



**HAL**  
open science

# Apport des méthodes méta-épidémiologiques pour la quantification du biais d'évaluation dans les études observationnelles : exemple avec les anticoagulants oraux directs

Maissa Safieddine

## ► To cite this version:

Maissa Safieddine. Apport des méthodes méta-épidémiologiques pour la quantification du biais d'évaluation dans les études observationnelles : exemple avec les anticoagulants oraux directs. Santé publique et épidémiologie. Université de Lyon, 2020. Français. NNT : 2020LYSES038 . tel-03266028

**HAL Id: tel-03266028**

**<https://theses.hal.science/tel-03266028>**

Submitted on 21 Jun 2021

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

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



N° d'ordre NNT : 2020LYSES038

**THESE de DOCTORAT DE L'UNIVERSITE DE LYON**

opérée au sein de

**UNIVERSITÉ JEAN MONNET - FACULTÉ DE MÉDECINE DE  
SAINT ETIENNE**

**Ecole Doctorale 488  
Sciences Ingénierie, Santé**

**Spécialité de doctorat :  
Discipline : Sciences de la vie**

Soutenue publiquement le 27/10/2020, par :  
**Maissa Safieddine**

---

**Apport des méthodes méta-  
épidémiologiques pour la quantification du  
biais d'évaluation dans les études  
observationnelles : exemple avec les  
anticoagulants oraux directs**

---

Devant le jury composé de :

Madame le Dr CORNU Catherine, PH, CHU de Lyon

Présidente

Madame le Pr LACUT Karine, PUPH, CHRU de Brest

Rapporteur

Monsieur le Dr SALVO Francesco, MCUPH, CHU de Bordeaux

Rapporteur

Monsieur le Pr MISMETTI Patrick, PUPH, CHU de Saint Etienne

Examineur

Monsieur le Dr FERDYNUS Cyril, PH, CHU de la Réunion

Examineur

Monsieur le Pr DELAVENNE Xavier, PUPH, CHU de Saint Etienne

Examineur

Madame le Dr LAPORTE Silvy, MCU-PH, CHU de Saint Etienne

Directrice de thèse

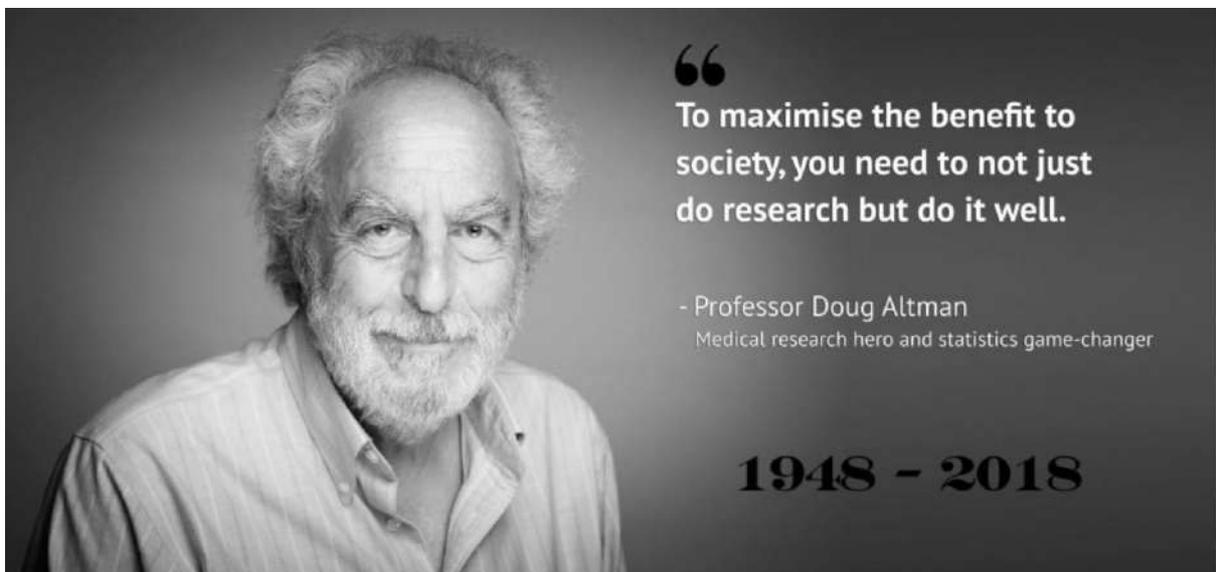
Monsieur le Pr CUCHERAT Michel, PUPH, CHU de Lyon

Co-directeur de thèse

# Avant-propos

Cette thèse a été préparée au sein de l'équipe INSERM DVH dirigée par le Pr Patrick Mismetti, équipe du laboratoire Sainbiose, UMR INSERM 1059.

Ce travail a été réalisé sous la direction du Dr Silvy Laporte, chef du service de l'Unité de Recherche Clinique Innovation et Pharmacologie (URCIP) au CHU de Saint-Etienne. Il a été co-dirigé par le Pr Michel Cucherat (UMR CNRS 5558 Laboratoire de Biométrie et Biologie Évolutive, équipe EMET). Cette thèse a été réalisée durant 4 ans (2016-2020). En 3<sup>ème</sup> année de thèse, j'ai été amenée à changer de département et suis partie m'installer sur l'Île la Réunion. Après avoir reçu l'avis favorable de ma directrice de thèse et de l'école doctorale de Saint-Étienne, j'ai été co-encadrée par le Dr Cyril Ferdynus dans le Centre d'Investigation Clinique (CIC) de la Réunion dirigé par le Pr Catherine Marimoutou.



# Remerciements

---

## **Aux membres du jury et aux rapporteurs de ma thèse**

Je remercie le Pr Karine Lacut et le Dr Francesco Salvo d'avoir accepté d'être rapporteurs de mon travail de thèse. Je remercie également le Pr Patrick Mismetti, le Pr Xavier Delavenne, le Dr Catherine Cornu et le Dr Cyril Ferdynus pour avoir accepté d'examiner ce travail. Vous me faites un grand honneur, et je vous en remercie sincèrement. Veuillez trouver ici l'expression de ma gratitude et de mon profond respect.

## **A ma directrice de thèse Madame le Dr Silvy Laporte**

Silvy, trouver les mots pour te remercier sera sans doute une des tâches les plus ardues de cette thèse. En plus de ton encadrement, tes conseils et ton soutien sans faille, j'ai été touchée par ton immense humanité et le profond respect que tu témoignes à ton équipe et à tes élèves : tu es un exemple à suivre. Je te remercie de m'avoir permis de réaliser ce travail. Sans toi je n'en serais pas là. Je te remercie de m'avoir acceptée comme élève, de m'avoir formée et de m'avoir transmis ta réflexion et ta méthodologie de travail. Tes qualités professionnelles et ta rigueur sont pour moi des exemples à suivre. Merci de m'avoir enseigné l'art de la statistique médicale, de la pharmaco-épidémiologie et des méta-analyses. Silvy, reçois l'expression de ma respectueuse gratitude. Merci pour hier, aujourd'hui et demain.

## **A mon co-directeur Monsieur le professeur Michel Cucherat**

Un immense merci pour vos idées, vos conseils et votre aide au cours de cette thèse. J'ai été honorée de publier à vos côtés.

## **A mon co-encadrant Monsieur le Dr Cyril Ferdynus**

Merci pour tes conseils, tes relectures et ton soutien.

## **A l'équipe de Recherche Clinique, Innovation et Pharmacologie**

Merci également à toutes les personnes qui ont été impliquées dans le travail de cette thèse au cours de ces années, et plus particulièrement à Céline Chapelle pour son travail, son aide, son temps et sa disponibilité depuis mon arrivée dans l'équipe. Je tiens également à remercier le Dr Edouard Ollier pour son aide, pour ses précieux commentaires sur ce travail et pour avoir pris

le temps de répondre à mes nombreuses questions. Je remercie aussi le Pr Laurent Bertoletti pour son aide et son esprit de recherche.

Merci à l'ensemble du service de l'Unité de Recherche Clinique, Innovation et Pharmacologie pour leur bonne humeur et leur gentillesse. Un très grand merci à toutes les personnes avec qui j'ai partagé un excellent moment scientifique, en particulier Emilie, Arnauld, Paul, Carine, Béatrice, Céline P, Caroline, Florence, Marie, Edith, Corine, Laurent, Michael, Maureen et Charly.

### **A ma Famille**

**A Radwan.** Je tiens à remercier tout particulièrement et chaleureusement mon époux Radwan pour son soutien sans faille durant ces années, et pour avoir cru en moi tout au long de cette expérience. Merci de m'avoir soutenue et supportée, merci d'être toujours là à mes côtés.

**A Wassef et Bassel.** Merci de m'avoir rendu heureuse toutes ces journées d'écriture studieuses et parfois fastidieuse. Si je suis là, c'est pour vous mes chéris.

**A mes parents Rima et Jamal.** Merci d'avoir été là à mes côtés à chaque instant de ma vie. Malgré les kilomètres qui nous séparent, je pense toujours bien à vous.

# Liste des productions scientifiques

---

## Directement issu du travail de thèse

- Safieddine M, Chapelle C, Ollier E, Bertoletti L, Bellet F, Mismetti P, Cucherat M, Laporte S. Quel apport des études de vraie vie par rapport aux études randomisées avec les anticoagulants oraux directs (AOD) ? RESP 2018; 66 (S3):S176
- Safieddine M, Chapelle C, Ollier E, Bertoletti L, Bellet F, Mismetti P, Cucherat M, Laporte S. Comparison of randomized controlled trials and cohort studies for the assessment of direct oral anticoagulants (DOAC). Fund Clin Pharmacol 2018; 32 (S1): 12.
- Safieddine M, Chapelle C, Ollier E, Ferdynus C, Bertoletti L, Mismetti P, Cucherat M, Laporte S. Observational studies versus randomized clinical trials with direct oral anticoagulants: a meta-epidemiological approach, J Clin Epidemiol 2020; soumis. Under Review.

## En lien avec la thèse

- Laporte S, Chapelle C, Trone JC, Bertoletti L, Girard P, Meyer G, Safieddine M, Cucherat M, Ollier E, Mismetti P, META-EMBOL Group. Early detection of the existence or absence of the treatment effect: A cumulative meta-analysis. Journal of Clinical Epidemiology. J Clin Epidemiol 2020; 124: 24-33.

## Autres publications durant la thèse

- Safieddine M, Kassir R. COVID 19 and the race to publish: an ethical issue. Br J Surg. 2020;10.1002/bjs.11966.
- Dargai F, Hoel G, Safieddine M, Payet E, Leonard R, Jaffarbanjee Z, Daoudi I. Ten-year radiological and clinical outcomes of capitulunate arthrodesis with scaphoid and triquetrum excision for advanced degenerative arthritis in the wrist: Single-center, retrospective case series with 10 patients. Hand Surg Rehabil 2020; 39: 41-47.
- El Soueidy, Kassir R, NAKHOUL M, Nunziante M, Safieddine M, Lainas P, Dagher I. Laparoscopic greater curvature plication for the treatment of severe obesity: an updated systematically review. Obes Surg 2020; en révision.
- El Soueidy T, Kassir R, Nakhoul M, Ballian A, Nunziante M, Safieddine M, Perlemuter G, Lainas P, Dagher I. Laparoscopic greater curvature plication for the treatment of severe obesity: a systematic review. Obes surg 2020; minor revision.

# Abréviations

---

AMM.....	Autorisation de mise sur le marché
AOD .....	Anticoagulants oraux directs
AVC .....	Accident vasculaire cérébral
AVK .....	Antivitamines K ou antagonistes de la vitamine K
CONSORT .....	Consolidated Standards of Reporting Trails
DR .....	Différence des risques
ECR.....	Essai contrôlé randomisé
EMA .....	Agence européenne du médicament
EP .....	Embolies pulmonaires, Embolie pulmonaire
ETEV.....	Evénements thromboemboliques veineux
FA.....	Fibrillation atriale ou auriculaire
FDA.....	Food and Drug administration
HR .....	Hazard ratio
INR .....	International Normalized Ratio
ITT.....	Intention de traiter
MTEV.....	Maladie thromboembolique veineuse
OR .....	Odds ratio
PP .....	Per protocole
RC.....	Rapport des cotes
RHR.....	Ratio des Hazard ratio
ROR.....	Ratio des odds ratio ou Rapport des rapports de cotes
RR.....	Risque relatif
SP .....	Score de propension
TEA.....	Accidents thromboemboliques artériels
TVP .....	Thrombose veineuse profonde
IC.....	Intervalle de confiance

# Table des matières

---

1	Introduction .....	16
1.1	Les études randomisées .....	16
1.1.1	Définition et objectifs .....	16
1.1.2	Les grands principes méthodologiques des essais randomisés .....	16
1.1.3	Outils d'évaluation du risque de biais des études randomisées .....	19
1.2	Les études observationnelles .....	21
1.2.1	Définitions et objectifs .....	21
1.2.2	Études de cohortes prospectives .....	22
1.2.3	Études de cohortes rétrospectives .....	22
1.2.4	Études observationnelles sur bases de données de santé informatisées .....	23
1.2.5	Utilisation des études observationnelles pour l'évaluation d'un traitement ....	26
1.3	Les études randomisées et les études observationnelles concernant les anticoagulants oraux directs dans le traitement des ETEV et dans la FA .....	33
1.3.1	La maladie thromboembolique veineuse : épidémiologie et impacts pour les patients .....	33
1.3.2	La fibrillation atriale : épidémiologie et impacts pour les patients .....	34
1.3.3	Contexte du développement des AOD dans la MTEV et la FA .....	37
1.3.4	Données disponibles avec les AOD dans la MTEV .....	41
1.3.5	Données disponibles avec les AOD dans la FA .....	45
1.4	Objectifs de la thèse .....	47

