initiale, au temps t_0 des patients entre les états.

- Aucun phénomène d'interaction ou de contagion n'est introduit entre les sujets

Ces éléments permettent de calculer très simplement l'état du système à n'importe quel moment du temps. Il est alors possible de déterminer la trajectoire du système et d'étudier son comportement asymptotique dans le temps. Les modèles de Markov sont généralement simulés jusqu'à ce que tous les patients d'une cohorte aient atteint un état absorbant (par exemple la mort) ou un nombre maximal de cycles.

Comme les arbres de décision, les modèles de Markov se prêtent à des analyses en sous-groupes qui permettent de prendre en compte l'hétérogénéité de la population. Quand le nombre de sous-groupes devient grand, le temps machine requis pour réaliser les analyses par sous-groupe, en utilisant à chaque fois des paramètres différents, devient vite élevé.

La faiblesse majeure des modèles de Markov agrégés est liée au fait que ce sont des modèles sans mémoire intégrée. En d'autres termes, les antécédents de l'individu ne sont pas pris en compte d'un cycle à l'autre, et à fortiori au cours des cycles suivants. De surcroît la diversité des terrains individuels et la variabilité des paramètres cliniques d'un patient à l'autre ne sont pas appréhendées. Les modèles d'échantillonnage individuel qui simulent le devenir des patients un par un permettent de contourner ces limitations.

6.1.3 Modèles individus centrés

6.1.3.1 Microsimulation de Markov

La méthode la plus adaptée pour tenir compte de la diversité des tableaux cliniques rencontrés dans la PR pourrait être de développer une des formes des modèles individus centrés (Barton et al., 2004).

Comme dans les modèles de Markov agrégés, il existe dans les modèles de microsimulation de Markov, un nombre fini d'états de santé et des pas de simulation fixes, mais les caractéristiques des patients y sont décrites au cas par cas et elles peuvent être modifiées au fil du temps via la mise en place de « variables compteurs » qui permettent de retracer les antécédents des malades suivis. Ces variables seront initialisées à zéro et s'accroîtront d'une unité pour chaque nouveau traitement de la PR qui aura été administré. Étant donné que chaque type de traitement présente des spécificités et des implications futures propres, nous créerons un compteur pour chaque séquence thérapeutique. Ces différents compteurs seront ensuite utilisés pour modifier les probabilités transitoires (dont les probabilités de décès) ainsi que les récompenses et les coûts des états. Ces compteurs nous permettent également de comptabiliser le nombre total de traitements de chaque type subi par l'ensemble des individus qui sont entrés dans le processus. Ce qui permet de modifier les probabilités de transition en fonction de l'anamnèse des malades suivis. Les modèles individus centrés permettent ainsi d'éviter de multiplier les états de santé pour suivre l'évolution des prises en charge au fil du temps.

6.1.3.2 Simulation à événements discrets

Les simulations à événements discrets (SED) se concentrent sur le cours des traitements d'un patient plutôt que sur la modélisation de l'évolution de la maladie elle-même. Des parcours individuels sont simulés un grand nombre de fois afin d'obtenir une estimation la plus précise possible du critère d'intérêt. Trois éléments permettent de définir une SED : les entités, les événements, les attributs.

Les entités constituent les items qui évoluent au cours de la simulation. Ce seront les patients auxquels seront affectés un certain nombre d'attributs (âge, sexe, nombre d'épisodes antérieurs de la maladie). Ces attributs sont spécifiques à chaque patient et ils évoluent au gré de la survenue des événements.

Le deuxième constituant majeur de la SED est l'événement. Il est défini comme tout ce qui peut arriver à un patient, c'est l'équivalent des états dans les modèles de Markov ; les événements se succèdent de

manière logique ; ils peuvent avoir un impact ou non sur les états de santé, ils peuvent infléchir ou non l'évolutivité de la maladie et modifier les chances d'occurrence de maladies future. Le système est désormais doté de mémoire.

La troisième composante est le temps. Une horloge globale permet de définir la durée de la simulation et des compteurs secondaires permettent de mémoriser le temps en fonction de critères prédéfinis par exemple le temps sans progression ou sans symptômes. Il n'est plus besoin de prédéfinir la longueur des cycles une fois pour toutes. La principale différence entre la méthode de microsimulation de Markov et la simulation à événements discrets réside est que le pas de la simulation est fixe dans le premier cas alors que dans le second il est conditionné par la date de survenue de l'événement morbide.

La SED permet de modéliser les comportements des agents économiques lorsqu'ils sont mis en concurrence pour le partage des ressources rares (Scholz and Mittendorf, 2014). Dans ce contexte, des « files d'attente » se forment pour accéder aux ressources en respectant des règles prédéfinies.

Un des principaux avantages d'une SED est son aptitude à intégrer la spécificité des profils des patients traités, ce qui permet de mieux refléter le caractère hétérogène des populations qui sont prises en charge (Choi et al., 2002).

Même si la SED est une des méthodes les plus avancées de toutes celles qui sont utilisées dans les évaluations économiques, elle n'est pas sans inconvénient. Une multiplicité d'événements étant possible pour chaque individu en fonction de ses propres caractéristiques et de son passé, de très nombreuses données doivent être disponibles et le temps nécessaire pour les réunir peut-être excessivement long.

6.1.3.3 Avantages des modèles individus centrés

Les modèles individus centrés où l'évolutivité de la maladie est fonction des caractéristiques de la personne malade sont mieux à même de refléter la réalité des prises en charge que les approches en termes de cohorte où les changements de traitement reposent sur des probabilités moyennes d'évolution d'un groupe d'individus. Les modèles de cohorte peuvent certes, être aménagés, pour qu'ils puissent mieux refléter les caractéristiques des patients - par exemple, en augmentant le nombre des branches dans un arbre de décision, ou en augmentant le nombre d'états de santé dans un modèle de Markov. Cependant, ces approches deviennent rapidement difficiles à gérer lorsqu'un grand nombre de caractéristiques doit être pris en compte. Le modélisateur est alors amené à multiplier les analyses en sous-groupes (Briggs et al., 2006), notamment lorsque certains traits caractéristiques des profils-patients sont fortement corrélés, par exemple l'âge et l'indice de masse corporelle. Dans le cadre d'une microsimulation, les corrélations entre les caractéristiques initiales peuvent être capturées automatiquement par le biais d'un jeu de données relevées en vie réelle.

Contrairement aux modèles de Markov, les modèles de microsimulation, qui simulent le devenir des patients un par un, permettent d'incorporer l'historique de chaque « individu » dans le modèle. Les événements qui se sont déjà produits pour un individu particulier sont mémorisés et les probabilités de transition sont modulées en fonction de ses antécédents et de son âge. Le modèle de microsimulation est un outil approprié pour prendre en compte les caractéristiques de la personne malade et créer une structure de modèle efficace et compact. La capacité d'intégrer les caractéristiques des patients permet également d'individualiser les coûts et QALY estimés, ce qui permet d'estimer quels sous-groupes de patients tirent le meilleur parti des traitements qui leur sont administrés, tout en étant à même de produire des résultats au niveau de la population tout entière dans le cadre d'un modèle unique.

6.1.3.4 Inconvénients des modèles individus centrés

L'un des inconvénients le plus évident des modélisations individuelles est qu'elles sont très consommatrices de données. Un autre inconvénient est le temps qu'il faut pour développer et exécuter le modèle. Bien que le temps d'exécution du modèle soit aujourd'hui un problème moins important avec l'augmentation de la puissance des ordinateurs, il n'en demeure pas moins que les modèles individuels centrés exigent plus de temps que les modèles de cohorte qui sont généralement utilisés par les économistes de la santé. Ce problème peut être un obstacle important si une analyse de sensibilité probabiliste doit être réalisée pour caractériser l'incertitude sur les paramètres au niveau de la population, car cela nécessite une boucle externe d'itérations, multipliant le temps nécessaire pour faire fonctionner un modèle. Un autre inconvénient de ce type de modèle est le temps requis par l'écriture du script. Lorsqu'un modèle devient plus complexe, le temps de programmation et de validation augmente de façon exponentielle.

Les auteurs de l'International Society for Pharmacoeconomics and Outcomes Research (ISPOR) préconisent de développer un modèle agrégé de cohorte, lorsque le nombre d'états de santé qu'il est nécessaire de mettre en place pour illustrer l'évolution pathologique avec suffisamment de nuances reste raisonnable. Si tel n'est pas le cas, alors le recours à une modélisation individuelle de type Markovien devrait être envisagée. La mise en œuvre des modèles de simulations à événements discrets serait plus particulièrement recommandée pour modéliser les problèmes causés par des ressources limitées (Siebert et al., 2012).

6.2 Hypothèse de travail sur le type de modèle à retenir

Pour faire le choix du modèle, il faut prendre en considération :

- la particularité de la maladie ;
- les données disponibles ;
- les délais pour l'exécution du modèle.

Le modèle de microsimulation individuel (Individual Patient Simulation, IPS) pourrait être envisagé pour estimer les coûts totaux et les QALYs de chaque stratégie de traitement de la PR après l'échec (ou d'intolérance) d'un traitement initial MTX ou anti-TNF. IPS permet efficacement prendre en compte l'hétérogénéité des patients et les impacts des risques sur les séquences de traitements.

L'appellation désigne le processus qui fait varier les probabilités de passage d'un état à un autre de façon aléatoire. Les sujets sont envoyés un par un dans le modèle. Au franchissement des embranchements de l'arborescence, on simulera une des réalisations possibles d'une loi uniforme [0,1] en adoptant des hypothèses sur le devenir des patients, en fonction de la valeur du HAQ par exemple. À titre purement illustratif, on peut imaginer les seuils suivants :

- Pour un tirage au sort dont la valeur est comprise entre les seuils [0, 0.1] de l'HAQ, le sujet conserve le même traitement ;
- Pour un tirage au sort dont la valeur est comprise [0.1, 0.3], le sujet conserve le même traitement, mais la fréquence des administrations est augmentée ;
- Pour une valeur est comprise [0.3, 1], le sujet change de traitement.

Lorsque l'expérience est répétée sur un grand nombre d'individus, la moyenne des résultats du modèle sur un grand nombre d'itérations, est la même que celle qui serait obtenue dans un modèle de simulation agrégé de cohorte. L'évaluation stochastique permet d'estimer la distribution des résultats en plus de leur valeur moyenne.

Un patient dans les modèles IPS présentera un faisceau de caractéristiques initiales définies à partir de l'âge, du sexe, du score HAQ, de l'ancienneté de la maladie et du nombre de traitements précédemment administrés. L'avantage d'un modèle IPS par rapport aux méthodes les plus couramment mises en œuvre : arbres de décision de cohorte et les modèles de Markov, tient au fait que l'avenir des patients dans ce type de modèle dépend de leur passé. Les traitements seront prescrits en fonction de la nature de traitements qui leur ont été précédemment administrés et du stade de la maladie dans lequel ils se trouvent. Le descriptif des profils des patients sera limité seulement par la possibilité d'avoir accès aux données qui permettront de les caractériser de façon précise. De nombreuses variables sont nécessaires pour prédire la réponse ou la non-réponse du patient au traitement, mais les données correspondantes ne sont pas toujours disponibles.