2	La méta-épidémiologie.....	49
2.1	La méta-analyse.....	49
2.1.1	Principes de la méta-analyse .....	49
2.1.2	Intérêts de la méta-analyse .....	50
2.1.3	Déroulement d'une méta-analyse .....	51
2.1.4	Analyse des effets.....	53
2.1.5	Illustrer les résultats .....	56
2.2	De la méta-analyse à la méta-épidémiologie.....	61
2.2.1	Principes de la méta-épidémiologie .....	61
2.2.2	Exemples « célèbres » d'études de méta-épidémiologie.....	67
2.2.3	Méta-épidémiologie, études observationnelles et ECR.....	71
2.2.4	Lignes directrices proposées pour les études méta-épidémiologiques.....	72
3	Quantification du biais d'évaluation dans les études évaluant les anticoagulants oraux dans la maladie thromboembolique veineuse.....	75
3.1	Recherche exhaustive et sélection des études .....	75
3.1.1	Critères d'éligibilité des études .....	76
3.1.2	Revue de la littérature .....	76
3.1.3	Analyses des études éligibles .....	77
3.1.4	Critères de jugement.....	77
3.1.5	Extraction des données.....	77
3.2	Description des études et des populations incluses.....	78

3.2.1	Identification et sélection des études.....	78
3.2.2	Comparaison des populations ECR et des études observationnelles .....	81
3.3	Approche méta-analytique : comparaison des estimations des effets du traitement entre les études randomisées et les études observationnelles.....	83
3.3.1	Méthodes .....	83
3.3.2	Résultats .....	84
3.4	Discussion des principaux résultats et limites.....	90
4	Quantification du biais d'évaluation dans les études évaluant les anticoagulants oraux dans la fibrillation auriculaire .....	93
4.1	Recherche exhaustive et sélection des études .....	93
4.1.1	Critères d'éligibilité des études .....	93
4.1.2	Revue de la littérature .....	94
4.1.3	Analyses des études éligibles .....	95
4.1.4	Critères de jugement.....	95
4.1.5	Extraction des données.....	95
4.2	Description des études et des populations incluses.....	96
4.2.1	Identification et sélection des études.....	96
4.2.2	Gestion des doublons .....	97
4.2.3	Comparaison des populations ECR et des études observationnelles .....	103
4.3	Approche méta-analytique : comparaison des estimations des effets des traitements entre les études randomisées et les études observationnelles.....	107
4.3.1	Méthodes .....	107

4.3.2	Résultats .....	108
4.3.3	Analyse de sensibilité.....	115
4.4	Discussion des principaux résultats et limites.....	115
5	Discussion générale.....	118
5.1	Une sélection moindre des patients dans les études observationnelles ?.....	118
5.2	Principaux résultats des études méta-épidémiologiques dans le traitement des ETEV et dans la FA.....	118
5.3	Limites de ce travail de pharmaco-épidémiologie .....	121
5.4	Perspectives .....	122
5.5	Les leçons actuelles de ce travail .....	122
6	Références .....	125
7	Annexes.....	144

# Liste des tableaux

---

Tableau 1: Principales caractéristiques pharmacocinétiques .....	40
Tableau 2: Tableau de calcul permettant de comparer deux rapports de risque .....	66
Tableau 3: Caractéristiques des patients des groupes AOD selon le type d'étude avant tout ajustement ou appariement.....	82
Tableau 4: Liste des études conduites sur des bases de données américaines .....	98
Tableau 5: Liste des études conduites sur la base danoise .....	101
Tableau 6: Caractéristiques des patients selon le type d'étude avant ajustement ou appariement .....	105
Tableau 7: Caractéristiques des patients selon le type d'étude après ajustement ou appariement (si disponible) .....	106

# Liste des figures

---

Figure 1: Cibles pharmacologiques des anticoagulants .....	38
Figure 2: Niveau de preuve de la médecine basée sur les faits .....	50
Figure 3: Exemple d'un forest plot .....	57
Figure 4: Illustration d'un funnel plot .....	58
Figure 5: Illustration d'un funnel plot symétrique et asymétrique .....	59
Figure 6: Exemple d'identification d'une interaction avec les caractéristiques des études. ....	60
Figure 7: Funnel plot symétrique illustrant le biais de publication .....	61
Figure 8: ROR estimé des essais sans double aveugle par rapport aux essais en double aveugle .....	67
Figure 9: ROR des essais sans double aveugle versus essais en double aveugle dans 4 méta- analyse différentes et sommation des ROR.....	68
Figure 10: ROR combiné entre les études observationnelles et les ECR.....	72
Figure 11: Diagramme de flux du processus de sélection des études .....	80
Figure 12: Graphique en entonnoir des résultats sur les récurrences d'ETEV en fonction des types d'études.....	85
Figure 13: Graphique en entonnoir des résultats sur hémorragies majeures en fonction des types d'études.....	85
Figure 14: Comparaison des estimations des effets du traitement en fonction des types d'études sur les récurrences d'ETEV .....	86
Figure 15: Comparaison des estimations des effets du traitement en fonction des types d'études sur les hémorragies majeures.....	87

Figure 16: RHR et IC à 95% des RHR pour quantifier la différence de l'effet traitement entre les études observationnelles et les ECR .....	89
Figure 17: Diagramme de flux du processus de sélection des études .....	97
Figure 18: Graphique en entonnoir des résultats sur les AVC ou les embolies systémiques en fonction du type d'études .....	109
Figure 19: Graphique en entonnoir des résultats sur les hémorragies majeures en fonction du type d'études .....	110
Figure 20: Comparaison des estimations des effets du traitement en fonction du type d'études sur les AVC ou les embolies systémiques.....	112
Figure 21: Comparaison des estimations des effets du traitement en fonction du type d'études sur les hémorragies majeures .....	113
Figure 22: RHR et intervalles de confiance à 95% des RHR pour quantifier la différence d'estimation de l'effet traitement entre les études observationnelles et les ECR .....	114
Figure 23: RHR et intervalles de confiance à 95% des RHR pour quantifier la différence d'estimation des effets traitement entre les études observationnelles et les ECR.....	120

# Partie introductive

---

Les essais contrôlés randomisés (ECR) sont considérés comme la méthode de référence pour évaluer l'efficacité et la sécurité d'un traitement. Toutefois, des études observationnelles sont également conduites pour estimer le bénéfice médical de ces traitements. Dans ce contexte, plusieurs questions sont légitimes : quel est l'intérêt de ces études observationnelles ? Comment sont pris en compte leurs biais potentiels ? Peut-on quantifier leurs biais par rapport à ceux des études randomisées ? Que nous apportent ces études observationnelles par rapport aux ECR dans l'évaluation d'un traitement ?

Afin de répondre à ces problématiques, l'objectif de cette thèse est de déterminer les différences entre les populations incluses (ECR vs études observationnelles) et de vérifier si les études observationnelles apportent des informations supplémentaires concernant le rapport bénéfice / risque d'un traitement en vie réelle. Pour étayer nos travaux, nous avons pris l'exemple des anticoagulants oraux directs (AOD) dans le traitement des événements thromboemboliques veineux (ETEV) et dans la fibrillation auriculaire (FA). Des méthodes méta-épidémiologiques ont été utilisées pour quantifier la différence d'estimation des effets du traitement, en termes d'efficacité et de sécurité, entre ces deux types d'études.

La première partie de cette thèse présente les objectifs et les grands principes méthodologiques des essais randomisés et des différents types d'études observationnelles. Nous y discuterons les biais des études observationnelles ainsi que les méthodes d'ajustement utilisées afin de contrôler ces biais. Pour cette partie, une revue de la littérature a été réalisée sur les études randomisées et les études observationnelles ayant comparé les effets des AOD à ceux des Anti-Vitamines K (AVK), qui sont la référence dans le traitement des ETEV et dans la FA.

La seconde partie de la thèse concerne l'approche des méthodes méta-analytiques et méta-épidémiologiques. Cette partie expose les objectifs des études méta-épidémiologies et les différentes méthodes utilisées afin de quantifier les biais potentiels par rapport aux ECR.

La troisième partie de la thèse cherche à quantifier les biais d'évaluation dans les études observationnelles évaluant les AOD dans le traitement des ETEV. L'objectif de cette partie est de comparer la population incluse dans les ECR à celle incluse dans les études observationnelles, et également de comparer l'effet du traitement par AOD à ceux du traitement par AVK (traitement de référence) entre les ECR et les études observationnelles. La récurrence

des ETEV était choisie comme le critère de jugement primaire et les hémorragies majeures comme critère secondaire. Des méta-analyses stratifiées selon les types d'études ont été réalisées. Ainsi, une quantification du biais d'évaluation a été effectuée en calculant le ratio des Hazard ratio (RHR).

La quatrième partie de cette thèse cherche à quantifier les biais d'évaluation dans les études observationnelles évaluant les AOD dans la FA. Les mêmes méthodes méta-épidémiologiques que dans la partie précédente ont été utilisées. En prenant l'exemple de la FA, l'objectif de ce travail dans la FA était de confirmer les résultats obtenus précédemment avec un plus grand nombre d'études. Le risque d'accidents thromboemboliques artériels (TEA) était choisi comme le critère de jugement primaire et les hémorragies majeures comme critère secondaire.

Nous terminerons cette thèse par une discussion générale.

# 1 Introduction

## 1.1 Les études randomisées

### 1.1.1 Définition et objectifs

La mise en évidence du bénéfice clinique apporté par un traitement est une étape importante dans le développement d'un traitement. Le développement optimal doit être protégé des biais qui feraient courir le risque de conclure à tort à l'efficacité / l'inefficacité du traitement. Des principes méthodologiques ont été développés pour éviter l'existence de ces biais et conduire à la méthode de l'essai contrôlé randomisé (ECR) en double aveugle. La méthodologie éprouvée de ces ECR est à la base de l'« evidence based medicine » (EBM), apportant le meilleur niveau de preuve scientifique. Elle est considérée comme la méthode de référence « gold standard » lors de l'autorisation de mise sur le marché (AMM) d'un nouveau médicament par la Food and Drug administration (FDA) aux États-Unis et l'agence européenne du médicament (EMA)<sup>(1,2)</sup>. Ces AMM se basent sur les résultats d'efficacité et de sécurité (effets indésirables) obtenus dans les ECR menés idéalement en double insu<sup>(3)</sup>.

L'objectif d'un essai thérapeutique est souvent d'établir une relation causale entre le(s) critère(s) de jugement et l'administration du traitement étudié. Cependant, il existe de nombreux facteurs interférant avec cette éventuelle relation. Ces facteurs ont des effets pouvant être confondus avec l'effet du traitement étudié : facteurs de confusion (« *confounding factor* »)<sup>(4)</sup>. Les principes méthodologiques des ECR ont pour objectifs de contrôler ces facteurs de confusion et d'éviter leur influence nuisible. La suite de cette partie décrit les principes méthodologiques permettant de contourner les différents biais qui surviennent dans la mesure de l'efficacité d'un traitement.

### 1.1.2 Les grands principes méthodologiques des essais randomisés

#### *A/ Approche prospective*

Pour répondre à la question principale de l'étude (objectif principal), un critère de jugement principal (« *primary endpoint* » ou « *outcome* ») doit généralement être défini. Les critères secondaires sont définis pour décrire d'autres effets du traitement (objectifs secondaires)

comme l'apparition d'événements indésirables. Ces critères de jugement (primaire ou secondaire) sont des variables dont la valeur numérique est susceptible de changer sous l'effet du traitement (« *treatment effect* »). Ils peuvent être de nature très variée : durée d'une maladie, fréquence de survenue d'un événement, taux de mortalité, score clinique, etc. Les individus inclus sont suivis dans le temps à partir d'une date de référence (*index date*) jusqu'à une date ultérieure afin de mesurer le(s) critère(s) de jugement recherché(s). Les données sont recueillies alors prospectivement et spécialement pour répondre à la question posée dans l'essai<sup>(5)</sup>.

#### *B/ Plan d'expérience (« design ») : comparatif*

Dans les ECR, le protocole expérimental de recherche clinique inclut des patients de plusieurs centres selon des critères d'éligibilité stricts. Ces patients sont répartis entre le groupe recevant le traitement étudié et le groupe contrôle. Ainsi, l'effet du traitement est déterminé par rapport au groupe contrôle pour prendre en compte les facteurs confondants comme l'évolution spontanée de la maladie, l'effet placebo, l'erreur de mesure ou l'influence des traitements concomitants. L'utilisation d'un groupe contrôle est indispensable pour éviter l'influence de ces facteurs mais cela n'est pas suffisant pour éviter tous les biais<sup>(6)</sup>.

#### *C/ La randomisation*

La randomisation est un procédé selon lequel l'attribution d'un traitement au patient est réalisée de façon aléatoire (c'est l'allocation aléatoire des traitements aux patients). Ainsi, elle crée deux groupes de patients initialement comparables : groupe contrôle et le groupe expérimental. La randomisation assure une répartition équilibrée des facteurs de pronostics (connus ou inconnus) entre ces deux groupes afin de s'assurer que la différence observée en fin d'étude entre les groupes soit attribuable au traitement reçu<sup>(5)</sup>. Le moyen de randomisation le plus simple est la liste constituée à l'aide d'un programme informatique et basée le plus souvent sur des blocs de taille variable. Une fois la liste générée, elle fournit, pour chaque patient inclus dans l'essai, le numéro de la boîte ou la nature du traitement. Cette attribution doit rester secrète jusqu'à l'inclusion du patient. Ainsi, une randomisation effectuée à l'aide d'enveloppes scellées renfermant le numéro ou la nature du traitement à donner au patient n'est pas recommandée pour les études en ouvert. En effet, tous les moyens de randomisation autorisant une prise de connaissance anticipée de la nature du traitement que devraient recevoir les futurs patients anéantissent l'intérêt de la randomisation car ne protègent plus du risque de biais de sélection. En effet, si la nature du traitement que recevra le futur patient est prévisible, l'investigateur a la

possibilité de retarder le moment de l'inclusion d'un patient jusqu'à ce qu'il reçoive le traitement<sup>(7,8)</sup>. Une procédure centralisée de randomisation évite cette faiblesse et garantit son imprévisibilité. Le numéro ou la nature de traitement est obtenu(e) par appel à un centre de randomisation ou par connexion sur un site internet, après enregistrement du patient inclus dans l'essai.