L'évaluation dans le modèle sera basée sur l'évolution du score HAQ du patient au fil du temps qui déterminera indirectement le nombre de QALY's gagnés ou perdus. Dans le cadre de la simulation, le score HAQ des patients sera évalué tous les 6 mois. À la fin de chaque cycle, un patient pourra être soit un « répondeur » au traitement auquel cas il restera sous ce traitement, ou non répondeur auquel il se verra administrer un autre traitement. Le niveau moyen d'amélioration HAQ pour les répondeurs sera quantifié à partir des données des essais cliniques publiées. Le modèle adoptera une perspective vie entière en intégrant la probabilité de décès, qui sera basé sur les données de survie des patients avec PR. La simulation sera répétée pour 10 000 patients sur des cycles de 6 mois. Le score HAQ sera converti en QALY en utilisant le modèle de régression de HAQ sur l'EuroQol (EQ-5D) présenté précédemment. Les résultats des 10 000 itérations fourniront les QALY et les coûts moyens pour chaque bras.

6.3 Réalisation pratique du modèle

6.3.1 Application de la simulation individuelle

Le modèle de microsimulation de type Markov est généralement structuré autour des états de santé mutuellement exclusifs, où les probabilités de transition assurent le mouvement à travers les états à chaque cycle discret. La principale différence avec le modèle de Markov de cohorte est la possibilité de simuler les individus, permettant le suivi de l'histoire de chaque individu simulé.

La microsimulation peut être programmée sur *Microsoft Excel* à l'aide de Visual Basic pour Applications VBA, qui a plusieurs avantages :

1. Premièrement, Excel est universellement disponible et largement accepté ;
2. L'utilisation de Visual Basic pour Applications (VBA) lui donne une grande flexibilité ;
3. Les feuilles de calculs rendent la mise en place d'interface du modèle relativement facile.

Par contre, le temps d'exécution du modèle peut être un problème quand il y a des interactions fréquentes entre VBA et la feuille de calcul. La réalisation du modèle de microsimulation demande beaucoup de programmation de VBA qui n'est pas assez transparent pour les agences d'ETS. Il manque aussi quelques fonctions statistiques nécessaires (par exemple la distribution Gompertz). La programmation du modèle de ce type peut être compliquée étant donné le nombre des états et des règles logiques requis. Dans une approche de calcul sur Excel, les erreurs de programmation peuvent se produire facilement, et elles sont difficiles à détecter.

D'autre part, l'utilisation de *TreeAge* (module « Microsimulation mode ») peut être plus adaptée, car il permet d'utiliser les variables-compteurs pour capturer l'histoire de la maladie et de construire la matrice de transition d'une manière plus transparente.

TreeAge a trois fonctionnalités, qui facilitent le développement du modèle de microsimulation :

1. Variables-compteurs (*tracker*) pour capturer les caractéristiques individuelles ;
2. Nœud logique (*logic node*) pour la transition ;
3. Matrice globale (*global matrix*).

Les variables-compteurs ou les variables locales $\{T\}$ sont utilisées pour sauvegarder l'information sur les événements et l'histoire de la maladie d'un individu. Elles capturent aussi les coûts, les événements indésirables, ou d'autres informations. Ces variables sont mises à jour durant la simulation, et elles contiennent l'information personnalisée qui détermine le parcours individuel du patient.

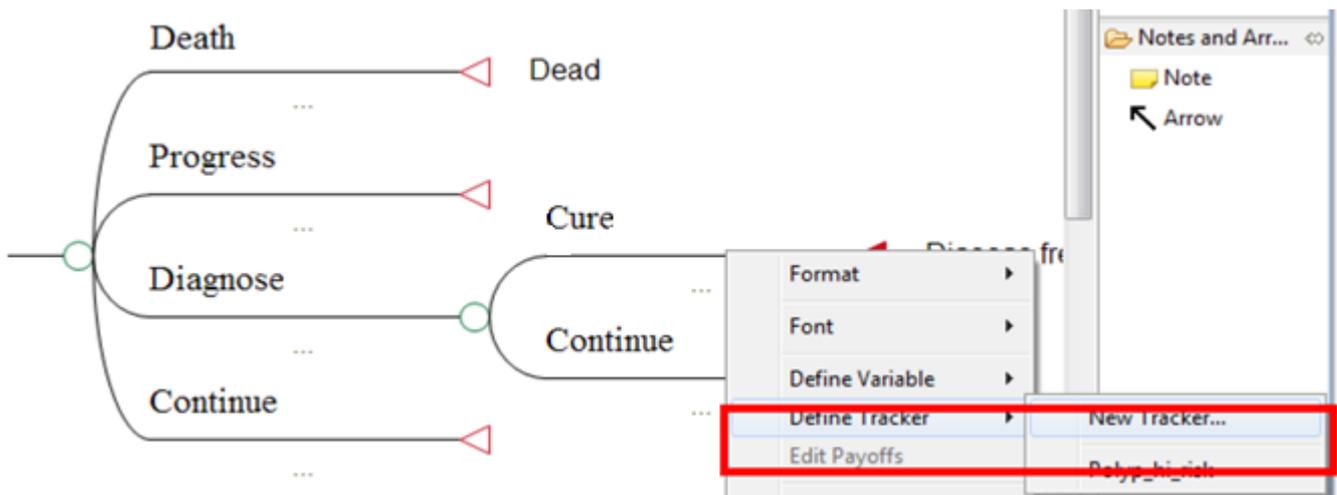


Figure 6. Exemple de définition d'un nouveau compteur sur un arbre de décision.

Les nœuds logiques permettent de choisir un parcours en fonction des conditions logiques. Par conséquent, les conditions logiques pour chaque alternative doivent être définies. À partir de la branche supérieure, les conditions sont évaluées pour chaque patient. Lorsque la condition logique est vraie, le patient suit ce chemin. Lorsque la condition logique est fausse, il évalue la condition logique de la branche suivante.

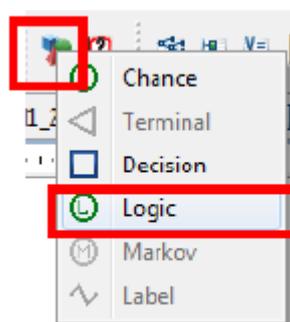


Figure 7. Définition d'un nouveau nœud logique.

La matrice globale stocke toutes les données durant la simulation :

$$M = \begin{pmatrix} Pat1 & HealthState(t0) & HealthState(t1) & \dots \\ Pat2 & HealthState(t0) & HealthState(t1) & \dots \\ \dots & HealthState(t0) & HealthState(t1) & \dots \\ Patm & HealthState(t0) & HealthState(t1) & \dots \end{pmatrix}$$

Plutôt que de représenter l'incertitude sur les paramètres pour l'analyse de sensibilité probabiliste, les distributions paramétriques sont utilisées pour représenter la variabilité individuelle dans le cadre du modèle de microsimulation.

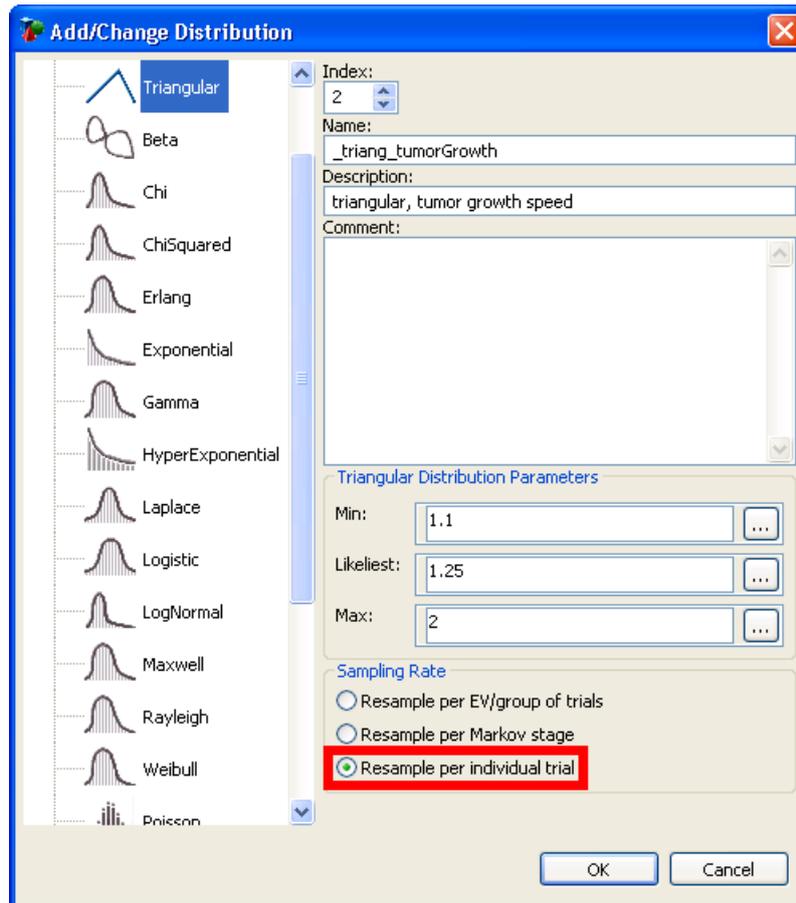


Figure 8. Exemple de la définition de la distribution.

Ces distributions peuvent être utilisées pour les valeurs qui varient entre les individus ou d'un cycle à un autre. Dans TreeAge, ceci est implémenté en changeant le taux d'échantillonnage (*Sampling rate*) dans les propriétés de la distribution.

Durant toute microsimulation (y compris une boucle interne dans une ASP à 2 dimensions), les distributions qui sont modifiées pour « *Resample per individual trial* » (figure 8) lors de la simulation vont automatiquement être tirées à partir de la distribution appropriée. Chaque individu va obtenir efficacement son propre tirage aléatoire.

Une fois la simulation est réalisée, les statistiques sont disponibles. Un exemple des résultats de la simulation de Monte-Carlo sont présentés sur la figure 9.

Monte Carlo Statistics	
Statistic	Population w/ Polyps
Mean	19,691
Std Deviation	1,7405513494
Minimum	2
2.5%	14
10%	20
Median	20
90%	20
97.5%	20
Maximum	20
Sum (n*Mean)	19691
Size (n)	1000
Variance	3,029519
Variance of Mean (var/n)	0,003029519
Std Error of Mean	0,0550410665

Figure 9. Résultats de la simulation Monte-Carlo.

La matrice globale qui trace les parcours des patients peut être facilement exportée sous fichier Excel :

A1		Trial								
	A	B	C	D	E	F	G	H	I	
1	Trial	State(t0)	State(t1)	State(t2)	State(t3)	State(t4)	State(t5)	State(t6)	State(t7)	
2	1	1	1	1	1	1	1	1	1	
3	2	1	1	1	1	1	1	1	1	
4	3	1	1	1	1	1	1	1	1	
5	4	1	1	1	1	1	1	1	1	
6	5	1	1	1	1	1	1	1	1	
7	6	1	1	1	1	1	1	1	1	
8	7	1	1	1	1	1	1	1	1	
9	8	1	1	1	1	1	1	1	1	
10	9	1	1	1	1	1	1	1	1	

Figure 10. Extraction Excel de la matrice globale.

6.3.2 Application de la simulation des événements discrets

Ce type du modèle représente une autre approche de simulation au niveau individuel. La SED permet de conceptualiser :

- l'évolution de la maladie ;
- la gestion de la maladie en fonction des événements qui peuvent se produire chez les individus ;
- l'effet de ces événements sur leur santé actuelle ;
- le risque en futur.

Les événements sont des éléments centraux dans SED et ils sont définis comme tout ce qui peut arriver à un individu au cours de la simulation. Le temps de l'événement, est au cœur de SED et il est continu (contrairement au modèle de cohorte, où le temps est divisé en intervalles discrets). Les mesures continues de l'état (par exemple changement du cours de la maladie) peuvent également être facilement gérées.

Les SED peuvent être programmées sous le logiciel de modélisation dédié comme *Rockwell Arena Simulation Software*. Ce logiciel est conçu pour gérer la logique du modèle très complexe, puisqu'il a le cadre intégré pour produire les modèles « *time to event* ». Le processus de construction d'un modèle dans Arena est simple avec l'expérience.

Le principe de fonctionnement du logiciel Arena est de suivre chacune des entités évoluant d'un état de santé vers un autre dans le modèle. L'ordonnement dans le temps des différents événements rattachés à l'évolution des entités dans les états constituant le modèle se fait au travers d'un échéancier.

Arena permet de construire un modèle en proposant des modules de représentation plus ou moins détaillée. Il permet également de créer des animations graphiques pour visualiser le comportement du modèle durant la simulation. Les blocs sont regroupés dans différentes bibliothèques (templates). Afin de bénéficier d'une animation, on peut utiliser les templates Basic Process et Advanced Process.

6.4 Les données à introduire dans le modèle

L'analyse de référence retient l'analyse coût-utilité et l'analyse coût-efficacité comme méthodes d'évaluation.

6.4.1 Population d'analyse

La population d'analyse, retenue dans l'analyse de référence, est l'ensemble des individus avec les symptômes de la PR après l'échec d'un traitement initial en France. Le modèle d'analyse choisi permet d'effectuer l'évaluation au niveau individuel de chaque patient et de considérer l'hétérogénéité des patients, donc cette analyse ne nécessitera pas l'analyse de sous-populations spécifiques.

6.4.2 Choix de l'horizon temporel

Un horizon d'analyse vie entière sera choisi dans l'analyse de référence pour intégrer l'ensemble des différentiels de coût et de résultat attendu. Des horizons temporels à court et à moyen terme seront considérés dans les analyses de sensibilité.

6.4.3 Choix du taux d'actualisation

Les coûts et les résultats futurs seront actualisés afin de les ramener à leur valeur présente.

Selon les recommandations du guide pour l'évaluation économique de la HAS, l'analyse de référence mobilise le taux d'actualisation public fixé à 4 % depuis 2005. Ce taux peut faire l'objet d'une réévaluation. Dans l'analyse de référence, la HAS préconise de considérer que le prix relatif du résultat de santé pour la collectivité est invariant au cours du temps. Les coûts et les résultats sont donc actualisés au même taux.

Une analyse de sensibilité sera réalisée pour apprécier la robustesse des conclusions de l'évaluation au choix de la valeur du taux d'actualisation.

7 Identification et quantification des incertitudes

7.1 Sources des incertitudes

Les notions de variabilité, d'hétérogénéité et d'incertitude sont essentielles dans l'évaluation quantitative du risque. La variabilité, encore appelée incertitude du premier ordre, incertitude aléatoire ou incertitude stochastique se réfère à une caractéristique intrinsèque du phénomène étudié à laquelle il est impossible d'échapper. Cette variabilité par essence est une donnée naturelle irréductible. Des différences à l'intérieur d'un groupe seront toujours observées. L'augmentation du nombre de données ne permettra jamais d'échapper à la réalité des fluctuations d'échantillonnage.

À partir du moment où les causes de la variabilité des résultats observés peuvent être rattachées à un faisceau de caractéristiques communes, la variabilité se transforme en hétérogénéité. La population cible peut alors être stratifiée en sous-groupes.

L'incertitude due à l'ignorance est au contraire de la variabilité, réductible par apport de données nouvelles, ce qui permet d'orienter les activités de recherches ultérieures et ouvre la voie aux applications de la théorie de la valeur escomptée de l'information.

7.1.1 Incertitudes stochastiques

Il s'agit d'incertitudes qui constatent le caractère éminent variable d'une réponse individuelle au traitement, qui peut être différent selon les individus même lorsque ceux-ci appartiennent à une sous population dont les caractéristiques sont globalement similaires. Ce type d'incertitude est représenté dans les modèles d'analyse de décision par les cercles qui symbolisent les nœuds de chance. Les deux branches qui se greffent sur celui lorsqu'on estime la probabilité de survivre ou non à une intervention, les deux branches qui se greffent sur le nœud de chance correspondant portent chacune les probabilités de décès et de survie qui reflète l'aléa auquel tout individu ne présentant aucun facteur de risque particulier est exposé : certains en seront les victimes, d'autres non. Une incertitude dont la nature est bien différente de celle qui conduit à s'interroger sur l'exactitude de l'estimation elle-même lorsque celle-ci provient d'un échantillon de taille réduite (incertitude paramétrique).

7.1.2 Incertitudes paramétriques

Les incertitudes paramétriques sont des incertitudes liées aux données d'entrée du modèle ; elles tiennent au fait que la valeur réelle d'un paramètre est inconnue. Il existe une grande différence entre l'aléa naturel et l'incertitude. L'aléa naturel est insensible à toute décision. L'incertitude par ignorance est décrite par un paramètre sur lequel l'apport de nouvelles données pourrait permettre de lever une partie du doute. L'incertitude des paramètres est abordée dans cette étude principalement par l'analyse de sensibilité probabiliste en faisant varier simultanément plusieurs paramètres cliniques et économiques clés.

7.1.3 Incertitudes structurelles

Ce sont les incertitudes qui résultent de simplifications, d'hypothèses liées à un manque d'information sur un système, une structure, des comportements. Leur prise en compte passe par une modélisation alternative. Le modèle choisi peut très bien ne pas correspondre à la meilleure description du phénomène

à l'étude pour les objectifs à atteindre. Dans un modèle économique, les incertitudes du modèle peuvent être nombreuses. Il est donc nécessaire d'examiner dans le détail, les sources possibles d'incertitude du construit et de les traiter en conséquence.

Au total les sources d'incertitudes par ignorance peuvent être réduites soit en partitionnant la population cible en sous-ensemble de patients, soit en colligeant des nouvelles observations, soit en formulant des hypothèses alternatives de modélisation et en assignant des probabilités subjectives aux sorties des modèles alternatifs.

7.1.4 Incertitudes par hétérogénéité

L'hétérogénéité se réfère à des situations où la diversité des réactions individuelles au traitement s'explique par un faisceau de caractéristiques communes non aléatoires et observables. Au fur et à mesure que les causes de la variété des résultats obtenus sont mieux connues, la variabilité se transforme en hétérogénéité et la population cible peut être subdivisée en sous-groupes de populations, chaque sous-groupe obtenant des résultats différents et présentant des incertitudes spécifiques.

7.2 Représentation des incertitudes paramétriques

L'approche fréquentiste considère que le paramètre est une valeur fixe, mais mal connue. Cette valeur est estimée se donnant un intervalle de confiance qui représente l'ensemble des valeurs possibles du paramètre. Aucune distribution de probabilité n'est utilisée. En revanche dans une distribution bayésienne, le paramètre d'intérêt est représenté par une distribution de probabilité. La distribution probabiliste est interprétée comme une probabilité subjective (un degré de croyance) représentant les niveaux de croyance associés aux valeurs possibles du paramètre. Ce qui exige que l'on pose des hypothèses sur la forme présumée des fonctions probabilistes à mettre en œuvre

7.2.1 La distribution normale

Cette loi reste la référence dans le domaine de la statistique de par le théorème central limite qui énonce que la moyenne d'une variable aléatoire tend vers une loi normale, quelle que soit la distribution initiale de la variable, lorsque la taille d'échantillon est suffisante (en pratique, ça donne de bons résultats pour une moyenne supérieure ou égale à 30). La loi normale est symétrique autour de son espérance et peut prendre toute valeur dans l'ensemble des réels. Elle est notée comme : $N(\mu; \sigma)$ où μ désigne la moyenne de la loi et σ^2 sa variance.

Affirmer que le nombre de survivants dans le modèle suit une loi de Weibull dont l'amplitude et la forme sont spécifiées par les valeurs de lambda et de gamma, c'est adopter une hypothèse fondamentale qui conditionne toute la structure du modèle. Proposer dans un second temps que les valeurs de lambda et de gamma soient fixées conditionnellement en fonction de la forme des lois de probabilité qui régissent leur réalisation, c'est faire une proposition sur un résultat anticipé à partir d'une idée a priori. Dans le cadre de l'analyse de sensibilité probabiliste, il est proposé de choisir une loi normale de moyenne μ et d'erreur-standard σ pour caractériser ces valeurs possibles des paramètres de la loi de Weibull. En ce sens μ et σ peuvent être qualifiés d'hyperparamètres. Ce sont en quelque sorte, des paramètres de second rang, qui règlent les valeurs possibles des distributions de probabilités quantifiant le degré de croyance sur les

valeurs possibles des paramètres qui spécifient le modèle, conditionnellement à la structure imposée par le choix d'une fonction de Weibull. Une structure complexe à deux étages typiquement bayésiens.

7.2.2 La distribution Beta

Il s'agit d'une loi très employée en statistique Baysésienne en tant que loi conjuguée d'une loi binomiale (ce qui veut dire que si la distribution *a priori* d'un paramètre est de forme Beta et que la vraisemblance des données observées est binomiale, alors la loi *a posteriori* du paramètre sera une loi Beta également). Il s'agit d'une loi particulièrement polymorphe qui est bornée sur $[0 ; 1]$. Elle permet donc de modéliser la distribution de probabilité d'une proportion sans craindre d'obtenir des proportions inférieures à 0 ou supérieures à 1 (comme ce serait le cas avec une loi normale, non bornée). La loi Beta est caractérisée par deux paramètres a et b , qui est notée : $Beta(a; b)$. Briggs et ses collègues recommandent d'utiliser la loi $Beta(a; b)$ avec $a = r$ (nombre de succès) et $b = n - a$ (nombre d'échecs $n = a + b$ effectifs), ou d'utiliser la méthode des moments, où l'on cherche à trouver a et b de façon à obtenir la variance et la moyenne observées dans les données disponibles.