#### *D/ Le double insu ou double aveugle (« double blinding »)*

Les biais sont évités non seulement par la randomisation mais aussi par la procédure de double aveugle. Cette procédure signifie que l'apparence des traitements de l'étude est identique. Elle permet aussi de maintenir la comparabilité entre les groupes tout au long du suivi et lors de la mesure du critère de jugement. Par conséquent, elle évite toute différence dans le suivi et l'évaluation des deux groupes, différence que l'on pourrait observer dans une étude en ouvert (« *open design* »), c'est-à-dire que le médecin et le patient connaissent la nature du traitement. Elle permet de minimiser les biais de suivi et de mesure/d'évaluation<sup>(6)</sup>. Le recours au double aveugle est particulièrement important quand le critère de jugement est de nature subjective. Pour rappel, il existe deux types d'aveugle<sup>(5,6)</sup> :

- Double aveugle : ni le patient, ni le médecin investigateur ne connaissent la nature réelle du traitement.
- Simple aveugle : le médecin connaît la nature du traitement mais pas le patient.

#### *E/ Analyse de la population*

Malgré les précautions prises dans un essai, des écarts au protocole peuvent survenir chez certains patients. Parmi les différents écarts au protocole, on note l'arrêt prématuré du traitement de l'étude ou l'absence de prise de ce traitement, la prise d'un traitement interdit ou de l'autre groupe, l'absence de données concernant le critère de jugement, la non présentation aux visites de suivi et l'inclusion de patients ne répondant pas aux critères d'inclusion. Le retrait de ces patients de l'essai risque d'introduire un biais d'attrition et de détruire la comparabilité initiale des groupes. Pour éviter le biais d'attrition et maintenir la comparabilité des groupes, un essai doit être analysé en intention de traiter (ITT)<sup>(6)</sup>. En effet, l'ITT permet d'analyser tous les patients randomisés dans leur groupe de randomisation. Le principe de l'analyse en ITT exige que tous les patients randomisés soient suivis jusqu'à leur décès ou à la fin de l'essai, ou jusqu'à l'observation de l'événement critère de jugement. Il existe un autre type d'analyse, l'analyse per protocole (PP), qui consiste en un sous-groupe de patients sélectionnés sur le respect du

protocole et la disponibilité des données. Elle analyse uniquement les patients qui ont été traités en pleine conformité avec le protocole. L'analyse en PP donne ainsi une meilleure estimation potentielle de l'effet du traitement<sup>(9)</sup>. Mais cette estimation est potentiellement biaisée puisqu'elle exclut des patients de l'analyse et ré-introduit donc un possible biais de sélection. Par conséquent, l'analyse principale d'un essai doit être faite en intention de traiter puisqu'elle garantit l'absence de biais, et permet ainsi l'estimation de l'effet du traitement dans les conditions proches de celle de la vie réelle en analysant tous les patients inclus dans l'essai<sup>(5)</sup>.

Au total, un essai randomisé avec randomisation imprévisible, conduit en double aveugle et analysé en intention de traiter permet globalement de garantir une estimation non biaisée de l'effet traitement. Toutes les études ne remplissant pas ces critères, il existe des outils permettant de classer ces études face à leur risque de biais.

### 1.1.3 Outils d'évaluation du risque de biais des études randomisées

*A/ L'outil de la Cochrane : « RoB 2 »*

En 2008, la Cochrane collaboration a lancé un outil d'évaluation de risque de biais (RoB) dans les essais randomisés inclus dans les revues systématiques et les méta-analyses<sup>(10)</sup>. Cet outil a remplacé les grilles habituelles de qualité d'une étude par une l'évaluation du risque de biais. Il prend en compte les biais qui surviennent à différentes étapes d'un essai (domaines de biais). Ces domaines ont été choisis sur la base de preuves empiriques et de considérations théoriques. Une nouvelle version de l'outil : RoB 2, a été mise en œuvre en 2016. Elle intègre les progrès de l'évaluation du risque de biais rendus possibles par les outils récemment développés<sup>(11,12)</sup>. En 2019, cet outil a été mis à jour en prenant compte les commentaires des utilisateurs et les dernières avancées scientifiques<sup>(13)</sup>.

L'outil actuel est basé sur les sources potentielles de biais pouvant apparaître dans les essais cliniques. Il fournit ainsi un cadre de référence permettant de prendre en compte le risque de biais dans les résultats des différents types d'essais randomisés. Cet outil est appliqué à des essais randomisés individuels, en groupes parallèles, en cluster et en cross-over. Une évaluation du risque de biais par l'outil RoB 2 consiste à : préciser le résultat à évaluer, évaluer la façon dont ce résultat est interprété, et à énumérer les sources d'information utilisées afin de documenter l'évaluation.

Le RoB 2 comporte cinq domaines de biais qui sont utilisés dans l'évaluation du risque de biais des ECR :

1. Biais émergeant au moment du processus de randomisation
2. Biais dû à des déviations par rapport aux interventions prévues
3. Biais dû à des données finales manquantes
4. Biais dû à la mesure du résultat
5. Biais dans la sélection des résultats rapportés

Des questions indicatives sont développées pour chaque domaine. Ces questions visent à obtenir des informations pertinentes pour l'évaluation du risque de biais. Les options de réponse sur ces questions indicatives sont : oui, probablement oui, probablement non, non, aucune information. Ensuite, un algorithme est mis en place afin d'associer les réponses aux questions spécifiques à une proposition de jugement du risque de biais pour chaque domaine. Les jugements sur le risque de biais sont : faible risque de biais, incertitude, risque de biais élevé.

Un jugement global quant au risque de biais peut être porté pour un résultat précis dans une étude donnée, à partir des qualificatifs suivants :

- Faible risque de biais : l'étude est jugée à faible risque de biais dans tous les domaines pour ce résultat ;
- Inquiétude : l'étude soulève certaines inquiétudes de biais dans au moins un domaine pour ce résultat ;
- Risque de biais élevé : l'étude est jugée à haut risque de biais dans au moins un domaine pour ce résultat ou l'étude est jugée soulever plusieurs incertitudes dans de multiples domaines d'une manière qui réduit considérablement la confiance dans le résultat.

En résumé, les étapes d'utilisation de l'outil RoB 2 permettent tout d'abord de déterminer le risque de biais dans chacun des cinq domaines et ensuite de déterminer un risque de biais global pour un résultat précis à travers les cinq domaines. Dans le cadre d'une méta-analyse, l'outil RoB 2 est le plus recommandé pour distinguer les résultats à haut risque de biais des résultats à faible risque de biais. Il permet de stratifier les résultats des méta-analyses en fonction du risque de biais.

## *B/ Les recommandations CONSORT « Consolidated Standards of Reporting Trails »*

Afin de faciliter l'évaluation des biais dans une étude, des lignes directrices CONSORT<sup>(14)</sup> ont été développées pour aider les auteurs à rédiger de façon standardisée la méthodologie et les résultats des ECR. Le « CONSORT » est composé d'une liste de contrôle de 25 points indispensables lors de la publication d'un ECR et d'un organigramme « Flow chart » qui montre l'évolution des patients tout au long de l'essai et le nombre des patients analysés. L'objectif de ces recommandations est d'obtenir une meilleure transparence dans la présentation et par suite la conception des essais randomisés. Par conséquent, ces lignes directrices offrent aux lecteurs une facilité de lecture et d'évaluation des biais potentiels dans les essais.

En synthèse, tous ces principes méthodologiques ont fait des ECR les meilleures études pour montrer l'efficacité et la sécurité d'un nouveau traitement dans une pathologie donnée. Néanmoins, une fois ces études conduites pour l'autorisation de la mise sur le marché d'un médicament, il n'est pas rare que les autorités de santé demandent la réalisation d'études observationnelles post-AMM afin de vérifier les conditions de prescription et de confirmer l'efficacité du traitement en vie réelle. Quelle est alors la place de ces études par rapport aux ECR ?

## 1.2 Les études observationnelles

### 1.2.1 Définitions et objectifs

Les études observationnelles utilisent des méthodes épidémiologiques et des connaissances pharmacologiques pour étudier l'efficacité et la sécurité des médicaments après leur commercialisation. L'objectif de ces études est de décrire, d'expliquer, contrôler et prévoir les utilisations et les effets bénéfiques ou indésirables des médicaments dans la population<sup>(15)</sup>. Contrairement aux ECR, l'évaluation des traitements repose sur des approches observationnelles. Les anglo-saxons ont bien distingué l'efficacité évaluée par les ECR dans des conditions optimales (*efficacy*) de l'efficacité évaluée par les études observationnelles (*effectiveness*). Dans ces études, le choix des traitements et leurs modalités d'administration ne sont pas déterminés par un protocole mais par la pratique clinique courante. Les données de santé de ces études sont générées au cours du suivi habituel des malades dans le cadre de la pratique clinique de routine, souvent appelée « en vie réelle », de soins. Ces données peuvent

être collectées par de multiples moyens : registres, cohortes prospectives, bases de données électroniques, objets connectés, remboursements des soins médicaux, web, etc.<sup>(1,16)</sup>. Si les données de santé recueillies le sont dans le cadre de la prescription et du suivi d'un médicament, c'est le domaine de la pharmaco-épidémiologie, post-AMM.

Les études observationnelles les plus fréquemment utilisées en pharmaco-épidémiologie sont les études de cohorte, car elles permettent d'estimer la fréquence d'un événement mais aussi l'existence d'une association entre un facteur d'exposition (le médicament) et l'événement recherché. Nous distinguerons ici trois grandes catégories selon leur schéma expérimental : études de cohortes prospectives, études de cohortes rétrospectives et études sur bases de données de santé informatisées.

### 1.2.2 Études de cohortes prospectives

En pharmaco-épidémiologie, le principe d'une étude de cohorte est le recrutement le plus rapide possible d'une population, indemne de l'événement d'intérêt, dans laquelle des sujets reçoivent ou non le traitement étudié. Les groupes sont ensuite suivis à partir d'une date de référence jusqu'à une date ultérieure suffisamment éloignée pour avoir de bonnes chances d'observer et de mesurer la fréquence de l'événement recherché. La collecte des caractéristiques des individus se fait à l'inclusion et l'incidence des événements étudiés est recueillie pendant le suivi. La fréquence des événements est mesurée ainsi prospectivement et elle est comparée entre les groupes de patients (exposés et non exposés au médicament évalué)<sup>(17,18)</sup>. Par conséquent, ces cohortes prospectives offrent une mesure prospective de l'exposition au médicament étudié et une évaluation du lien entre cette exposition et plusieurs événements. Cette comparaison s'apparente à une comparaison de groupes dans un essai randomisé, à la différence majeure que dans ces études observationnelles, même prospectives, le traitement n'est pas tiré au sort, et est toujours administré en ouvert.

### 1.2.3 Études de cohortes rétrospectives

Lorsque l'événement mesuré peut survenir plusieurs années après l'exposition au traitement (par exemple un effet non connu des statines après 15 ans de prescription), il est possible d'avoir une estimation de l'association entre médicament et événement en ayant recours aux études rétrospectives. Contrairement aux cohortes prospectives, la question de recherche est lancée après l'exposition au traitement et la survenue de l'événement étudié. Ces cohortes sont

conduites sur des données archivées (dossiers hospitaliers, dossiers médicaux informatisés, dossiers administratifs, questionnaires, etc.). L'inclusion des individus peut se faire rétrospectivement à partir d'une date choisie dans le passé. Ainsi, la totalité de leur période de suivi se situe dans le passé. Comme dans les études de cohortes prospectives, les individus sont classés en fonction de leur exposition ou non au traitement, afin de comparer l'incidence des événements étudiés entre les deux groupes : groupe exposé au traitement, groupe non exposé au traitement. Par conséquent, et contrairement aux études prospectives, ces études rétrospectives proposent des mises en place d'études plus rapides et moins coûteuses. Néanmoins, elles sont limitées par les données déjà existantes dans les dossiers de santé. Ces cohortes sont aussi sujettes aux biais de sélection/confusion et d'évaluation, mais également à d'autres types de biais liés au caractère rétrospectif de l'étude et le non recueil systématique des données.

#### 1.2.4 Études observationnelles sur bases de données de santé informatisées

##### *A/ Définition*

Comme dans les études de cohortes rétrospectives, la question de recherche des études portant sur des bases de données bases de données de santé informatisées est lancée après l'exposition au traitement. Le suivi des patients est donc fait totalement dans le passé. Ces études observationnelles sont conduites sur des bases de données volumineuses à l'inverse des cohortes rétrospectives qui sont généralement conduites sur des petites bases de données (base de données de l'hôpital). Ces bases de données sont créées à partir des informations électroniques collectées dans le cadre du recours aux soins des populations.