7.2.3 La distribution de Dirichlet

La distribution Dirichlet est l'équivalent multinomial de la distribution Beta. La distribution Dirichlet est appropriée quand le patient peut passer d'un état à plus de deux autres états, et elle peut également être établie en fonction de données-échantillons. Elle est caractérisée par un nombre de paramètres (α_i) équivalent au nombre d'états potentiels. Quand la transition aboutit à trois états potentiels (A, B ou C), la distribution Dirichlet sera désignée Dirichlet ($\alpha_A, \alpha_B, \alpha_C$). Les paramètres représentent le nombre de personnes de l'échantillon qui se sont acheminées vers un état particulier. La probabilité de passer (transition) à l'état A est le nombre de personnes de l'échantillon qui sont passées à l'état A (α_A) divisé par le nombre total de personnes des échantillons ($\alpha_A + \alpha_B + \alpha_C$).

7.2.4 La distribution gamma

C'est une autre loi très appréciée en statistique bayésienne, car elle se conjugue bien. Tout comme la loi log-normale elle est asymétrique à droite et bornée sur $[0 ; +\infty[$. On la notera Gamma ($\alpha ; \beta$) où β est le paramètre d'échelle et α le paramètre de forme.

Sa densité est donnée par :

$$f(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}$$

Il n'existe pas de forme simple pour la valeur de la médiane.

Les paramètres α et β s'expriment en fonction de l'écart-type σ :

$$\beta = \frac{\mu}{\sigma}$$

$$\alpha = \left(\frac{\mu}{\sigma}\right)^2$$

Alors que pour estimer les paramètres d'une loi log-normale on passe par la moyenne et la médiane, pour une loi gamma on emploie la moyenne et l'écart-type. Outre cette première différence, la loi gamma a la particularité de donner plus de probabilité aux valeurs du paramètre proches de zéro que la loi log-normale. Elle est donc adaptée pour les paramètres qui prennent souvent des valeurs nulles, mais peuvent également prendre occasionnellement des valeurs importantes. Ces caractéristiques en font un instrument à privilégier pour décrire l'incertitude liée à l'estimation des coûts.

7.2.5 Méthodes d'estimation des paramètres d'une loi

Une loi de probabilité se caractérise par les paramètres qui permettent de la définir : la loi normale par exemple se caractérise par une moyenne μ et un écart-type σ . Afin d'estimer les paramètres d'une loi à partir de données observées, plusieurs méthodes sont possibles. Celles qui reposent sur le maximum de vraisemblance ou sur les moindres carrés sont inutilisables dans le cas présent : nous n'avons pas accès aux données individuelles. Les méthodes basées sur les moments sont par contre accessibles : il s'agit, à partir de caractéristiques observées dans l'échantillon (moyenne, variance, médiane, quantiles, coefficient d'asymétrie...) de calculer les paramètres de la loi en question.

7.3 Traitements des incertitudes paramétriques

7.3.1 Analyse de sensibilité déterministe (ASD)

Les conclusions que l'on peut tirer d'une analyse décisionnelle sont conditionnées par les valeurs choisies pour décrire les possibilités envisagées, tant en termes de probabilités de survenue des événements que de valorisation des résultats finaux. Les valeurs sur lesquelles se fait l'analyse proviennent le plus souvent d'enquêtes épidémiologiques et sont entachées d'incertitudes. Il importe donc de vérifier si des variations au niveau des probabilités et résultats affectés ne changent pas les conclusions. L'analyse de sensibilité va quantifier l'impact de la variation de chaque facteur sur la variation de la sortie. Cela permet de mettre en évidence les facteurs dont la variation influence le plus les variations de la sortie. Une telle étude peut être utile pour savoir, par exemple, si la solution trouvée est robuste (c'est-à-dire si elle ne va pas être modifiée pour des petites variations des facteurs) ou encore de savoir sur quels facteurs il est important de concentrer les "efforts" d'estimation pour diminuer les variations de la sortie.

Une analyse de sensibilité déterministe peut être conduite de plusieurs manières différentes.

- La méthode la plus simple est l'analyse de sensibilité univariée, où l'on modifie une par une les valeurs de chaque variable en gardant fixes, les valeurs des autres variables du modèle. Pour ce faire, il faut varier successivement (Tornado) les paramètres clés du modèle afin d'observer l'effet d'une telle modification sur les conclusions fournies. La démarche se propose d'identifier les paramètres d'entrée qui contribuent le plus à l'incertitude sur la variable de sortie et ont un impact sur la décision au regard du critère de choix qui a été retenu. Cette méthode permet l'identification des variables auxquelles le modèle est sensible.

- L'analyse de sensibilité de type seuil consiste à faire varier la valeur d'une variable jusqu'à atteindre le point où l'alternative choisie n'apparaît plus comme préférable aux autres alternatives évaluées. Elle est particulièrement utile si l'on cherche à définir des groupes cibles d'un traitement à partir des valeurs d'une certaine variable.
- Les analyses de sensibilité multivariées consistent à faire varier simultanément les valeurs de plusieurs variables. Les valeurs attendues sont alors calculées pour chaque combinaison de valeurs des variables sujettes à l'analyse.

7.3.2 Limites de l'analyse de sensibilité déterministe

Les analyses de sensibilité classiques souffrent de plusieurs limites :

- Le nombre de variables que l'on peut explorer simultanément est assez faible. L'interprétation d'une analyse de sensibilité où l'on fait varier trois paramètres simultanément est déjà assez malaisée. Pour les modèles incluant un grand nombre de paramètres, une analyse de sensibilité complète est ardue.
- Les variables non sujettes à variation sont considérées comme constantes. Par conséquent, les conclusions d'une analyse de sensibilité pour un groupe de variables donné ne sont valables que pour une combinaison donnée de valeur des variables restantes.
- L'analyse de sensibilité classique ne tient en outre pas compte du fait que certaines valeurs d'une variable sont plus probables que d'autres.
- Il est parfois difficile lors d'une analyse de sensibilité de tenir compte des corrélations entre les variables.

7.3.3 Analyse de sensibilité probabiliste (ASP)

L'analyse de sensibilité probabiliste La démarche est simple dans son principe. À chaque variable du modèle est associée non plus une valeur moyenne ponctuelle, mais une distribution de valeurs possibles associées à cette valeur moyenne ; autrement dit, à chaque probabilité est associée une distribution de probabilité.

Les fondements de l'analyse de sensibilité probabiliste sont très proches du paradigme bayésien, il est possible de classer cette méthode parmi les techniques quasi bayésiennes. Le principe de l'analyse de sensibilité probabiliste est en effet de caractériser chaque paramètre du modèle par une distribution de probabilité (un peu comme l'analyse bayésienne caractérise les paramètres étudiés par des distributions a priori, le fait de considérer un paramètre à estimer comme une variable aléatoire et non comme une valeur ponctuelle, la « vraie valeur » du statisticien fréquentiste, est d'ailleurs typiquement bayésien).

On résout ainsi un des problèmes de l'analyse de sensibilité classique, qui considère que les valeurs prises par un paramètre sont équiprobables. L'analyse de sensibilité probabiliste peut d'ailleurs rejoindre l'analyse classique lorsqu'elle caractérise les paramètres par une loi uniforme.

Une fois chaque paramètre caractérisé par une distribution, il faut simuler n réalisations de chacune des lois par Monte-Carlo et calculer les espérances d'efficacité et de coût de chaque stratégie à partir de ces réalisations. Nous obtenons ainsi un échantillon de valeurs d'efficacité et de coût qui est accessible à tous les traitements statistiques connus, en particulier ceux du datamining et de l'analyse multivariée. Ainsi on peut résoudre le problème majeur de l'analyse de sensibilité classique qui ne sait comment prendre en compte la variation simultanée de l'ensemble des paramètres. Une seconde analogie, plus hasardeuse, peut

être faite avec la statistique bayésienne : nous obtenons en effet une caractérisation des critères de jugement. Les calculs exigés pour combiner plusieurs lois de probabilités sont impossibles à réaliser manuellement ou à l'aide des tableurs classiques. Les analyses de Monte-Carlo permettent de résoudre le problème. Un exemple de la réalisation d'analyse de sensibilité probabiliste sous VBA est présenté dans l'annexe 1.

7.4 Propagation des incertitudes

À l'occasion de la mise en œuvre de ces simulations numériques, la distinction entre variabilité et incertitude retrouve toute son importance. Elle justifie notamment la séparation dans les analyses de Monte-Carlo de deux types de simulation : les simulations sur paramètres variables (nommées modélisations de premier ordre) et les simulations sur paramètres incertains (qualifiées de modélisation de second ordre).

7.4.1 Simulation de Monte-Carlo d'ordre 1

La simulation de premier ordre consiste à tirer au hasard le devenir d'un patient dans l'arborescence et à répéter l'opération sur un grand nombre de patients. C'est bien la variabilité des devenirs individuels qui est décrite. En tirant au hasard le devenir d'un grand nombre de patients, les valeurs attendues des coûts et de l'efficacité auront la même moyenne que celles obtenues en déroulant l'arbre. Dans une simulation de premier ordre, il n'existe aucune incertitude sur les valeurs choisies pour caractériser les paramètres de l'arbre. Les résultats permettent juste de déterminer la probabilité de survenue de chacun des événements décrits dans l'arbre et d'obtenir un écart-type pour les conséquences étudiées ; dans les modélisations du second ordre, une distribution bien définie est attribuée à chacune des variables de l'arbre. À chaque tirage, les variables ont une valeur unique et différente du précédent tirage. L'espérance des valeurs de chaque alternative sera par conséquent également différente. Le choix de la stratégie optimale pourra donc varier selon les tirages. Les simulations du second ordre sont donc les seules qui prennent réellement en compte l'incertitude. Ce sont celles qu'il convient d'utiliser dans le cadre des analyses probabilistes.

7.4.2 Simulation de Monte-Carlo d'ordre 2

Les étapes opérationnelles de la mise en œuvre d'une modélisation de second ordre peuvent être résumées de la façon suivante. 1) Coder les incertitudes portant sur des grandeurs incertaines par des distributions de probabilité *a priori* qui décrivent leur fréquence d'apparition anticipée. 2) Pour une famille arrêtée de courbe, caractériser la valeur de ses paramètres théoriques non observables qui simule le mieux la réalité observée. 3) Tirer au sort chaque réalisation de la variable. 4) Évaluer le modèle avec les variables tirées au sort. 5) La moyenne des sorties du modèle approche leur espérance lorsque le nombre de tirages est grand. En réalisant B échantillons sur les couples coût ; années de vie sauvées ajustées ou non sur la qualité de vie sur la base de ces principes, il est possible de construire B rapports différentiels coût/résultat qui peuvent être représentés dans le graphique classique à quatre quadrants.

La simulation de Monte-Carlo de troisième ordre consiste en un mélange de la première et de la seconde stratégie. Le cheminement d'un patient sera déterminé comme lors d'une simulation de premier ordre, mais à partir de probabilités tirées au hasard comme lors d'une simulation de second ordre.