Grâce à cette caractéristique de données massives, ces dernières sont souvent nommées « big data »<sup>(19,20)</sup>, même s'il n'est pas toujours évident de savoir à quoi cette expression renvoie précisément. Dans les critères des big data, on note des critères initiaux dits des trois « V » (volume, vitesse, variété) auxquels se sont progressivement ajoutés trois autres « V » (valeur, variabilité et véracité). Les big data sont donc des données représentant une grande quantité d'information, rapidement mise à jour et rapidement utilisable. Néanmoins, elles sont hétérogènes. La « conformité » en termes de qualité et de pertinence clinique doit être prise en compte et évaluée<sup>(19)</sup>. Pour l'évaluation pharmaco-épidémiologique, on distingue plusieurs types de données informatisées utilisables : les données informatisées de santé (health data), de nature administrative (administrative data), les données médicales, à savoir les dossiers

médicaux informatisés (electronic health record) et les données médico-administratives (regroupement des deux premières).

On parle de données administratives de remboursement (remboursement data, claims data ou insurance data) lorsque l'enregistrement de ces informations a pour objectif de permettre le remboursement des soins de santé pour les patients assurés<sup>(21-23)</sup>. Dans la littérature, les bases de données correspondantes sont souvent sans référence précise à la nature des données qu'elles contiennent. Elles sont désignées par « base ou système national de données de santé » (national healthcare database ou national research database) voire par le terme inapproprié de registres (registry) pour les bases nord-européennes (risque de confusion possible en épidémiologie)<sup>(24)</sup>.

Parmi les bases de données de santé informatisées les plus utilisées en pharmaco-épidémiologie, on note :

- Les programmes gouvernementaux d'assurance maladie Medicare (personnes  $\geq 65$  ans,  $< 65$  ans et handicapées, ou atteintes d'insuffisance rénale terminale), Medicaid (personnes et familles à faible revenu et disposant de peu de ressources), ou la Veterans' Health Administration aux États-Unis<sup>(25,26)</sup>. Le système de santé américain est complexe, faisant intervenir différents acteurs des secteurs public et privé. Il n'existe pas aux États-Unis de régime général d'assurance-maladie. Les assurances publiques ne couvrent que certaines catégories de personnes et le mode de financement s'appuie sur les assurances privées. Compte-tenu du nombre de publications d'études pharmaco-épidémiologiques réalisées sur ces programmes et afin de comprendre les résultats de ces études observationnelles, qui seront exploités dans le cadre de cette thèse, nous avons effectué une recherche spécifique sur ce système de santé et les bases de données de santé américaines. Cette recherche est décrite en Annexe 1.
- les bases de l'Europe du Nord<sup>(27)</sup>
- la base Clinical Practice Research Datalink (CPRD, données de soins primaires issues de médecins généralistes) au Royaume Uni<sup>(28)</sup>.

La plupart du temps, ces bases de données de santé sont reliées entre elles afin de construire un réseau de bases avec mise en commun des données de santé « anonymisées » issues de bases d'origines différentes. Ces réseaux permettent la restructuration ou la réorganisation de ces données suivant un modèle commun, garantissant ainsi une parfaite similitude des analyses réalisées dans chacune des bases. Via un unique script informatique développé au sein d'un

centre de coordination, ce modèle permet l'exploitation directe des informations de chaque base sans qu'un transfert des données sources ne soit nécessaire. Ces réseaux accroissent de manière importante la quantité et la diversité des données des patients. L'objectif principal de ces réseaux est donc l'accélération et la détection de signaux de sécurité des médicaments. Les deux réseaux les plus reconnus sont : le système de surveillance post-AMM « Sentinel » (193 millions de personnes)<sup>(29,30)</sup>, lancé en 2008 aux Etats-Unis et financé par la FDA et le projet « EU-ADR »<sup>(31)</sup>, financé par l'Union européenne la même année, associant huit bases de quatre pays européens : Danemark, Italie, Pays- Bas et Royaume-Uni (30 millions de personnes).

Parmi les autres réseaux, on note les regroupements de bases issus :

- du projet européen Arrhythmogenic potential of drugs - ARITMO avec six bases de cinq pays : Allemagne, Danemark, Italie, Pays-Bas et Royaume-Uni (29 millions de personnes) ;
- du projet international de l'Asian Pharmacoepidemiology network - AsPEN, regroupant Australie, Corée, Japon, Taïwan, Suède et États-Unis (200 millions de personnes)<sup>(32)</sup> ;
- du projet canadien du Canadian Network for Observational Drug Effect Studies - CNODES<sup>(33)</sup> avec les données des sept provinces canadiennes et celles des bases américaine 'MarketScan' et britannique 'CPRD' (plus de 100 millions de personnes).

### *B/ Avantages et limites*

Le schéma d'une étude observationnelle conduite sur bases de données de santé informatisées présente plusieurs avantages pour l'évaluation pharmaco-épidémiologique :

- Inclusion possible d'un grand nombre de patients.
- Inclusion de données collectées en temps réel de manière systématique et indépendamment des objectifs de recherche.
- Utilisation rapide de données récentes et recueillies exhaustivement dans les bases de données.
- Possibilité d'évaluation d'événements rares et inclusion des groupes de population peu ou pas représentés dans les ECR.
- Possibilité d'évaluation longitudinale rapide d'événements de santé nécessitant d'être étudiés à long terme avec des enregistrements portant parfois sur plusieurs décennies.
- Faible coût de production comparé à volume égal aux données des études prospectives<sup>(20,26,34)</sup>.

Concernant les limites des études sur bases de données de santé informatisées, on peut noter les faiblesses suivantes :

- Complexité des données recueillies.
- Relatif manque d'information médicale dans les bases de données : données manquantes.
- Difficultés d'accès et d'utilisation de certaines bases de données.
- Validité des analyses longitudinales des études qui peut être limitée par des changements dans les modes de prise en charge générale des patients par le système assurantiel<sup>(19)</sup>.
- Validité incertaine des informations cliniques des bases de données de remboursement, notamment concernant les événements étudiés.
- Absence de connaissance de l'indication d'une prescription, absence quasi-totale d'information concernant les traitements reçus en cours d'hospitalisation, et absence d'information concernant les résultats des explorations paracliniques.

A ces limites, s'ajoutent les limites des bases de données administratives qui sont des prérequis indispensables à une utilisation optimale de ces bases<sup>(19,35,36)</sup> : la description détaillée des données, la validation des événements mesurés dans les bases par retour au dossier patient<sup>(37-39)</sup>, l'enrichissement des données administratives par des données médicales et le soin apporté au protocole d'étude des données.

#### 1.2.5 Utilisation des études observationnelles pour l'évaluation d'un traitement

Pour l'évaluation d'un traitement, que ce soit pour son efficacité ou sa toxicité, ces cohortes sont sujettes au biais de sélection / confusion et au biais de mesure / d'évaluation<sup>(6)</sup>. En effet, toute association entre un effet indésirable et la prise du traitement peut être due au traitement certes, mais aussi aux comorbidités du patient, aux autres traitement concomitants, à ses caractéristiques cliniques...

## *A/ Biais sous-jacents*

Les autorités de santé (EMA, FDA, haute autorité de santé HAS, etc) encouragent la conduite des études observationnelles post-AMM afin de connaître, en situation réelle, les modalités d'utilisation, le bénéfice et l'impact d'un médicament. En effet, même si les ECR offrent la meilleure méthode pour mesurer l'efficacité et la sécurité d'un médicament, les ECR sont critiquées car la population incluse est considérée ainsi comme non représentative de la population suivie en pratique clinique. D'autre part, le complément voire même parfois le remplacement des ECR par des études observationnelles dites de vraie vie sont favorisés par la disponibilité des bases de données informatisées, ainsi que la possibilité de les analyser à un coût relativement faible<sup>(40,41)</sup>. D'autres facteurs sont aussi à l'origine de ce constat, notamment le souhait de l'industrie de réduire les coûts et le temps nécessaire pour obtenir des résultats.

Toutefois, la principale critique des études observationnelles est le risque de biais pouvant affecter la validité des résultats. En effet, l'absence de la randomisation<sup>(42)</sup> dans ces études observationnelles fait que les patients sont affectés à différents groupes de traitement selon une méthode non aléatoire. Les écarts observés entre les résultats peuvent être ainsi expliqués par les caractéristiques différentes des deux groupes. Par conséquent, l'affectation non aléatoire peut conduire à des biais lors de l'évaluation de l'effet d'un traitement et fausser les résultats. Les principaux biais rencontrés dans les études observationnelles sont le biais de sélection / confusion et le biais de mesure/d'évaluation.

### ➤ Le biais de sélection / confusion (confounding)

#### • *Biais de confusion*

Dans les études observationnelles, le statut exposé/non-exposé ne découle pas d'une assignation par un processus aléatoire. Il découle d'une décision thérapeutique pouvant être motivée par de nombreux facteurs de confusion. Ces facteurs faussent la relation causale (cause à effet). En effet, une association trouvée entre une exposition et un événement peut être expliqué par une relation directe alors qu'elle résulte d'une relation indirecte via un tiers facteur de confusion. D'une part, un biais de confusion initiale survient lorsque des variables pronostiques pré-intervention (variables qui prédisent le résultat recherché) prévoient également si une personne reçoit l'une ou l'autre des interventions à l'étude. D'autre part, il existe des variables de confusion qui varient tout au long de la période de comparaison. Ces facteurs surviennent lorsque les individus passent d'une intervention à l'autre et lorsque des variables pronostiques

post intervention entraînent une déviation de l'intervention initiale (prévue). Certains facteurs de confusion sont difficiles à identifier, d'autres demeurent probablement inconnus et peuvent être à l'origine d'une confusion résiduelle. Le risque de biais de confusion est sérieux quand des domaines de confusions importants n'ont pas été pris en compte, ou encore quand ils étaient peu susceptibles d'être bien identifiés par les variables mesurées. Ces facteurs de confusion doivent être déterminés et pris en compte dans l'analyse statistique de l'étude.

- *Biais dans la sélection des participants à l'étude*

L'absence de la randomisation entraîne une absence de la comparabilité des groupes à l'inclusion. En effet, la procédure de sélection des sujets est différente entre les deux groupes (traités/non traités), ce qui affecte la constitution de l'échantillon et produit un biais de sélection.

Un biais de sélection apparaît lorsque l'exclusion des participants ou la période de suivi de certains participants font en sorte que l'association entre l'intervention et le résultat diffère de ce qui aurait été observé dans l'essai cible. Ce phénomène est différent de la confusion. Ce biais peut se produire quand des utilisateurs existants (prévalent) plutôt que des nouveaux utilisateurs (incident), sont inclus dans les analyses (le biais des utilisateurs existants). Le biais de sélection fait référence seulement aux biais qui sont internes à l'étude et non aux questions de généralisabilité. Il inclut le biais du commencement, le biais du temps d'avance au diagnostic et le biais de temps immortel. Ce biais de sélection apparaîtra si la sélection dans l'étude est influencée par l'intervention et par le résultat.

La sélection en dehors de l'étude est aussi une préoccupation couverte dans le domaine des données manquantes.

- Le biais de mesure / d'évaluation

- *Biais dans la classification des interventions*

Il résulte de la différence systématique dans la façon de recueillir l'information des événements entre les deux groupes comparés, et entraîne donc une erreur de classification des exposés/non exposés à l'origine d'une erreur dans l'estimation de l'effet de l'exposition sur la maladie. Ces biais d'évaluation sont identiques à ceux rencontrés dans toutes les études conduites en ouvert. L'erreur de classification en soi n'est pas un problème tant qu'elle reste aléatoire. Par contre, si elle est systématique en faveur d'une catégorie de patients, elle peut modifier le sens de l'association exposition-maladie d'où la notion de biais. Un biais de classification est dit

différentiel lorsqu'il affecte différemment les groupes à comparer. Il se produit quand la classification erronée du statut d'intervention est liée au résultat ou au risque de résultat. Ce biais est dit non-différentiel quand il affecte de la même manière les deux groupes à comparer. Il n'est pas lié aux résultats et généralement il biaise l'effet estimé de l'intervention en le rapprochant du zéro.

A titre d'exemples de biais de mesure dans les études observationnelles, on inclut :

- Le biais « d'investigateur » : il survient lorsque les investigateurs interrogent ou recueillent différemment les informations des patients exposés et non exposés sur l'événement d'intérêt
- Le biais lié à la qualité des données : ce biais survient quand les données concernant les exposés sont meilleures ou plus accessibles que celles qui concernent les non exposés.
- *Biais dans la mesure des résultats*

Biais introduit par des erreurs différentielles ou non différentielles dans la mesure des résultats. Il survient quand la mesure du critère de jugement n'est pas standardisée dans les deux groupes. Il survient aussi lorsque la connaissance du traitement peut influencer le diagnostic et il entraîne ainsi une modification dans la prise en charge (biais de suivi).

- *Biais dû à une déviation par rapport aux interventions prévues*

Ce biais survient lorsqu'il y a des différences systématiques dans les soins courants entre le groupe expérimental et le groupe comparateur. Ces différences reflètent des déviations par rapport à l'intervention prévue. L'évaluation du biais dans ce domaine dépendra du type d'effet recherché comme l'effet de l'adhésion à l'intervention ou l'effet du fait de commencer et de se conformer à l'intervention. Les déviations par rapport à l'intervention prévue ne sont pas importantes lorsqu'on étudie l'effet de l'adhésion à l'intervention. Par exemple certaines personnes ne répondent pas aux interventions de dépistage.

- *Biais dû à des données manquantes*

Biais dû à l'absence de suivi au cours de l'étude des patients inclus et suivis initialement (comme la perte de suivi due à des facteurs de pronostic). Ce biais provient de l'exclusion des personnes ayant des informations manquantes sur le statut de l'intervention ou d'autres variables comme les facteurs de confusion

- *Biais dans la sélection des résultats rapportés (biais de déclaration)*

Ce biais provient de la sélection des résultats rapportés en fonction des conclusions des études. Ce biais empêche l'estimation d'être incluse dans une méta-analyse (ou autre synthèse).