8 Caractérisation de l'incertitude finale

8.1 Courbe retraçant les proportions de cas gagnants (CPCG)

Le rapport différentiel coût-résultat (RDCR) est le critère d'intérêt dans toute analyse coût-efficacité confrontant différents traitements. Le choix d'un traitement par rapport à un autre est conditionné par trois inconnues : différentiel de résultats (ΔR), différentiel de coûts (ΔC) et la valeur de l'effort financier socialement acceptable (EFSA). Les valeurs des deux premiers paramètres dépendent du modèle élaboré et de sa fiabilité. L'effort jugé socialement acceptable par la société peut être considéré comme une donnée exogène, sur laquelle l'évaluateur n'a pas à se prononcer. Plusieurs méthodes permettent de représenter l'incertitude liée à l'estimation ponctuelle du rapport. L'une d'entre elles consiste à construire une courbe représentant la proportion des résultats obtenus au décours de la simulation numérique qui sont efficaces par rapport à leur comparateur, pour un niveau donné de l'effort social consenti. Cette courbe est communément désignée sous le nom de courbe d'acceptabilité on peut la désigner de façon plus informelle, sous le nom de « courbe des proportions de cas gagnants ».

Pour les décisions concernant seulement deux traitements, la CPCG est tracée en général que pour le nouveau traitement. Cependant, la CPCG de l'alternative thérapeutique à laquelle il a été comparé peut-être présentée. Puisque les traitements sont mutuellement exclusifs et collectivement exhaustifs, la courbe d'acceptabilité des résultats du comparateur est de forme et de localisation opposée. Les deux courbes se croisent lorsque les probabilités d'efficacité sont égales.

La courbe d'acceptabilité des résultats permet d'identifier l'incertitude sur l'état actuel des connaissances et les données pour lesquels l'incertitude devrait être réduite, mais cette mesure ne fournit pas d'information sur l'impact des incertitudes sur un indicateur de gestion clair en fonction duquel la décision pourrait être prise. Au niveau décisionnel, c'est la mesure du bénéfice monétaire net qui permet de mesurer l'impact des données d'entrée sur le critère de résultat déterminant le choix.

8.2 Frontière des choix qui maximisent le bénéfice de santé net (FdCMax)

Lorsque les décideurs se proposent de maximiser l'état de santé de la population, la décision doit être prise en choisissant, parmi toutes les options thérapeutiques disponibles, celle qui maximise le bénéfice net de santé publique, quelles que soient les chances de faire le bon choix ou le risque de se tromper (Barton et al., 2008). L'intervention optimale pour un niveau d'effort financier socialement acceptable donné est celle qui maximise le bénéfice de santé net entendu comme la différence moyenne entre le bénéfice clinique attendu valorisé sur la base de l'effort social jugé acceptable et le coût qui lui est associé ($BNS = \lambda * B - C$). C'est en effet, la seule intervention qui, comparée à toutes les autres interventions, dégage par rapport à chacune d'entre elles un bénéfice différentiel de santé net positif ($BDNS \geq 0$).

Mille échantillons sont construits par itération en effectuant des tirages successifs à partir des distributions de probabilité des différents paramètres du modèle. Le nombre de fois où un traitement dégage un bénéfice de santé net positif par rapport à ses comparateurs sur l'ensemble des simulations effectuées, pour une valeur donnée du niveau d'effort financier socialement acceptable, est calculé. Cette opération est répétée pour toutes les valeurs de l'effort financier socialement acceptable par année de vie gagnée. Les traitements étant mutuellement exclusifs et collectivement exhaustifs, les courbes qui représentent leur probabilité d'être efficace par rapport à leurs comparateurs se superposent verticalement les unes par rapport aux autres et leur sommation verticale est égale à l'unité.

Ces courbes permettent de visualiser pour une valeur donnée de lambda (λ) quel est le traitement dont le bénéfice de santé net est supérieur à celui de tous les autres. La plage de ces valeurs de λ , pour lesquelles

le bénéfice net moyen d'un traitement est maximal et les «points de retournement » (i.e. quand il y a un changement dans le choix optimal) correspondants aux RDCR des différents traitements les uns par rapport aux autres, peuvent être isolés (Fenwick et al., 2004). La borne inférieure de la fourchette de valeur pour laquelle un traitement est considéré comme optimal correspond au RDCR de ce traitement, et la borne supérieure indique du RDCR du traitement qui lui est immédiatement supérieur en termes d'efficience.

Une fois que le traitement maximisant le bénéfice de santé net a été identifié pour chaque niveau de λ , la probabilité que ces traitements maximisent l'intérêt collectif peut alors être tracée (axe y) pour différentes valeurs de λ (axe x). Cette probabilité de maximiser l'intérêt de santé publique sans risquer de se tromper est représentée par la proportion des résultats qui permettent de maximiser la valeur marchande du bénéfice collectif associée à cette plage de valeurs. Son complément correspond à la probabilité de prendre une mauvaise décision.

Contrairement aux CPCG, la $FDCM_{ax}$ évalue la probabilité que le traitement soit efficace pour différentes valeurs de λ dans le seul cas où celles-ci permettent de maximiser simultanément le bénéfice net moyen. Les options optimales étant les seules options représentées, le tracé de la FPPE ne correspond pas toujours aux interventions pour lesquelles la probabilité d'être coût-efficace soit la plus élevée (Barton et al., 2008)

9 Expériences et compétences pour l'exécution du projet

9.1 Qui est REES France ?

REES France, créé en 1998, est un bureau d'études spécialisé dans l'évaluation des technologies et des actions de santé. La Société, dont le professeur Robert Launois assure la Direction scientifique, a pour objectif d'intensifier les liaisons entre la recherche, l'industrie et les milieux médicaux, en mettant à la disposition des professionnels de santé des techniques innovantes qui leur permettent de mieux maîtriser leur environnement. REES France a publié plus 250 articles consacrés à l'évaluation médico-économique ou aux systèmes de santé et réalisé plus de 70 études dans ces différents domaines. Son équipe de 2 personnes est composée d'un économiste, et d'un biostatisticien.

REES France met à la disposition des professionnels de santé des techniques d'analyse et de suivi innovantes. Il les accompagne de la genèse du projet d'étude à la communication des résultats dans les congrès ou les revues appropriées. Il propose des plans d'expérience randomisés ou observationnels, des études de qualité de vie, de satisfaction ou de mesure des préférences des acteurs de santé, des modèles médico-économiques d'aide à la décision ou de simulation d'impact budgétaire, des logiciels à l'usage des décideurs ou des médecins exploitant les résultats de ses travaux.

De nombreux outils lui permettent de répondre aux attentes de ses interlocuteurs : techniques statistiques avancées (bootstrap, simulation de Monte Carlo, propensity score, ...), modèles économiques, cliniques ou marketing (choix discrets, régressions linéaires généralisées, analyse conjointe...), outils informatiques (CRF électronique, programmation de didacticiels...).

Recherche et travaux récents :

- **Études observationnelles** : Programme de soutien aux Innovations diagnostiques et thérapeutiques coûteuses, Direction des Hôpitaux et de l'organisation des Soins (DHOS) : Protocole EMOCAR «Évaluation médico-économique de la surveillance par oxymétrie

cérébrale(INVOS™) en peropératoire lors d'endartériectomies carotidiennes (STIC 2010-2016) ; Protocole CECILE, Centre Hospitalier Universitaire de Poitiers (2010-2016) ; ministère de la Santé. Programme de Soutien aux Innovations, Diagnostics et Thérapeutiques Coûteuses : « Évaluation du traitement endoscopique de l'emphysème pulmonaire (TEEP) comme technique alternative à la chirurgie de réduction du volume pulmonaire » (STIC-DHOS 2007-2008) - Ministère de la Santé. Programme de Soutien aux Innovations, Diagnostics et Thérapeutiques Coûteuses : «Évaluation de la mise en place temporaire (6 mois) d'un ballon gastrique chez les patients présentant une obésité morbide sévère dans le cadre d'une prise en charge multidisciplinaire avant chirurgie bariatrique » (STIC-DHOS 2007-2008) ;évaluation du Réseau MORPHÉE (DRDR-IdF 2006-2007) ; Office Parlementaire des Politiques de Santé Assemblée Nationale-Sénat : « Prise en charge précoce des accidents vasculaires cérébraux »(OPEPS 2006-2007). Institut National du Cancer : « Évaluation du coût du dépistage organisé du cancer colorectal et du cancer du sein » (INCA 2006-2007) ; Office Français des Drogues et Toxicomanies. « Évaluation du Réseau des soins en microstructures en Alsace (OFDT 2006-2007) ; ICEREA (14 services de réanimation PHRC 2006-2008) Evaluation du Réseau MORPHEE (FAQSV 2004-2006) ; INCA Évaluation du coût du dépistage organisé du cancer colorectal et du cancer du sein (2006) ; Programme de soutien aux Innovations diagnostiques et thérapeutiques coûteuses, Direction des Hôpitaux et de l'organisation des Soins (DHOS) : PREMISS (SFAR, SRLF, 85 services de réanimation,1096 patients inclus 2002-2006) ; Programme de soutien aux Innovations diagnostiques et thérapeutiques coûteuses, Direction des Hôpitaux et de l'organisation des Soins (DHOS) : HER.ME.S (AP-HP, CRLCC, 12 services d'oncologie 2001 – 2005) ; Evaluation du Réseau Asthme RESALIS (CPAM. Évreux, Alliance Médica 1998- 2002) ; Evaluation des Réseaux Groupama Partenaire Santé (Mutualité Sociale Agricole, Groupama 1999-2001) ; Conception et Développement de Cahier d'Observation en ligne, Contrôle qualité, Analyse statistique, Rédaction du rapport de fin de l'étude.

- **Méta-analyses décisionnelles dans de nombreuses indications** : Cancer du sein métastatique, cancer du poumon métastatique, schizophrénie, AVC et unités neurovasculaires, cancer de la prostate, dépistage du cancer colorectal, etc. (voir site www.rees-france.com). Modèles d'impact budgétaire de la mise sur le marché de nouvelles molécules (schizophrénie, antiémétiques dans les chimiothérapies fortement et moyennement émétisantes).
- **Méta-analyse multitraitement bayésienne** : Cancer du poumon, cancer du sein, cancer colorectal, polyarthrite rhumatoïde.

Dans le domaine de la santé, REES travaille pour les principaux acteurs du secteur : des instances nationales comme le Ministère de la Santé, la HAS, ou l'INCa, ainsi que les plus grands industriels de la pharmacie et des dispositifs médicaux, ou encore les acteurs des biotechnologies (voir Annexe).

REES dispose en particulier de l'expérience et des compétences suivantes : économie / financement des Systèmes de Santé, méthodologie d'évaluation de programmes (voir références dans la partie suivante).

REES possède des compétences statistiques et de modélisation éprouvées (DEA Biostatistiques, ENSAI ou Mastères ENSAE) déjà mises à profit sur des projets de synthèse des données médicales et économiques dans le cadre d'analyses observationnelles prospectives de longue durée.