L'existence de ces différents biais, souligne l'importance des choix méthodologiques opérés dans les études observationnelles afin d'optimiser le niveau de preuve. Les approches et les méthodes de correction des biais peuvent être différentes selon la question posée dans l'étude. Dans ce contexte, des méthodes d'ajustement ont été développées afin de contrôler ces biais.

#### *B/ Méthodes de minimisation des biais par ajustement*

Des méthodes d'ajustement sont utilisées afin de contourner le problème des facteurs de confusion et minimiser le biais de sélection lié au choix de traitement dans les études observationnelles. Les méthodes d'ajustement utilisées sont : 1/ L'analyse de régression multiple (régression linéaire, régression logistique, régression de Cox), qui est choisie s'il existe une forte corrélation entre le facteur de confusion et le résultat. 2/ La méthode du score de propension, qui est utilisée afin de limiter le nombre de variables d'ajustement ou le nombre de variables sur lesquelles on veut matcher.

Ces méthodes d'ajustement permettent une distribution comparable des variables de base mesurées entre les sujets traités et non traités, pour rendre les deux groupes de traitement comparables<sup>(43)</sup>.

L'utilisation d'un modèle de régression multiple permet un ajustement sur les facteurs de confusion potentiels connus et mesurés. Ce modèle permet de déterminer l'association entre un traitement et un résultat après prise en compte des autres variables explicatives. Toutes les variables pouvant potentiellement influencer le choix du traitement ou le résultat sont intégrées dans le modèle. Le choix du modèle dépend de la nature de la variable à expliquer. Dans le cas d'une variable continue, une régression linéaire est d'usage et l'estimation se fait par la méthode de moindre carré. Lorsque la variable est qualitative, l'estimation est faite par la méthode du maximum de vraisemblance dans le cadre d'une régression logistique. Une variable discrète nécessite un modèle de comptage tel que le modèle de poisson. Le résultat dans les trois types d'estimations cités précédemment est la probabilité de survenue de l'événement. En revanche, il existe une autre méthode de régression appelée régression de Cox dont la variable expliquée s'exprime en temps passé avant la survenue d'un événement. Cette régression permet aux taux de risques (survenue d'un événement) de varier dans le temps, mais elle suppose qu'ils sont

proportionnels entre les deux groupes, traités et non traités<sup>(43)</sup>.

L'ajustement peut également être réalisé via un score de propension (SP). Fréquemment utilisée, cette méthode a été initialement développée par Rosenbaum et Rubin en 1983. L'objectif du SP est de synthétiser toutes les caractéristiques observables des patients (clinique et socio démographiques) en un paramètre unique dont la valeur permettra de prédire le traitement choisi pour chaque patient. Il permet ainsi d'assurer la comparabilité de deux groupes<sup>(44)</sup>. Pour un individu (i), le SP correspond à la probabilité de recevoir le traitement étudié ( $Z=1$ ), connaissant ses caractéristiques observables connues à l'inclusion ( $X_i$ )<sup>(45)</sup>

$$e(x_i) = P(Z_i = 1 / X_i = x_i)$$

Le calcul du score de propension se fait en 3 étapes. La première étape consiste à sélectionner exhaustivement les facteurs significativement associés au choix du traitement évalué. La deuxième étape est de calculer, en analyse univariée, le SP avec les facteurs principaux. Dans une troisième étape, les variables significativement associées au choix thérapeutique sont entrées dans un modèle de régression logistique afin de ne conserver que les plus importantes<sup>(43)</sup>.

Ensuite, le SP calculé est intégré comme variable d'ajustement dans l'estimation de l'association exposition au traitement et risque de survenue de l'événement, afin de limiter un potentiel biais de sélection lié au choix du traitement. Ce modèle est identique aux modèles de régression standard, le résultat étudié est considéré comme une variable dépendante alors que l'effet de traitement et le SP sont estimés comme des variables indépendantes. L'effet du traitement est ainsi ajusté sur le SP et indirectement sur toutes les caractéristiques des patients inclus dans le SP<sup>(43,46)</sup>.

Cette méthode d'ajustement sur SP est identique aux modèles de régression standard en cas de déséquilibre sur un facteur dans un ECR. Cette méthode de SP est assez adaptée en cas d'un effectif de l'échantillon réduit. En effet, le SP permet de résumer plusieurs variables en une seule variable. Néanmoins, si l'effectif est important, cette méthode d'ajustement sur SP fournira les mêmes conclusions que le modèle de régression multiple. L'ajustement sur SP est plus robuste que le modèle de régression classique car, entre autres, il n'est pas limité par le nombre de variables à tester dans un modèle de régression.

La première limite de cette méthode est qu'elle suppose que le modèle de régression est correctement spécifié. Sa deuxième limite est que le facteur de confusion du traitement est souvent aussi un facteur de risque de l'événement. Pour éviter d'introduire des facteurs corrélés en variables explicatives du modèle, il faut un SP résumant plusieurs facteurs de confusion. Enfin, l'ajustement sur SP a aussi une limite en commune avec les modèles multivariés : l'estimation de l'effet traitement est basée sur un OR (odds ratio). Ce dernier n'est pas toujours un bon estimateur du RR (risque relatif). Il existe donc 3 méthodes d'utilisation du SP présentées ci-dessous :

### *C/ Autres méthodes d'utilisation du score de propension*

Le SP peut être utilisé pour faire de **l'appariement (Matching)** : le principe de cette méthode est de créer des groupes comparables en termes de probabilité de recevoir le traitement. Il s'agit d'apparier chaque patient du groupe traité avec le patient du groupe non traité présentant le SP le plus proche. Une fois l'appariement effectué, une estimation de l'effet du traitement dans la population appariée est calculée. Cette méthode est intéressante dans les études conduites sur des grandes base de données. Néanmoins, l'exclusion des patients non appariés dans ce modèle, entraîne une perte d'informations et de précision<sup>(44,45,47,48)</sup>.

Le matching peut également se faire par **stratification** : contrairement à l'appariement sur SP, l'approche de stratification utilise l'ensemble des observations disponibles en partitionnant les individus traités et non traités suivant les valeurs estimées du SP. Elle consiste à diviser l'échantillon en strates de même effectif. Par exemple pour un échantillon de 1000 patients, l'échantillon est divisé en 10 strates de 100 patients. Au sein de chaque strate, les individus ont des valeurs similaires de SP. L'estimation de l'effet moyen du traitement s'appuie sur une analyse au sein de chaque strate, puis du calcul de l'effet moyen agrégé. Contrairement à la méthode d'appariement, l'analyse stratifiée sur le SP a l'avantage de ne pas exclure des patients<sup>(43,44,47,48)</sup>.

Enfin, le SP peut être utilisé comme facteur de pondération avec la **méthode de pondération inverse (IPTW : inverse probability of treatment weighting)**. Cette méthode consiste en la création d'une pseudo-population dans laquelle l'effet du traitement est estimé en donnant des poids différents aux patients (d'où la création d'une population virtuelle) : les patients traités auront un poids égal à l'inverse du score de propension  $p(X)$ , et les patients non traités par  $(1-p(X))$ . L'estimation de l'effet du traitement s'effectue directement par régression linéaire<sup>(43,46-48)</sup>.

L'utilisation des méthodes d'ajustement a permis aux études observationnelles de se rapprocher des conditions de comparabilité des ECR et d'avoir un schéma d'étude qui ressemble celui des ECR. Mais ces différentes approches ne peuvent être recevables que si l'existence d'une confusion résiduelle est écartée.

#### *D/ Notion de confusion résiduelle*

La prise en compte de facteurs d'ajustement ou d'un SP ne corrige pas systématiquement l'influence de facteurs de confusion. En effet, le SP ne prend pas en considération les facteurs de confusion non observés et non inclus dans le modèle utilisé pour estimer le SP. Ce schéma d'étude ajustée ne corrige donc que les facteurs de confusion observés<sup>(48)</sup>. Par conséquent, l'absence de biais n'est jamais garantie avec ces méthodes d'ajustement.

En résumé, les études observationnelles utilisent des données de vie réelle pour estimer le bénéfice médical d'un traitement. Ces études observationnelles sont sujettes au biais de sélection/confusion dû à l'absence de randomisation. Même si des méthodes d'ajustement statistique sont appliquées pour contrôler ce biais, elles ne permettent pas de garantir l'absence de confusion résiduelle. Quelles peuvent être alors les conséquences de l'utilisation des études observationnelles sur l'estimation de l'effet traitement ?

### 1.3 Les études randomisées et les études observationnelles concernant les anticoagulants oraux directs dans le traitement des ETEV et dans la FA

Afin de répondre à ces problématiques, nous allons comparer les ECR et les études observationnelles évaluant les anticoagulants oraux directs (AOD) par rapport aux antivitamines K (AVK) dans le traitement des événements thromboemboliques veineux (ETEVE) et dans la fibrillation auriculaire (FA).

#### 1.3.1 La maladie thromboembolique veineuse : épidémiologie et impacts pour les patients

La maladie thromboembolique veineuse (MTEV), définie par une thrombose veineuse profonde (TVP) et/ou une embolie pulmonaire (EP) est une maladie fréquente. La TVP est l'obstruction plus ou moins complète d'une veine profonde par un thrombus constitué in situ. La stase veineuse et les lésions endothéliales prédisposent à la thrombose. L'histoire naturelle

de la TVP est indissociable de celle de l'EP. D'où le concept de MTEV qui est donc une entité clinique. Son incidence augmente exponentiellement avec l'âge : elle est de moins de 1 cas/1000 personnes/année chez les patients de moins de 50 ans, elle s'élève à plus de 6 cas / 1000 personnes/année chez les patients de plus de 80 ans<sup>(49)</sup>. La mortalité de la MTEV est également plus élevée chez le sujet âgé : environ 4% à 1 mois chez les patients de moins de 65 ans, elle s'élève à 11% chez les patients de plus de 65 ans<sup>(50)</sup>.

Concernant la prise en charge diagnostique et thérapeutique de la MTEV, des recommandations de bonne pratique ont été actualisées en 2019 sous l'égide de la Société de Pneumologie de Langue Française (SPLF)<sup>(51)</sup>.

- Le diagnostic de MTEV est difficile car les signes cliniques sont inconstants et non spécifiques. La stratégie diagnostique se fait alors en plusieurs étapes : évoquer l'hypothèse, estimer la probabilité clinique, prescrire des examens complémentaires adaptés (D-dimères, angioscanner spiralé thoracique, échographie veineuse des membres inférieurs).
- Le traitement de la MTEV est l'anticoagulation efficace qui a pour but de prévenir une récurrence. L'anticoagulation initiale repose soit sur les AOD d'emblée soit sur les héparines de bas poids moléculaire, le fondaparinux ou l'héparine non fractionnée. Les AOD sont désormais recommandés en première intention. Ils sont plus simples d'utilisation (administration orale à dose unique sans contrôle biologique). La durée du traitement anticoagulant est variable selon le contexte. En présence de facteurs majeurs transitoires, la durée sera courte (3 à 6 mois maximum). Néanmoins, en présence de facteurs persistants (cancer) ou en l'absence de facteurs majeurs, le traitement sera long (6 mois voire durée non limitée).

### 1.3.2 La fibrillation atriale : épidémiologie et impacts pour les patients

- *Données épidémiologiques majeures*

La FA est l'arythmie cardiaque la plus fréquente. Elle représente environ un tiers des hospitalisations dues à des troubles cardiaques<sup>(52)</sup>. Ce trouble du rythme supraventriculaire peut être suspecté cliniquement mais il est diagnostiqué par un électrocardiogramme (ECG). Il est souvent associé à des facteurs de risque non cardiovasculaires (maladie thoracique, infection) et des pathologies cardiovasculaires comme l'hypertension, le diabète, l'insuffisance cardiaque, les maladies vasculaires et la cardiopathie valvulaire. Ces pathologies augmentent à leur tour le

risque de complications lié à l'arythmie<sup>(53,54)</sup>. L'étude ARIC (Atherosclerosis Risk in Communities) a montré aussi que 56.5% des nouveaux cas de FA peuvent être associés à des facteurs de risque cardiovasculaires tels que l'hypertension, le diabète, l'obésité et le tabagisme<sup>(55)</sup>.

La prévalence et l'incidence augmentent partout dans le monde avec le vieillissement des populations et l'allongement de l'espérance de vie. Ainsi, elle touche particulièrement les personnes âgées. Plus élevée aux États-Unis, en Australie et en Europe (1% de la population adulte), la prévalence et l'incidence de la FA sont généralement moins élevées ou sous-estimées dans les pays en voie de développement<sup>(54,56)</sup>. La prévalence mondiale de la FA est de 37 574 millions de cas (0,51 % de la population mondiale), et a également augmenté de 33 % au cours des 20 dernières années<sup>(57)</sup>.

En 2010, la FA est une épidémie mondiale qui a touché environ 33,5 millions d'individus (596 pour 100 000 hommes et 373 pour 100 000 femmes)<sup>(52,58)</sup>. En 2017, un total de 3,046 millions nouveaux cas de FA dans le monde ont été enregistrés avec un taux d'incidence estimé à 403/millions d'habitants. En France, entre 600 000 et un million de personnes seraient atteintes de FA, dont deux tiers de plus de 75 ans. L'incidence de la maladie est estimée entre 110 000 et 230 000 nouveaux cas par an<sup>(59)</sup>. Il est prévu qu'avec le vieillissement de la population mondiale, la FA affectera 6 à 12 millions de personnes aux États-Unis en 2050 et 17,9 millions en Europe en 2060<sup>(60,61)</sup>. Les projections futures suggèrent ainsi que la charge absolue de la FA pourrait augmenter de plus de 60 % en 2050<sup>(57)</sup>. Cette augmentation de prévalence peut être expliquée par l'amélioration du dépistage avec des programmes de dépistages ciblés et la prévention des facteurs de risques spécifiques comme le suggèrent les données de l'étude Framingham Heart Study<sup>(62)</sup>. Néanmoins, la FA pourrait être non-diagnostiquée, en particulier chez les sujets âgés et chez les patients avec insuffisance cardiaque<sup>(54)</sup>. Malgré le fait que la prévalence de la FA soit plus faible chez les femmes, le risque de décès lié à cette pathologie est plus élevé<sup>(54)</sup>. La gravité de la FA est liée au risque d'insuffisance cardiaque par altération du remplissage ventriculaire et dépend de la cardiopathie sous-jacente<sup>(53)</sup>.

La FA reste une des principales causes d'accident vasculaire cérébral, de défaillance cardiaque, de décès et de morbidité cardiovasculaire dans le monde, malgré l'amélioration de la prise en charge des patients qui en sont atteints. Elle est un facteur de risque indépendant et important de survenue d'accidents thromboemboliques artériels (TEA) : l'accident vasculaire cérébral (AVC) et l'embolie systémique artérielle. Ils surviennent à la suite de la migration d'un thrombus intra-auriculaire formé par stagnation du sang, conséquence de l'altération de la contraction cardiaque. Elle augmente ainsi de cinq fois plus le risque d'un accident vasculaire cérébral (AVC) par rapport aux personnes qui ne souffrent pas de FA. Plusieurs études épidémiologiques ont montré que 20 à 30 % des patients victimes d'un AVC ont une FA diagnostiquée avant, pendant ou après la survenue de ces accidents<sup>(63-65)</sup>.

La FA augmente également la sévérité des AVC, le risque de décès et le risque de développer une insuffisance cardiaque. Aux États-Unis, 140 000 décès par AVC sont répertoriés chaque année<sup>(66)</sup>. En France, 31 346 décès également par AVC ont été estimés en 2013 et 141 652 personnes domiciliées avaient été hospitalisées pour AVC ou accident ischémique transitoire en 2014. Les trois quarts de ces patients étaient âgés de 65 ans et plus.

- *Impacts pour les patients et le système de santé*

De point de vue économique, la FA et ses complications engendrent des coûts élevés de prise en charge médicamenteuse. Des données provenant des États-Unis et du Royaume-Uni indiquent que la FA est considérée comme un problème de santé publique coûteux<sup>(54)</sup>. Les coûts directs liés à cette prise en charge étaient estimés entre 6 et 26 milliards de dollars US aux États-Unis en 2010<sup>(67)</sup> et elles représentaient environ 1 % du total des dépenses de santé au Royaume-Uni en 2008<sup>(68)</sup>. Avec plus de 795 000 AVC par an, le coût annuel des AVC chez les patients avec FA est estimé aux États-Unis à 34 milliards de dollars<sup>(69)</sup>. Le coût moyen de la prise en charge hospitalière d'un épisode d'AVC en France est de l'ordre de 5 000 à 30 000 euros. Le coût dépend de la gravité de l'accident<sup>(70)</sup>.

La prise en charge de la FA est donc un défi majeur pour les systèmes de santé<sup>(54)</sup>. Cette prise en charge permanente commence par les 3 étapes suivantes : l'identification, la prévention et le traitement des facteurs de risque et des maladies associées. Afin de réduire les symptômes associés à la FA, le contrôle de la fréquence cardiaque et du rythme cardiaque font partie intégrante de cette prise en charge. Dans le même sens, l'enjeu principal du traitement médicamenteux de la FA reste la prévention de la survenue des événements TEA par un

traitement AOD. Dans ce sens, plusieurs études randomisées et épidémiologiques ont été conduites pour évaluer l'efficacité et la sécurité de ces traitements AOD chez les patients atteints de FA pour la prévention de ces accidents.

### 1.3.3 Contexte du développement des AOD dans la MTEV et la FA

#### *A/ Propriétés pharmacologiques des Antivitamines K*

Les inhibiteurs de la synthèse des facteurs vitamine K dépendants ou plus communément (mais termes non appropriés) appelés les AVK sont des anticoagulants oraux "indirects" qui inhibent la synthèse hépatique des facteurs de la coagulation vitamine K dépendants (facteurs II, VII, IX, X, protéines C et S) (Figure 1)<sup>(71)</sup>. Le mécanisme d'action des AVK est un mécanisme indirect donc plus long, ils nécessitent donc un certain délai avant d'être efficaces et ne sont donc pas utilisables seuls en urgence mais en relais d'un autre traitement anticoagulant. Parmi les éléments essentiels des propriétés pharmacocinétiques des AVK, on note une forte liaison aux protéines plasmatiques et un métabolisme hépatique important. En effet, l'hypoprothrombinémie induite par les AVK est obtenue dans les 24 à 36 heures après leur administration et leur effet peut persister quelques jours après l'arrêt du traitement. Les AVK sont éliminés avec un fort passage hépatique (mixte et majoritairement rénale)<sup>(72)</sup>. De ce fait, ils sont contre-indiqués en cas d'insuffisance hépatique sévère. Le tableau 1 ci-dessous résume les principales caractéristiques pharmacocinétiques. Les médicaments appartenant à la classe des AVK sont au nombre de trois : la fluindione (demi-vie d'élimination est  $\geq 31$  heures, Préviscan®), la warfarine (demi-vie d'élimination est  $\geq 35$  heures, Coumadine®), et l'acénocoumarol (demi-vie d'élimination est de 8 à 11h, Sintrom®).

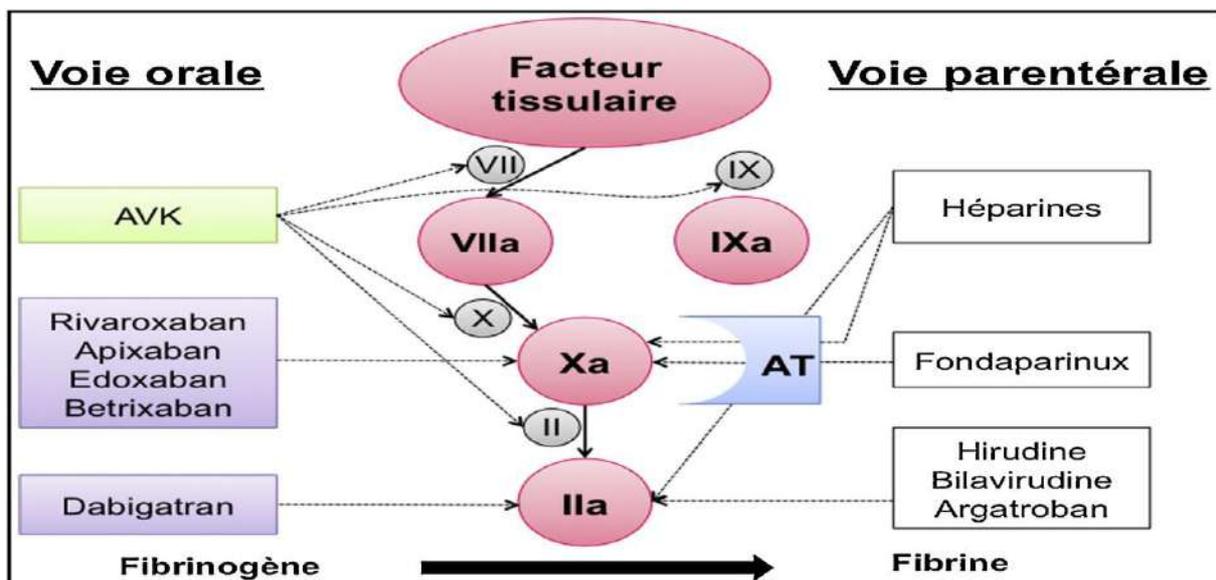
Les AVK sont notamment indiqués dans le traitement des TVP/ EP en relais d'une héparinothérapie, dans la prévention de leur récurrence ainsi que la prévention des accidents thromboemboliques artériels chez les patients avec FA non valvulaire, mais aussi dans la prévention en cas de valve mitrale. Pour rappel, la FA non valvulaire est définie comme une FA chez un patient n'ayant pas de valvulopathie mitrale rhumatismale, pas de prothèse valvulaire cardiaque et/ou pas eu de réparation valvulaire.

La première particularité de l'effet pharmacologique de ce traitement est la présence d'une importante **variabilité inter-individuelle** (liaison aux protéines plasmatiques et métabolisme hépatique) **et intra-individuelle**. Cette variabilité est problématique car les AVK ont un index

thérapeutique étroite (entre la dose anticoagulante et la dose hémorragique), toute bascule dans une sous ou sur-exposition au traitement est donc un risque d'événement thrombotique clinique pertinent (récidive thromboembolique dans le cas de la MTEV, AVC et embolie systémique dans le cas de la FA) ou d'hémorragie majeure. De ce fait, la dose d'équilibre d'un patient doit être déterminée en adaptant la dose en fonction de la mesure standardisée du temps de coagulation, l'International Normalized Ratio (INR) égal à 1 chez un sujet sain sans anticoagulation). Cet INR doit être compris entre 2 et 3 pour chacune des indications (MTEV et FA), puisque c'est lorsque l'INR se situe dans cet index que le risque d'événements indésirables est le plus faible. Une diminution des posologies associée à une surveillance clinique et biologique accrue est recommandée en cas d'insuffisance hépatique légère ou modérée car elle est associée à une augmentation du risque de surdosage, du fait de la diminution du métabolisme mais aussi de la synthèse des facteurs de la coagulation.

De **nombreuses interactions** avec ces molécules ont été décrites : médicamenteuses et alimentaires, du fait de leurs propriétés pharmacologiques (liaison aux protéines plasmatiques, métabolisme hépatique). Ces interactions peuvent être responsables d'une difficulté d'obtention un INR stable. Ainsi, il existe un risque de sous ou surdosage. En cas de surdosage, l'effet indésirable le plus fréquent de cette classe pharmacologique est le risque hémorragique avec notamment la survenue d'hémorragies intracrâniennes<sup>(73)</sup>. Il peut être alors nécessaire d'avoir recours à l'antidote des AVK, la vitamine K, qui permet un retour progressif de la synthèse des facteurs de la coagulation vitamine K dépendants.

Figure 1: Cibles pharmacologiques des anticoagulants



Source : Presse Med. 2013- PMID :42 : 1206–1212 ; AT : antithrombine

Compte tenu de ces contraintes pharmacologiques, l'anticoagulation des patients doit ainsi faire l'objet d'un **suivi biologique** régulier par le dosage de l'INR et d'une adaptation posologique. Ceci rend le traitement compliqué dans sa gestion, autant pour les patients que pour les médecins, avec des conséquences importantes en termes d'événement clinique dès que le patient sort de l'index thérapeutique.

### *B/ Développement des anticoagulants Oraux Directs*

Les AOD ou DOAC en anglais (Direct Oral AntiCoagulants), ont été développés comme une alternative aux AVK. Les AOD ont obtenu une autorisation de mise sur le marché dès 2008 en France. Contrairement aux AVK, les AOD agissent de façon spécifique et directe sur les facteurs de la coagulation activés (Figure 1). Ces médicaments ont une faible variabilité individuelle et présentent une marge thérapeutique large, donc l'opposé des AVK. Il n'y a donc pas de délai d'action. Le tableau 1 résume les principales caractéristiques pharmacocinétiques. En fonction de la cible d'action, ils sont divisés en deux classes :

- Les inhibiteurs directs de la thrombine ou anti-IIa. La thrombine est l'enzyme clé de la coagulation qui transforme le fibrinogène en fibrine. Parmi les anti-thrombines, on trouve l'argatroban, les hirudines et pour les AOD, le dabigatran étextilate (PRADAXA®). Cette dernière molécule est la seule molécule active par voie orale disponible à ce jour.
- Les inhibiteurs directs du facteur Xa ou anti-Xa. Le facteur Xa est aussi la cible de choix des héparines et des pentasaccharides (fondaparinux, idraparinux). Il existe quatre représentants parmi les AOD : le rivaroxaban (XARELTO®), l'apixaban (ELIQUIS®), l'edoxaban (LIXIANA®) et le bétrixaban (BEVYXXA®)

Ces molécules se caractérisent par une pharmacocinétique dépendante de transporteurs et un métabolisme hépatique sous l'action des cytochromes P450 pour certains d'entre eux<sup>(74)</sup>. Le métabolisme hépatique ne constitue pas une voie d'élimination exclusive comme pour les AVK, l'élimination se fait majoritairement sous forme active pour le dabigatran et sous forme mixte de métabolites ou active pour le rivaroxaban et apixaban. Du fait de cette élimination, leur utilisation est contre-indiquée en cas d'insuffisance rénale sévère (clairance de la créatinine inférieure à 30 mL/min selon Cockcroft et Gault) et en cas d'insuffisance hépatique.