9.2 Expertise des intervenants en matière médico-économique

1. Robert LAUNOIS, Directeur scientifique REES France

Robert LAUNOIS a fait ses études à Rennes, Paris et Cambridge Mass. Agrégé des facultés de Sciences Economiques, diplômé de l'Institut d'Études Politiques de Paris et Harkness Fellow du Commonwealth

Fund of New York (Harvard University), il enseigne l'évaluation des technologies de santé dans les masters de l'Université René Descartes (Paris V), Pierre et Marie Curie (Paris VI) et de l'université du Mans , après avoir été en poste à la Faculté de Médecine de l'Université de Paris XIII et à la faculté de sciences économiques de Rennes I. Fondateur et Directeur scientifique du Réseau d'Évaluation en Économie de la Santé (REES France) depuis 1998, il est consultant pour le compte des administrations et des industries de santé. Ses recherches portent sur l'étude des stratégies diagnostiques et thérapeutiques, l'évaluation pharmaco-économique. Membre du comité de rédaction du Journal Health Economics (1981-1990), membre de la commission de la Nomenclature Générale des Actes Professionnels (NGAP Ministère de la Santé 1991-1995), Rapporteur puis Président du chantier santé de l'Institut La Boétie, membre de la commission de surveillance de l'hôpital Laennec (Assistance Publique des Hôpitaux de Paris 1994-1999), expert économique auprès du comité de co-pilotage de la classification commune des actes médicaux (Caisse Nationale d'Assurance Maladie des travailleurs salariés 2004), membre du Fond d'Amélioration de la Qualité des Soins de Ville Ile de France (FAQSV 2003-2006), membre de la commission des Comptes de la Santé (1996-2004). Expert auprès de la HAS (2004 -2016) Validateur auprès du Centre Fédéral d'Expertise des Soins de Santé en Belgique (2008), expert auprès du National Institute for Health Research Health Technology Assessment programme (2008-2010) il est Président de la Société Française d'Économie de la Santé (SFES) depuis 2012. Robert Launois a réalisé en 2010, une étude pour la HAS dédiée au dépistage de la rétinopathie diabétique et pour la Direction générale de la santé du ministère de la Santé et des Sports (DGS), un modèle sur le dépistage de la surdité néonatale et en 2012 une étude pour l'institut national du cancer (INCA) sur le coût du dépistage organisé du cancer colorectal et de son évolution. Robert Launois coordonne actuellement la partie économique de l'essai randomisé EMOCAR dans le cadre du programme de recherches medico économiques (PRME) de la Direction Générale de l'organisation des soins(DGOS) du ministère de la Santé et des Sports. Il développe des programmes innovants de Méta-analyse multitraitements en Cancérologie et dans les maladies chroniques. Il est lauréat du prix Winthrop Européen d'Économie de la Santé (1990), du prix d'Économie de la Santé de l'association créée (2003),

Il a publié plus de 250 articles dans les journaux professionnels économiques ou médicaux à comité de lecture.

2. Anastasiia KABESHOVA, REES France

Ingénieure biostatisticienne à REES France, Anastasiia Kabeshova s'occupe de développement et adaptation des différents types de modèles de décision économiques en santé, de mise en oeuvre de techniques statistiques et probabilistes (simulations MCMC, bootstrap, analyses de sensibilité probabiliste et déterministe), de la rédaction des plans d'analyse et des rapports critiques, et de préparation des articles et des résumés pour les publications. Elle apportera ses compétences techniques dans la modélisation des séquences des biothérapies dans la prise en charge de la polyarthrite rhumatoïde. La maîtrise des logiciels statistiques et des outils de simulation de Mme Kabeshova permettra d'effectuer le développement efficace du modèle choisi.

Formation :

Désignation du(des) diplôme(s)	Niveau du (des) diplôme(s)	Année d'obtention Du (des) diplôme(s)	Établissement ayant délivré le (les) diplôme(s)
Thèse en Biostatistiques et Mathématiques Appliquées à la Santé	Bac + 8	2015	Université d'Angers (France)
Master de Mathématiques Appliquées	Bac + 5	2013	Université Nationale de Dnipropetrovsk (Ukraine)
Master de Statistique et Économie Appliquées à la Santé	Bac + 5	2012	Université du Maine, Le Mans (France)

Expériences :

<u>Objet de la (des) mission(s)</u>	<u>Date de commencement</u>	<u>Durée de la (des) expérience(s)</u>
N°1) Etude Médico-Economique dans la prise en charge des patients âgés présentant un Myélome Multiple non éligibles à la greffe en 1ère ligne de traitement <ul style="list-style-type: none"> - Évaluation médico-économique - Modélisation séquentielle - Analyse coût-efficacité - Réalisation d'analyse de sensibilité 	01/10/2014	12 mois
N°2) Etude Médico-Economique dans la Prise en Charge de l'Encéphalopathie Hépatique <ul style="list-style-type: none"> - Adaptation du modèle coût-utilité pour le marché français - Réalisation d'analyse de sensibilité - Rédaction d'article pour la publication dans une revue médicale 	05/10/2015	6 mois
N°3) Évaluation médico-économique dans le traitement des formes avancées et sévères de maladie de parkinson <ul style="list-style-type: none"> - Rédaction du plan d'analyse - Analyses intermédiaires de la base de données - Imputation des données manquantes - Évaluation médico-économique 	1/02/2016	7 mois
N°4) Évaluation multidimensionnelle du projet de télémédecine <ul style="list-style-type: none"> - Rédaction du plan d'analyse - Réalisation d'analyse médico-économique 	5/10/2015	12 mois
N°5) Étude sur le Syndrome des Jambes Sans Repos (SJSR) Analyse critique de la méthodologie et l'exploitation des résultats des analyses transmises	6/07/2015	3 mois

3. Soumaia BEN HARIZ, REES France

Titulaire d'un master en économie de la santé (Université Paris Est Créteil), Soumaia Ben Hariz a rejoint Rees France en 2014 en tant qu'économiste de la santé-statisticienne et contribue depuis aussi bien à la réalisation des études d'évaluation médico-économique (analyse coût-efficacité et analyse d'impact budgétaire, TreeAge, Excel) qu'à la réalisation de synthèses de la littérature médico-économique et des méta-analyses (ReviewManager, WinGUGS, R).

Également chargée en tant que data-manager de gérer la qualité des données pour deux essais randomisés multi centrés conduits en France, elle dispose de forte compétence en programmation SAS.

Son expérience à la Caisse Primaire de l'Assurance Maladie du Val d'Oise lui a permis d'acquérir une bonne connaissance du fonctionnement du système de santé et des bases de données notamment le SIAM ERASME et le PMSI lui permettant de disposer d'une compétence toute particulière dans l'analyse des bases de données médico-administratives françaises et une solide capacité d'analyse ainsi qu'une bonne connaissance des techniques de traitement de l'information appliquées au domaine de la santé.

Formation :

<u>Désignation du(des) diplôme(s)</u>	<u>Niveau du (des) diplôme(s)</u>	<u>Année d'obtention Du (des) diplôme(s)</u>	<u>Établissement ayant délivré le (les) diplôme(s)</u>
Approches de modélisation pour l'évaluation des	Certification	2014	UMIT, Institute of Public Health, Medical Decision Making and Health

technologies de la santé (HTA)			Technology Assessment (Anvers, Belgique)
Master en économie de la santé	Bac + 5	2013	Université Paris Est Créteil (UPEC)

Expériences :

<u>Objet de la (des) mission(s)</u>	<u>Date de commencement</u>	<u>Durée de la (des) expérience(s)</u>
<p>N°1) Etude Médico-Economique dans la prise en charge des patients âgés présentant un Myélome Multiple non éligibles à la greffe en 1ère ligne de traitement</p> <ul style="list-style-type: none"> - Modèle d'impact budgétaire (BIM) - Réalisation d'analyse de sensibilité - Méta analyse bayésienne en réseau pour l'analyse coût-efficacité (ACE) - Réalisation de synthèses de la littérature (données cliniques et économiques) - Rédaction des rapports techniques de BIM et ACE 	01/10/2014	12 mois
<p>N°2) Etude Médico-Economique dans la Prise en Charge de l'Encéphalopathie Hépatique</p> <ul style="list-style-type: none"> - Adaptation du modèle coût-utilité pour le marché français - Réalisation d'analyse de sensibilité - Rédaction d'article pour la publication dans une revue médicale 	02/01/2015	8 mois
<p>N°3) Essais randomisés multi centriques conduits en France</p> <ul style="list-style-type: none"> - Évaluation médico-économique dans l'endartériectomie carotidienne - Rédaction du plan d'analyse - Imputation des données manquantes - Réalisation d'un mapping du questionnaire SF-36 vers EQ-5D-3L - Analyses intermédiaires de la base de données - Préparation des présentations des résultats intermédiaires 	02/04/2014	24 mois
<p>N°4) Essais randomisés multi centriques conduits en France</p> <ul style="list-style-type: none"> - Évaluation médico-économique dans le traitement des formes avancées et sévères de maladie de parkinson - Data management : gestion de queries, tests de cohérence, préparation de monitoring 	07/07/2014	21 mois