Contrairement aux AVK, les AOD :

- ont une dose fixe de prescription
- ne nécessitent pas de surveillance biologique de routine
- ont un délai d'action rapide
- ont une demi-vie plus courte
- sont commercialisés sans antidote (sauf l'antidote spécifique du dabigatran a été commercialisé depuis 2016)<sup>(75)</sup>

Tableau 1: Principales caractéristiques pharmacocinétiques

	<b>Antivitamines K</b>	<b>Anticoagulants oraux directs</b>		
<b>Molécules</b>	Warfarine (COUMADINE®) Acénocoumarol (SINTROM®, MINISINTROM®) Fluindione (PREVISCAN®)	Dabigatran (PRADAXA®)	Rivaroxaban	Apixaban
<b>Cibles pharmacologiques</b>	Inhibition des facteurs vitamine K dépendants (II, VII, IX, X) et des protéines inhibitrices C et S	Inhibiteur direct de la thrombine	Inhibiteur direct du facteur Xa de la coagulation	
<b>Demi-vie</b>	≥35 heures (warfarine) ≥31 heures (fluindione) 8 à 11h (acénocoumarol)	7 à 17 heures	7 à 11 heures	8 à 115 heures
<b>Élimination</b>	Mixte, majoritairement rénale	80 % rénale, forme active 20 % rénale et biliaire, métabolites	36 % rénale, forme active, 32 % rénale, métabolites, 32 % fécale, métabolites	25 % rénale, forme active
<b>Surveillance biologique</b>	Oui, mesure de l'INR	Non	Non	Non
<b>Antidote</b>	Vitamine K	Anticorps monoclonal idarucizumab	Non	Non

	<b>Antivitamines K</b>	<b>Anticoagulants oraux directs</b>
<b>Indication selon libellé de l'AMM dans l'indication de la MTEV</b>	Traitement curatif des thromboses veineuses et des EP, ainsi que dans la prévention de leur récurrence, en relais de l'héparine.	Traitement des TVP et des EP, et prévention des récurrences de TVP et d'EP chez l'adulte.
<b>Indication selon libellé de l'AMM dans l'indication de la FA</b>	Traitement préventif de la thrombose ou de l'embolie systémique dans les situations à risque suivantes : -cardiopathies emboligènes (FA paroxystique, valvulopathies mitrales rhumatismales, prothèses valvulaires), -infarctus du myocarde compliqué, syndrome des anticorps antiphospholipides (en cas de thrombose)	Prévention de l'AVC et de l'embolie systémique chez les patients adultes atteints de FA non valvulaire et présentant un ou plusieurs facteur(s) de risque tels que : antécédent d'AVC ou d'accident ischémique transitoire, âge $\geq 75$ ans, hypertension artérielle, diabète, insuffisance cardiaque symptomatique (classe NYHA $\geq$ II).

Source : Vidal

Les AOD sont indiqués dans le traitement des TVP et des EP, la prévention de leur récurrence ainsi que la prévention des TEA chez les patients avec FA non valvulaire. Ils sont également indiqués dans la prévention primaire des ETEV en chirurgie programmée pour prothèse totale de genou ou de hanche.

#### 1.3.4 Données disponibles avec les AOD dans la MTEV

Nous allons présenter ici les grands essais ayant permis pour la plupart la mise sur le marché de ces AOD dans la MTEV ainsi que les études post-inscription.

## *A/ Les grands essais randomisés et méta-analyse*

Les AOD ont été évalués par six ECR dans le traitement des ETEV. Ces essais ont comparé l'efficacité (récidives de ETEV) et la sécurité (hémorragies) des AOD par rapport aux AVK dans le traitement des ETEV. Le taux d'hémorragie prend en compte 2 types d'hémorragies distincts : les hémorragies majeures et les hémorragies non-majeures mais cliniquement significatives. Selon la définition ISTH (International Society on Thrombosis and Haemostasis)<sup>(76)</sup>, une hémorragie majeure est définie par toute hémorragie aiguë extériorisée accompagnée d'au moins un des éléments suivants : chute de 2 g/dL du taux d'hémoglobine, transfusion de deux concentrés, touchant un organe critique ou entraînant le décès. L'hémorragie non-majeure mais cliniquement significative est une hémorragie modifiant la prise en charge médicale du patient en dehors des critères, cités ci-dessus d'une d'hémorragie majeure. Dans les paragraphes suivants, pour chacune des molécules, nous allons faire un bref rappel de ces études pilotes.

- **Le dabigatran** (inhibiteurs directs de la thrombine) a été évalué dans deux études: RECOVER<sup>(77)</sup> et RE-COVER II<sup>(78)</sup>, ayant inclus au total 5107 patients avec MTEV. Le plan expérimental des deux études était un essai en double aveugle de non-infériorité. Le dabigatran était évalué (vs AVK) uniquement dans le traitement d'entretien. Ainsi, tous les patients inclus avaient reçu au moins 5 jours de traitement parentéral par héparine avant d'être randomisés dans le groupe dabigatran 150 mg 2 fois par jour ou le groupe AVK pour six mois. En total, le dabigatran est non-inférieur aux AVK en termes de risque de récidives de ETEV. Néanmoins, il est associé à une baisse du risque d'hémorragie majeure.
- **Le rivaroxaban** (inhibiteurs du facteur Xa) a été étudié dans 2 études principales : l'étude EINSTEIN-TVP<sup>(79)</sup> ayant inclus de 3449 patients avec TVP et l'étude EINSTEIN-EP<sup>(80)</sup> ayant inclus de 4832 patients avec EP. Le plan expérimental utilisé dans chacune de ces deux études était un essai ouvert de non-infériorité. Le rivaroxaban (15 mg 1 fois par jour pendant 3 semaines puis 20 mg/jour) était non-inférieur au traitement standard (enoxaparine/AVK) en termes de récidives thromboemboliques. Chez les patients avec TVP, il n'y avait pas de différence entre AOD et AVK en termes d'hémorragie majeure. Cependant, chez les patients ayant une EP, le traitement par rivaroxaban était associé à une réduction du risque d'hémorragie majeure.
- **L'apixaban** (inhibiteurs du facteur Xa) a été évalué dans le traitement des ETEV dans un essai de non infériorité: AMPLIFY<sup>(81)</sup>. Contrairement au rivaroxaban, l'étude est réalisée

en double-aveugle incluant 5395 patients avec TVP et/ou EP. L'apixaban est utilisé comme traitement unique à la dose de 10 mg 2 fois par jour pendant les sept premiers jours, puis à dose d'entretien de 5 mg 2 fois par jour versus un traitement conventionnel.

- **L'édoxaban** (inhibiteurs du facteur Xa) a été étudié dans le traitement des ETEV dans un essai de non infériorité en double-aveugle: HOKUSAI<sup>(82)</sup>. 8240 patients avec TVP et/ou EP ont été inclus. Comme le dabigatran, l'edoxaban n'a pas été testé comme traitement unique, mais seulement versus AVK à la dose 60 mg une fois par jour, après une phase initiale par héparinothérapie.

Tous ces ECR ont montré la non-infériorité des AOD par rapport aux AVK dans le traitement des ETEV, avec des bornes de non-infériorité différentes d'une étude à l'autre. Les AOD étaient associés à une meilleure sécurité et à une diminution du risque d'hémorragies majeures. Par conséquent, les AOD semblent être une alternative possible aux AVK, sans monitoring biologique, avec un bénéfice en termes de risque d'hémorragie majeure.

Afin de réaliser une synthèse des résultats des ECR présentés ci-dessus, une méta-analyse a été conduite<sup>(83)</sup>. Elle a étudié l'efficacité et la sécurité des AOD par rapport aux AVK dans le traitement des ETEV. Il s'agit d'une méta-analyse réalisée sur 27 096 patients inclus dans 6 ECR conduites entre 2006 et 2012. Cette méta-analyse a montré que l'effet des AOD était similaire à l'effet des AVK en termes d'efficacité (HR = 0.91 [0.79; 1.06] sur le risque de récurrences de ETEV si on retient la borne de non-infériorité la plus stricte de ces 6 essais c'est-à-dire HR=1.5). De plus, les AOD réduisent le risque d'hémorragies majeures par rapport au traitement standard (HR = 0.62 [0.45 ; 0.85]).

Deux autres méta-analyses ont aussi évalué l'efficacité et la sécurité des AOD dans le traitement des ETEV<sup>(84,85)</sup>. Néanmoins, ces deux méta-analyses n'ont pas inclus les 6 essais et avaient un nombre plus limité de sujets. Elles ont montré également que les AOD ont une efficacité comparable aux AVK avec une réduction significative des hémorragies majeures.

### *B/ Quelques études observationnelles*

La transposabilité dans la pratique clinique courante des résultats des ECR reste l'objectif principal des études observationnelles. Il s'agit donc de déterminer si les résultats de ces études (AOD vs AVK) réalisées sur des populations de patients plus hétérogènes vont être similaires à ceux des essais de phase III. Plusieurs études observationnelles (18 études) ont été conduites

afin de comparer et confirmer le bénéfice médical des AOD par rapport aux AVK dans le traitement des ETEV.

Les différents types d'études observationnelles post-AMM conduites sur ce sujet étaient : des cohortes prospectives, des cohortes rétrospectives et des études sur bases médico-administratives.

- Toutes les cohortes prospectives post-AMM disponibles ont comparé le rivaroxaban aux AVK dans le traitement des ETEV. Cet AOD était associé à une réduction du risque des ETEV et du risque des hémorragies majeures.
- Les cohortes rétrospectives et les études sur bases médico-administratives ont évalué les quatre AOD (rivaroxaban, le dabigatran, l'apixaban et l'edoxaban). Les AOD étaient associés à une réduction du risque de ETEV et une réduction du risque d'hémorragies majeures dans plus de la moitié des études.

Globalement, les AOD aussi bien dans la plupart des études observationnelles que dans les ECR, étaient associés à une meilleure sécurité. Cependant, et contrairement aux ECR, les AOD étaient associés à une meilleure efficacité dans les études observationnelles. En effet, les résultats en termes d'efficacité des études observationnelles ne sont pas similaires aux résultats des ECR, probablement du fait du biais de sélection/confusion associé à l'absence de randomisation.

Ces études observationnelles ont utilisé des méthodes d'ajustement afin de corriger le biais de sélection et l'influence des facteurs de confusion. On note douze études ayant ajusté le résultat avec un SP, quatre études ayant ajusté le résultat avec des méthodes de régression multivariée et deux études sans méthode d'ajustement. Malgré l'utilisation de ces méthodes d'ajustement, il persiste une discordance des résultats entre les études observationnelles et les ECR en termes d'efficacité, qui peut être expliquée soit par une non transposabilité des résultats à la population cible, soit par la persistance des facteurs de confusion résiduels dans les études observationnelles.

### 1.3.5 Données disponibles avec les AOD dans la FA

Nous allons faire la même synthèse avec les études conduites dans la FA.

#### *A/ Les 4 grands essais randomisés et méta-analyse*

Quatre ECR de phase III ont été conduits afin d'évaluer les AOD sur de grands effectifs chez des patients avec une FA non valvulaire. L'évaluation de l'efficacité de ces molécules porte sur la survenue des TEA alors que l'évaluation de leur sécurité porte sur le taux d'hémorragies.

- Le **dabigatran** a été étudié dans l'essai RE-LY<sup>(86)</sup>. L'essai a comparé aux AVK, deux doses de dabigatran (110 mg et 150 mg) chez des patients avec une FA non valvulaire. Le plan expérimental utilisé dans cette étude était un essai en ouvert de non-infériorité. Cette étude a montré la supériorité du dabigatran 150 mg par rapport aux AVK en termes d'efficacité. Cette étude a montré aussi la non-infériorité du dabigatran 110 mg par rapport aux AVK en termes d'efficacité et sa supériorité en termes d'hémorragies majeures. Le dabigatran (aux deux posologies) est associé à une réduction des hémorragies intracrâniennes. Néanmoins, le caractère ouvert de cette étude rend incertain l'appréciation attendue de la quantité d'effet (avec un possible biais en sa faveur). Les résultats de l'essai RE-LY ont conduit à retenir aux États-Unis la posologies de 150 mg (75 mg si la clairance de la créatinine est entre 15 et 30 ml/min/1,73m<sup>2</sup>) et pour l'Union européenne, les doses de 150 mg et 110 mg<sup>(86)</sup>.
- Le **rivaroxaban** dont l'évaluation principale repose sur l'étude ROCKET AF<sup>(87)</sup>. Cette étude a comparé en double aveugle le rivaroxaban (20 mg ou 15 mg en cas d'insuffisance rénale modérée) aux AVK chez des patients avec une FA non valvulaire. Elle a établi la non-infériorité du rivaroxaban par rapport aux AVK en termes de survenue d'un AVC ou d'une embolie systémique. L'incidence des hémorragies majeures et/ou « non majeures cliniquement significatives » a été similaire dans les deux groupes. Dans le groupe rivaroxaban, il y a eu moins d'hémorragies mortelles/intracrâniennes/localisation critique que dans le groupe AVK<sup>(87)</sup>.
- L'**apixaban** a été évalué dans l'étude ARISTOTLE<sup>(88)</sup>. Cet essai a comparé en double-aveugle l'apixaban (5 mg ou 2,5 mg pour les patients > 80 ans, de < 60 kg ou insuffisance rénale modérée) aux AVK chez des patients avec une FA non valvulaire. Cette étude a démontré la supériorité de l'apixaban par rapport aux AVK en termes d'efficacité avec un

impact significatif sur la mortalité (critère secondaire) renforçant la crédibilité du résultat sur le critère principal. Une réduction significative du risque d'hémorragies majeures (incluant les hémorragies intracrâniennes) est aussi établie<sup>(88)</sup>.

- L'**edoxaban** a été évalué au cours d'un ECR de non infériorité en double-aveugle versus AVK dans la FA non valvulaire : ENGAGE AF-TIMI 48<sup>(89)</sup>. L'edoxaban (60 mg ou 30 mg chez les patients présentant certains facteurs de risque de surdosage) a été non-inférieur aux AVK en termes d'efficacité. Les hémorragies majeures (et les non majeures cliniquement pertinentes) sont survenues moins fréquemment sous edoxaban que sous AVK. L'edoxaban était associé à une réduction significative du risque d'hémorragie et de décès par maladie cardiovasculaire<sup>(89)</sup>.