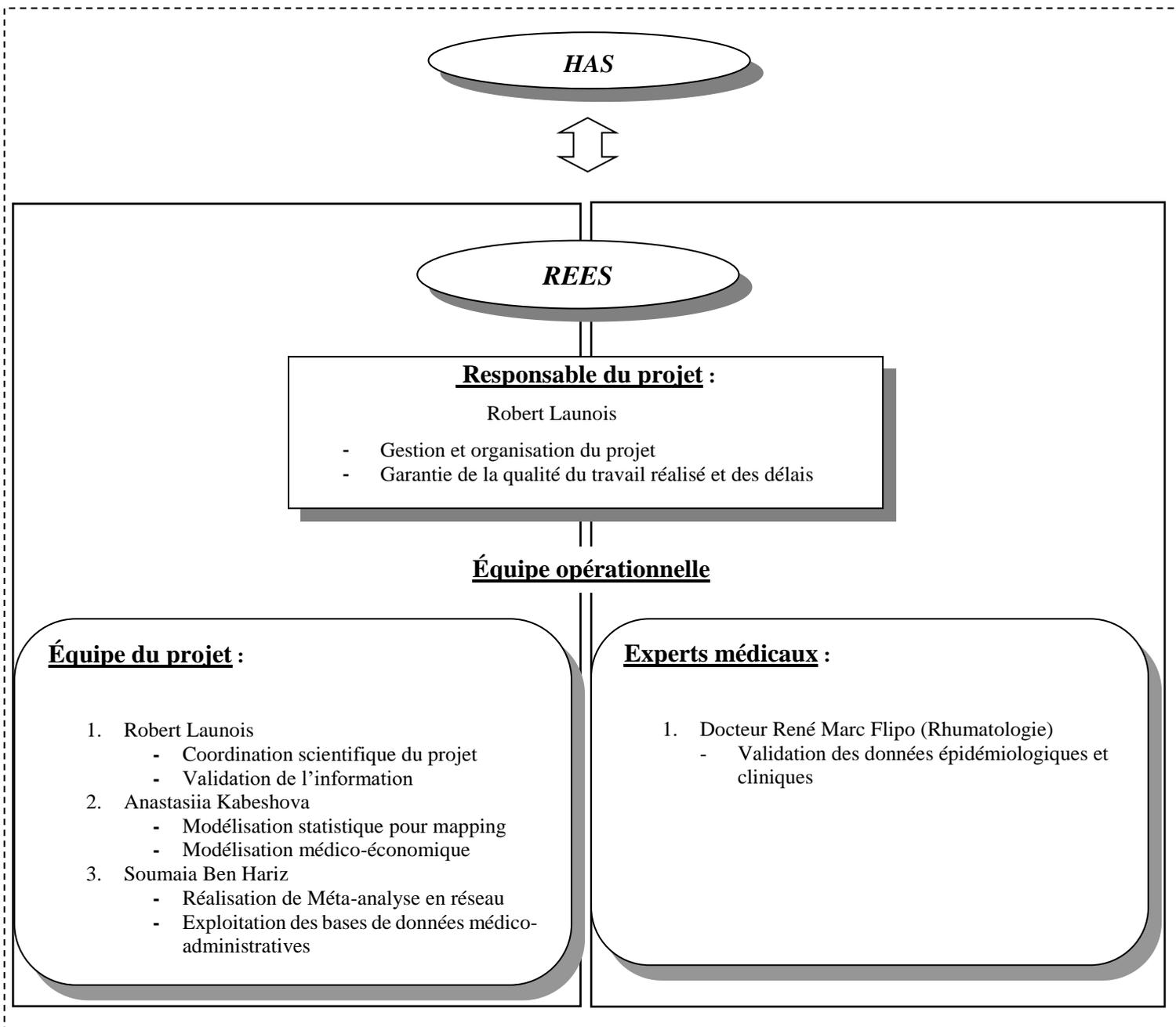
10 Calendrier prévisionnel

PHASES	DESCRIPTION	ÉCHÉANCE
	<ul style="list-style-type: none"> - Notification du marché - Réunion de lancement avec le Titulaire 	Mai 2016
Phase 1	<ul style="list-style-type: none"> - État des lieux de l'identification de la littérature portant sur le type de modèle économique - Présentation et validation des choix méthodologiques - Développement du plan d'analyse économique et validation du plan d'analyse économique 	Juillet 2016
Phase 2	<ul style="list-style-type: none"> - Réalisation des revues systématiques de données, réalisation de la méta-analyse en réseau 	Décembre 2016

	<ul style="list-style-type: none"> - Validation des données portant sur tous les paramètres du modèle : cliniques, coûts et qualité de vie (utilités) - Paramétrage du modèle - Production des premiers résultats du modèle 	
Phase 3	- Correction des anomalies du modèle, validation finale des résultats du modèle, présentation des résultats et rédaction du rapport technique par le Titulaire	Janvier 2017
	- Présentation des résultats de l'analyse en groupe de travail et en sous-commission économie	Janvier-Février 2017
	- Analyses de sensibilité/analyses complémentaires	Février 2017
	- Validation du rapport technique du Titulaire	Février-Mars 2017
Phase finale propre au projet HAS	- Rédaction du rapport final HAS (prise en compte des commentaires du groupe de travail) / Élaboration d'une synthèse à intégrer directement dans le rapport	Février-Mars 2017
	- Passage en CEESP	Mars 2017
	- Intégration des modifications éventuelles suite au passage en CEESP	Mars-Avril 2017
	- Passage en Collège - Intégration des modifications éventuelles suite au passage en Collège	Avril 2017
Publication	- Publication de l'évaluation économique dans une revue clinique disposant d'un comité de lecture et/ou une revue d'économie de la santé	Avril-Juin 2017

11 Déroulement et organisation de la mission

11.1 Organisation



REES met en place une équipe qui réunit plusieurs compétences clés pour le succès de cette mission :

- des compétences d'économie de la Santé
- des compétences médicales et biologiques
- des compétences d'analyse statistique

11.2 Démarche qualité chez REES France

La démarche qualité chez REES intervient à plusieurs étapes du déroulement d'une mission :

- Dès le lancement de la mission, la qualité est assurée par la bonne répartition des tâches entre les consultants avec une définition précise des timing associés à chaque tâche et des livrables associés. Cette répartition est orchestrée par le responsable de mission.
- Chaque tâche est ensuite suivie avec soin dans sa réalisation avec un échange deux à trois fois par semaine entre les consultants et le responsable de mission.
- Par ailleurs, des échanges téléphoniques sont régulièrement organisés avec la HAS afin qu’il soit tenu régulièrement au courant de l’avancée des travaux.
- À l’échéance de chaque tâche, les livrables sont discutés et retravaillés entre le coordinateur et l’équipe opérationnelle (responsable de mission et experts), puis consolidés entre eux.

Les résultats consolidés d’un bloc sont ensuite présentés à un Comité interne composé :

- de l’équipe opérationnelle ;
- du comité d’expertise composé des experts associés à la réalisation de la mission.

Suite à ce Comité, des modifications sont apportées au livrable avant que celui-ci soit présenté au client.

12 References bibliographiques

- Aletaha, D., Neogi, T., Silman, A.J., Funovits, J., Felson, D.T., Bingham, C.O., Birnbaum, N.S., Burmester, G.R., Bykerk, V.P., Cohen, M.D., Combe, B., Costenbader, K.H., Dougados, M., Emery, P., Ferraccioli, G., Hazes, J.M.W., Hobbs, K., Huizinga, T.W.J., Kavanaugh, A., Kay, J., Kvien, T.K., Laing, T., Mease, P., Ménard, H.A., Moreland, L.W., Naden, R.L., Pincus, T., Smolen, J.S., Stanislawska-Biernat, E., Symmons, D., Tak, P.P., Upchurch, K.S., Vencovský, J., Wolfe, F., Hawker, G., 2010. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* 62, 2569–2581. doi:10.1002/art.27584
- Anderson, J., Caplan, L., Yazdany, J., Robbins, M.L., Neogi, T., Michaud, K., Saag, K.G., O'Dell, J.R., Kazi, S., 2012. Rheumatoid arthritis disease activity measures: American College of Rheumatology recommendations for use in clinical practice. *Arthritis Care Res.* 64, 640–647. doi:10.1002/acr.21649
- Arnett, F.C., Edworthy, S.M., Bloch, D.A., McShane, D.J., Fries, J.F., Cooper, N.S., Healey, L.A., Kaplan, S.R., Liang, M.H., Luthra, H.S., 1988. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum.* 31, 315–324.
- Barton, P., Bryan, S., Robinson, S., 2004. Modelling in the economic evaluation of health care: selecting the appropriate approach. *J. Health Serv. Res. Policy* 9, 110–118. doi:10.1258/135581904322987535
- Briggs, A.H., Claxton, K., Sculpher, M.J., 2006. *Decision Modelling for Health Economic Evaluation*. Oxford University Press.
- Chevalier, J., 2010. *Mesure de l'utilité attachée aux états de santé. Valorisation de l'index d'utilité EQ-5D-3L et évolution de l'échelle actuelle en France*. (Thèse de Doctorat en Sciences Economiques, Université Paris IX Dauphine, France).
- Choi, H.K., Seeger, J.D., Kuntz, K.M., 2002. A cost effectiveness analysis of treatment options for methotrexate-naïve rheumatoid arthritis. *J. Rheumatol.* 29, 1156–1165.
- Durieux S, Saroux A, Guillemin F, 2003. Critères d'activité et de sévérité de la polyarthrite rhumatoïde : proposition d'un groupe d'experts rhumatologues français, élaborée par la méthode Delphi. *Communication orale O.85*.
- Edwards, J.C., Cambridge, G., 2001. Sustained improvement in rheumatoid arthritis following a protocol designed to deplete B lymphocytes. *Rheumatol. Oxf. Engl.* 40, 205–211.
- Efficacy of B-Cell-Targeted Therapy with Rituximab in Patients with Rheumatoid Arthritis — NEJM [WWW Document], n.d. URL <http://www.nejm.org/doi/full/10.1056/NEJMoa032534> (accessed 4.8.16).
- EMA, 2005. Guideline on the choice of the non-inferiority margin.
- Felson, D.T., Anderson, J.J., Boers, M., Bombardier, C., Chernoff, M., Fried, B., Furst, D., Goldsmith, C., Kieszak, S., Lightfoot, R., Paulus, H., Tugwell, P., Weinblatt, M., Widmark, R., James Williams, H., Wolfe, F., 1993. The American college of rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. *Arthritis Rheum.* 36, 729–740. doi:10.1002/art.1780360601
- Fries, J.F., Spitz, P., Kraines, R.G., Holman, H.R., 1980. Measurement of patient outcome in arthritis. *Arthritis Rheum.* 23, 137–145.
- Gaujoux-Viala, C., Gossec, L., Cantagrel, A., Dougados, M., Fautrel, B., Mariette, X., Nataf, H., Saroux, A., Trope, S., Combe, B., 2014. Recommandations de la Société française de rhumatologie pour la prise en charge de la polyarthrite rhumatoïde. *Rev. Rhum.* 81, 303–312. doi:10.1016/j.rhum.2014.04.009
- Gray, A.M., Rivero-Arias, O., Clarke, P.M., 2006. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Med. Decis. Mak. Int. J. Soc. Med. Decis. Mak.* 26, 18–29. doi:10.1177/0272989X05284108
- Guillemin, F., Brainçon, S., Pourel, J., 1991. *Mesure de la capacité fonctionnelle dans la polyarthrite rhumatoïde: Adaptation Française du Health Assessment Questionnaire (HAQ)*. (Measurement of

- functional capacity in rheumatoid arthritis: French adaptation of the Health Assessment Questionnaire (HAQ). *Rev. Rhum. Mal. Ostéo-Articul.* 58, 459–65.
- Guillemin, F., Briançon, S., Klein, J.M., Sauleau, E., Pourel, J., 1994. Low incidence of rheumatoid arthritis in France. *Scand. J. Rheumatol.* 23, 264–268.
- Guillemin, F., Saraux, A., Guggenbuhl, P., Roux, C.H., Fardellone, P., Le Bihan, E., Cantagrel, A., Chary-Valckenaere, I., Euller-Ziegler, L., Flipo, R.-M., Juvin, R., Behier, J.-M., Fautrel, B., Masson, C., Coste, J., 2005. Prevalence of rheumatoid arthritis in France: 2001. *Ann. Rheum. Dis.* 64, 1427–1430. doi:10.1136/ard.2004.029199
- Hansen, K.E., Hildebrand, J.P., Genovese, M.C., Cush, J.J., Patel, S., Cooley, D.A., Cohen, S.B., Gangnon, R.E., Schiff, M.H., 2004. The efficacy of switching from etanercept to infliximab in patients with rheumatoid arthritis. *J. Rheumatol.* 31, 1098–1102.
- HAS (Haute Autorité de Santé), 2006a. Métoject : avis de la commission de la transparence.
- HAS (Haute Autorité de Santé), 2006b. Mabthera : avis de la commission de la transparence.
- HAS (Haute Autorité de Santé), 2005. Enbrel : avis de la commission de la transparence.
- Hernández Alava, M., Wailoo, A.J., Ara, R., 2012. Tails from the Peak District: Adjusted Limited Dependent Variable Mixture Models of EQ-5D Questionnaire Health State Utility Values. *Value Health* 15, 550–561. doi:10.1016/j.jval.2011.12.014
- Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* 21, 1539–1558. doi:10.1002/sim.1186
- Karnon, J., 2003. Alternative decision modelling techniques for the evaluation of health care technologies: Markov processes versus discrete event simulation. *Health Econ.* 12, 837–848. doi:10.1002/hec.770
- Kass, R., Raftery, A., 1995. Bayes Factor. *J. Am. Stat. Assoc.* 773–795.
- Kremer, J.M., Westhovens, R., Leon, M., Di Giorgio, E., Alten, R., Steinfeld, S., Russell, A., Dougados, M., Emery, P., Nuamah, I.F., Williams, G.R., Becker, J.-C., Hagerty, D.T., Moreland, L.W., 2003. Treatment of rheumatoid arthritis by selective inhibition of T-cell activation with fusion protein CTLA4Ig. *N. Engl. J. Med.* 349, 1907–1915. doi:10.1056/NEJMoa035075
- Leandro, M., Edwards, J., Cambridge, G., 2002. Clinical outcome in 22 patients with rheumatoid arthritis treated with B lymphocyte depletion. *Ann. Rheum. Dis.* 61, 883–888. doi:10.1136/ard.61.10.883
- Longworth, L., Rowen, D., 2011. NICE DSU technical support document 10: the use of mapping methods to estimate health state utility values. Sheff. UK Decis. Support Unit ScHARR Univ. Sheff. b4.
- Mariette, X., 2004. Biothérapies émergentes dans la polyarthrite rhumatoïde. *Rev. Rhum.* 71, 848–853. doi:10.1016/j.rhum.2004.07.012
- Piaggio, G., Elbourne, D.R., Altman, D.G., Pocock, S.J., Evans, S.J.W., CONSORT Group, 2006. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA* 295, 1152–1160. doi:10.1001/jama.295.10.1152
- Proskorovsky, I., Lewis, P., Williams, C.D., Jordan, K., Kyriakou, C., Ishak, J., Davies, F.E., 2014. Mapping EORTC QLQ-C30 and QLQ-MY20 to EQ-5D in patients with multiple myeloma. *Health Qual Life Outcomes* 12, 35. doi:10.1186/1477-7525-12-35
- Riise, T., Jacobsen, B.K., Gran, J.T., Haga, H.J., Arnesen, E., 2001. Total mortality is increased in rheumatoid arthritis. A 17-year prospective study. *Clin. Rheumatol.* 20, 123–127.
- Rothmann, M.D., Tsou, H.-H., 2003. On non-inferiority analysis based on delta-method confidence intervals. *J. Biopharm. Stat.* 13, 565–583. doi:10.1081/BIP-120022775
- Roux, C.H., Saraux, A., Le Bihan, E., Fardellone, P., Guggenbuhl, P., Fautrel, B., Masson, C., Chary-Valckenaere, I., Cantagrel, A., Juvin, R., Flipo, R.-M., Euller-Ziegler, L., Coste, J., Guillemin, F., 2007. Rheumatoid arthritis and spondyloarthropathies: geographical variations in prevalence in France. *J. Rheumatol.* 34, 117–122.
- Rowen, D., Brazier, J., Roberts, J., 2009. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health Qual. Life Outcomes* 7, 27. doi:10.1186/1477-7525-7-27
- Scholz, S., Mittendorf, T., 2014. Modeling rheumatoid arthritis using different techniques - a review of model construction and results. *Health Econ. Rev.* 4, 18. doi:10.1186/s13561-014-0018-2
- Siebert, U., Alagoz, O., Bayoumi, A.M., Jahn, B., Owens, D.K., Cohen, D.J., Kuntz, K.M., ISPOR-SMDM Modeling Good Research Practices Task Force, 2012. State-transition modeling: a report

- of the ISPOR-SMDM Modeling Good Research Practices Task Force--3. *Value Health J. Int. Soc. Pharmacoeconomics Outcomes Res.* 15, 812–820. doi:10.1016/j.jval.2012.06.014
- Smolen, J.S., Landewé, R., Breedveld, F.C., Buch, M., Burmester, G., Dougados, M., Emery, P., Gaujoux-Viala, C., Gossec, L., Nam, J., Ramiro, S., Winthrop, K., Wit, M. de, Aletaha, D., Betteridge, N., Bijlsma, J.W.J., Boers, M., Buttgereit, F., Combe, B., Cutolo, M., Damjanov, N., Hazes, J.M.W., Kouloumas, M., Kvien, T.K., Mariette, X., Pavelka, K., Riel, P.L.C.M. van, Rubbert-Roth, A., Scholte-Voshaar, M., Scott, D.L., Sokka-Isler, T., Wong, J.B., Heijde, D. van der, 2013. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. *Ann. Rheum. Dis.* annrheumdis–2013–204573. doi:10.1136/annrheumdis-2013-204573
- Snapinn, S.M., 2000. Noninferiority trials. *Curr. Control. Trials Cardiovasc. Med.* 1, 19–21. doi:10.1186/cvm-1-1-019
- Spiegelhalter DJ, Abrams KR., Myles JP., 2004. Bayesian approaches to clinical trials and health care evaluation. Chichester Hoboken NJ Wiley 55.
- van der Heijde, D.M., van 't Hof, M.A., van Riel, P.L., Theunisse, L.A., Lubberts, E.W., van Leeuwen, M.A., van Rijswijk, M.H., van de Putte, L.B., 1990. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann. Rheum. Dis.* 49, 916–920.
- Wiens, B.L., 2002. Choosing an equivalence limit for noninferiority or equivalence studies. *Control. Clin. Trials* 23, 2–14.

Annexe 1 : Realisation d'analyse de sensibilité probabiliste sous VBA

```

Option Private Module
Dim First_Row As Integer
Dim Last_Row As Integer

Dim Parameter_Name_Column As String
Dim Value_Column As String
Dim Default_Formula_Column As String
Dim Upper_Bound_Column As String
Dim Lower_Bound_Column As String
Dim PSA_Flag_Column As String
Dim RD_Flag_Column As String
Dim Sheet_Name_Column As String
Dim Location_Cell_Column As String
Dim Distribution_Shape_Column As String
Dim Formula_Copy_Column As String
Dim Data_Copy_Column As String
Dim Mean_Column As String
Dim SA_Flag_Column As String
Dim SA_Mean_Column As String
Dim SA_LB_Column As String
Dim SA_UB_Column As String

Dim Control_Panel_Sheet_Name As String
Dim PSA_Output_Sheet_Name As String
Dim PSA_Settings_Sheet_Name As String
Dim SA_Output_Sheet_Name As String
Dim SA_Settings_Sheet_Name As String

Dim StDev_Column As String
Dim Iterations As Double
Dim First_Iteration As Double
Dim Last_Iteration As Double

Dim Results_First_Row As Integer
Dim Results_Last_Row As Integer

Dim pi As Double
Dim Run_Map_Control_Panel As Boolean

Sub PSA()
'~~~~~
' Purpose: Run probabilistic analysis
' Notes: If an error occurs in a distribution calculation, the macro recalculates
'        the trial and checks again. An infrequently occurring error (e.g., <100 times)
'        may not be significant to address but if the macro observes a frequently occurring
'        error, it will exit the macro and return an error msg in the status bar
'~~~~~

Dim x As Integer, trial As Integer, a As Integer, infiniteX As Integer, i As Integer
Dim goodTrials As Integer
Dim rge As Range
Dim bErrorInTrial As Boolean 'new
Dim FlagNetBen As Boolean
Dim iterNumber As Integer

Application.ScreenUpdating = False
FlagNetBen = False
Range("bPSA").Value = True

```

```

goodTrials = 0 'used to track the trials without sampling errors
infiniteX = 0 'Used to track an error in trial result

CurrentSimNumber = -1000000000 'Number used to set the seed. The number was chosen to
avoid using a seed that is used for the other random numbers
Call GenerateRandomNumbersPSA

iterNumber = 1000

For i = 0 To iterNumber '1099

    Range("PSA.Randcolumn").Value = Application.Index(RandomFullPSA, , i + 1) 'Puts random number
generation formula into the defined range

    For Each rge In Range("macroPsaTrial") 'Test trial for any errors
        If Application.WorksheetFunction.IsError(rge) Then
            bErrorInTrial = True
            Exit For
        End If
    Next rge
    If bErrorInTrial = False Then 'If there is no error in trial then save trial result
        Range(Range("macroPsaStart").Offset(goodTrials, 1), Range("macroPsaStart").Offset(goodTrials, 9)).Value
= Range("macroPsaTrial").Value
        Application.StatusBar = "Simulation: " & Round(goodTrials / 1000 * 100) & "% complete"
        goodTrials = goodTrials + 1
        If goodTrials = 1000 Then Exit For 'The required trials have been reached, then stop the loop

    Else 'There is an error in trial then update the number of errors
        infiniteX = infiniteX + 1
        If infiniteX > 100 Then 'If error in trial persists for more than 100x, then exit sub and report error
            Application.StatusBar = "Error found in PSA parameter. Please check your distributions."
            GoTo InfiniteLoopExit
        End If
        bErrorInTrial = False 'Reset bErrorInTrial tracker
    End If
    Call UpdateProgressBar(Round(((goodTrials - 1) / 1000) * 100) / 100, False)
Next i

'Net benefit

'Saves the outcome used on Net Benefit calculations
If Range("PSA.Outcome").Value = "QALYs" Then
    Sheets("PSA Results").Range("D86").Value = "QALY"
Else
    Sheets("PSA Results").Range("D86").Value = "Life-year"
End If

For i = 0 To 39

    Range("macroLam").Value = Range("macroLamValue").Offset(i, 0).Value
    Range("macroLamValue").Offset(i, 1).Resize(1, 3).Value = Range("macroPercentCE").Value
    Application.StatusBar = "Net Benefit Calculation: " & i + 1 & "% complete"
    Call UpdateProgressBar(Round((i / 100) * 100) / 100, True)
Next i

For i = 1 To 3
    Sheets("PSA Output").Range("macroPercentCE").Cells(1, i).EntireColumn.Hidden = Not (Sheets("PSA
Output").Range("PSA.netben.inclusion").Cells(1, i).Value)
Next i

Application.StatusBar = False

```

```
InfiniteLoopExit: 'goto used only if there is an infinite loop in PSA trials  
Range("bPSA").Value = False  
Sheets("PSA Results").Select  
Application.CutCopyMode = False
```

```
End Sub
```