Une méta-analyse a été réalisée à partir des résultats des quatre essais présentés ci-dessus ayant comparé les AOD aux AVK sur la prévention des accidents vasculaires cérébraux chez les patients avec FA non valvulaire<sup>(90)</sup>. Cette méta-analyse a permis de quantifier l'effet respectif des AOD dans leur ensemble par rapport aux AVK. Elle a été réalisée sur 71 683 patients inclus dans ces quatre grandes ECR conduites entre 2005 et 2010. Elle a montré que les AOD offrent un profil bénéfice/risque favorable par rapport aux AVK. En effet, les AOD réduisent significativement le risque d'AVC et d'embolie systémique (HR = 0.81 [0.73 ; 0.91]), les hémorragies intracrâniennes (HR = 0.48 [0.39 ; 0.59]) et la mortalité (HR = 0.90 [0.85 ; 0.95]) sans modifier significativement le risque d'hémorragies majeures. Néanmoins, les AOD seraient associés à une augmentation significative du risque d'hémorragies gastro-intestinales (HR = 1.25 [1.01 ; 1.55]). L'efficacité et la sécurité des AOD étaient relativement stables quelles que soient les caractéristiques des patients avec FA, comme cela a été observé dans d'autres travaux<sup>(91)</sup>. Les conclusions de cette méta-analyse offrent donc aux cliniciens une meilleure compréhension des AOD en tant qu'option thérapeutique pour réduire le risque embolique chez les patients avec une FA non valvulaire.

Plusieurs autres méta-analyses publiées ont évalué l'efficacité et la sécurité des AOD versus AVK chez les patients avec FA<sup>(91-93)</sup>. Néanmoins, la plus grande étude « ENGAGE AF-TIMI » de ces ECR n'a pas été incluse dans ces méta-analyse car plus récente.

### *B/ Un grand nombre d'études observationnelles*

De nombreuses études observationnelles ont été mises en place visant à évaluer l'usage des AOD versus AVK dans la prévention des AVC chez les patients avec FA en conditions réelles

d'utilisation. L'objectif de ces études a été et reste toujours la validation de la transposabilité des résultats des ECR à la pratique clinique courante<sup>(94,95)</sup>. Elles reflètent l'utilisation de ces médicaments dans la population générale de patients non sélectionnés. Ces études ont été majoritairement réalisées à partir de bases de données de santé (notamment des bases de remboursements nord-européennes et nord-américaines), mais aussi avec la constitution de registres ad hoc, dont ceux des sociétés savantes. La quasi-totalité des études observationnelles s'est concentrée sur les deux types d'événements évalués au cours des ECR : les événements TEA d'une part, et les événements hémorragiques d'autre part.

Les études observationnelles réalisées concluaient majoritairement à une efficacité et une sécurité des AOD au moins égales à celles des AVK. Cependant, ces études observationnelles ont été assez hétérogènes en termes de méthodologie et également en termes de résultats. Les résultats issus de ces études en termes d'efficacité et de sécurité sont même parfois contradictoires. Par rapport aux AVK, certaines études ont montré que les AOD étaient associés à une réduction du risque d'événements TEA et hémorragiques, alors que d'autres études ont montré l'inverse, et d'autres ont montré des taux d'événements similaires entre les deux groupes. Cette hétérogénéité des résultats inter-études observationnelles mais aussi leurs discordances avec les ECR peuvent être expliquées par la nature diverse des données utilisées et par les principes méthodologiques inhérents aux études observationnelles.

La majorité de ces études observationnelles ont utilisé des méthodes d'ajustements statistiques afin de minimiser les biais de sélection et les facteurs de confusion. Malgré l'application des méthodes d'ajustement, la confusion résiduelle n'est pas écartée dans ces études observationnelles. Il est donc difficile de tirer de ces études des résultats fiables et des conclusions robustes. Pour cela, il faut que ce biais de confusion inhérent à ces études soit estimé. Nous proposons ici une approche méta-épidémiologique pour tenter de quantifier ce biais de confusion résiduelle, par la différence de l'effet du traitement estimé, en termes d'efficacité et de sécurité, entre les études observationnelles et les ECR<sup>(96)</sup>.

#### 1.4 Objectifs de la thèse

L'évaluation de l'efficacité et de la sécurité d'un nouveau traitement repose sur les résultats des ECR réalisées idéalement en double aveugle. En effet, le schéma expérimental de ces essais assure le meilleur niveau de preuve scientifique en évitant, de par la randomisation, les biais de

sélection ou biais de confusion. Les résultats des ECR ont une forte validité interne et ces études sont donc nécessaires pour obtenir l'AMM d'un nouveau traitement. Afin de vérifier les conditions de prescription et de confirmer l'efficacité du traitement en vie réelle, les autorités de santé suggèrent la réalisation des études observationnelles post-AMM. Néanmoins, ces études observationnelles sont sujettes au biais de sélection/confusion dû à l'absence de randomisation. En effet, même si des méthodes d'ajustement statistique sont appliquées pour contrôler ces biais, elles ne permettent pas de garantir l'absence du biais de confusion résiduelle.

L'objectif de cette thèse est d'illustrer cette problématique en prenant comme exemple les études évaluant les AOD dans le traitement des ETEV et dans la FA. En effet, si les résultats obtenus par les ECR évaluant les AOD par rapport aux AVK sont tous homogènes et conduisent à la même conclusion, les résultats des études observationnelles sont parfois contradictoires et hétérogènes en termes de méthodologie et de résultat. Cette différence dans les résultats peut être expliquée par le biais de confusion résiduelle. D'où la nécessité de l'estimation de ce biais.

Afin d'atteindre l'objectif de cette thèse dans la quantification de ce biais de confusion inhérent dans les études observationnelles, une approche méta-épidémiologique a été utilisée. Cette approche méta-épidémiologique permet de quantifier la différence de l'effet du traitement estimé, en termes d'efficacité et de sécurité. Cette différence ne résumera pas seulement la confusion résiduelle, car il sera difficile de l'isoler de la non-transposabilité des ECR en pratique clinique, mais les caractéristiques des patients des deux types d'études pourraient nous permettre une approximation. Nous avons comparé d'une part la population incluse (ECR vs les études observationnelles) et d'autre part l'effet du traitement des AOD par rapport aux AVK (ECR vs les études observationnelles). Des méta-analyses stratifiées selon le type des études ont été réalisées afin de quantifier la différence d'estimation de l'effet traitement par le rapport des hazard ratio.

## 2 La méta-épidémiologie

Des méthodes méta-épidémiologiques et méta-analytiques ont été utilisées dans cette thèse afin de quantifier et d'estimer le biais des études observationnelles par rapport aux ECR. Dans cette partie, nous abordons la méthode de méta-analyse classique et ses principales étapes.

### 2.1 La méta-analyse

#### 2.1.1 Principes de la méta-analyse

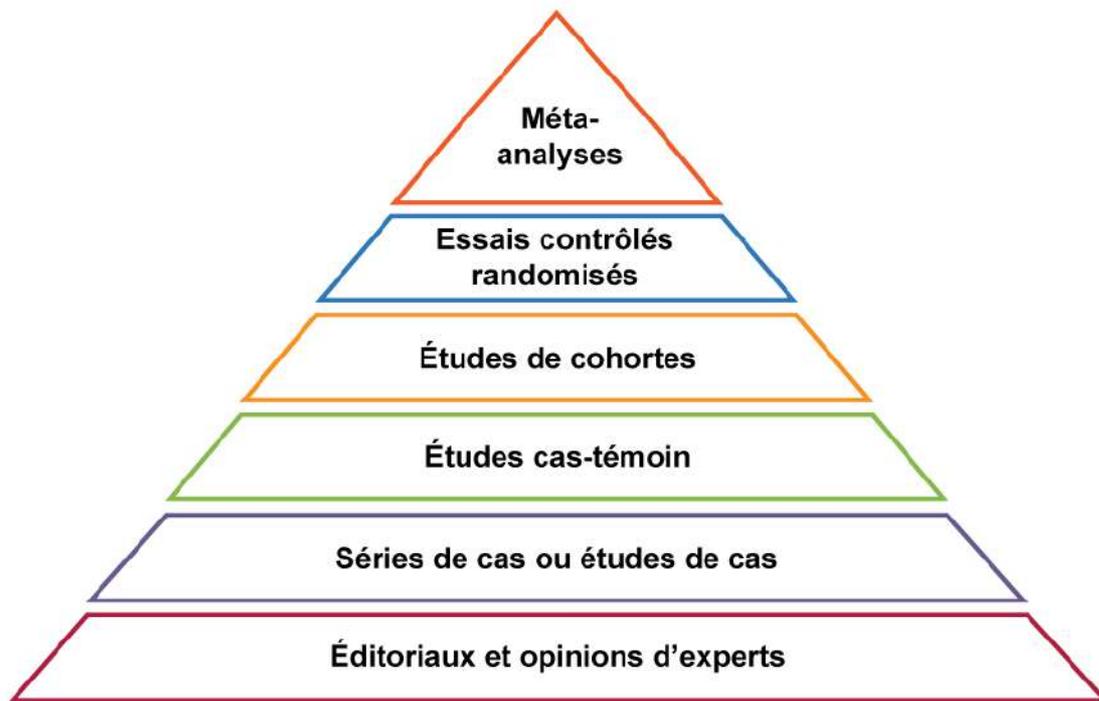
La méta-analyse en recherche clinique est une synthèse systématique et quantifiée des essais thérapeutiques suivant une méthodologie rigoureuse afin d'assurer sa reproductibilité et son impartialité<sup>(97)</sup>. Elle est donc basée sur 3 grands principes :

- **L'exhaustivité.** La méta-analyse s'appuie sur une recherche exhaustive de tous les essais publiés ou non publiés, favorables ou non au traitement étudié, significatifs ou non. Elle permet ainsi de combiner quantitativement les résultats de toutes les études abordant une question de recherche similaire.
- **La sélection argumentée des études.** Compte tenu de la grande quantité de revues scientifiques /articles publiés, la méta-analyse doit se baser sur une sélection argumentée des études. En effet, les auteurs doivent expliquer clairement sur quels critères les études ont été retenues ou non pour l'analyse.
- **La quantification.** La méta-analyse est une procédure statistique qui intègre les résultats de plusieurs études indépendantes. Elle donne une estimation de la taille de l'effet selon des calculs statistiques, en se basant sur l'hypothèse d'un effet commun : les études sont des échantillons fournissant des estimations de l'effet traitement, la combinaison de ces effets permet d'approcher la valeur commune. En théorie, le regroupement de données provenant de nombreux essais devrait améliorer la précision et la fiabilité de tout résultat mis en commun. La méta-analyse est donc un modèle d'étude quantitatif pour synthétiser les résultats des études d'une façon précise et fiable.

La méta-analyse est la référence en recherche clinique la plus fréquemment citée, ayant le niveau de preuve scientifique le plus haut (Figure 2). Durant ces 10 dernières années, on compte

plus de 8000 méta-analyses publiées par an. Le succès de la méta-analyse provient, entre autres, du fait qu'elle répond à un besoin ressenti par de nombreux acteurs de santé. Les médecins ont de plus en plus besoin de données synthétiques qui intègrent efficacement l'ensemble des informations existantes pour assurer une base rationnelle à leur décision médicale<sup>(98)</sup>.

Figure 2: Niveau de preuve de la médecine basée sur les faits



Source : EUPATI - Académie européenne des patients

### 2.1.2 Intérêts de la méta-analyse

Par rapport à une analyse d'un seul essai ou de plusieurs pris individuellement, la méta-analyse a de multiples avantages :

- Elle **permet d'augmenter la probabilité de trouver un effet bénéfique (ou néfaste) significatif** si l'effet existe, et **d'augmenter la précision de l'estimation de la taille de l'effet du traitement**. En effet, elle se base sur un plus grand nombre de sujets, ce qui réduit la taille de l'intervalle de confiance de l'effet du traitement. **L'avantage d'une méta-analyse est l'augmentation de la puissance statistique** grâce à l'augmentation du nombre de patients inclus. Elle permet donc de détecter une différence d'efficacité non mise en évidence dans chaque étude prise individuellement. Plus l'efficacité du traitement étudié est faible comparée à celle du traitement de

référence, plus le nombre de patients nécessaires pour mettre en évidence cette différence doit être important.

- Elle peut montrer **l'effet du traitement sur un critère clinique non mis en évidence** par manque de puissance sur les études prises isolément. Elle **peut donc répondre à une question non posée initialement** par les études et dans les cas où les essais déjà réalisés pris individuellement ne donnent pas de résultat statistiquement significatif.
- Dans certains essais, la méta-analyse permet **de réconcilier** des résultats apparemment discordants et **de lever ainsi le doute, et d'expliquer la variabilité des résultats** entre les études, variabilité parfois liée à la méthodologie utilisée.
- La méta-analyse procure un moyen **d'approcher le "patient moyen tout venant" de la population de diffusion**. En effet, chaque essai pris individuellement sélectionne avec beaucoup de soin les sujets inclus. En regroupant ces essais portant sur des groupes de sujets de caractéristiques différentes, la méta-analyse permet de tester et d'augmenter la généralisation d'un résultat à un large éventail de patients.

### 2.1.3 Déroulement d'une méta-analyse

#### *A/ Poser une question de recherche et définir les critères de sélection*

La méta-analyse cherche à répondre à une question de recherche claire et précise. Cette question de recherche est formulée à partir des études déjà réalisées, des questions résolues ou partiellement résolues et des problèmes restés en suspens. La question posée conditionnera d'une façon importante la sélection des études à inclure dans la méta-analyse. Les critères de sélection découlent directement de la question de recherche. Ils permettent de ne retenir que les études qui apportent des informations nécessaires pour répondre à la question posée par la méta-analyse. Ils se définissent par la maladie, la population étudiée (caractéristiques des patients), les traitements (doses, etc.) ou la stratégie thérapeutique utilisée et le type des critères de jugement utilisés.

#### *B/ Recherche des données : revue de la littérature*

- Ne rien manquer : exhaustivité

Après avoir formulé une question de recherche précise et défini les critères de sélection, la deuxième étape de la méta-analyse consiste à faire une revue systématique des études

existantes. Il convient donc de faire une recherche bibliographique exhaustive et de prendre en considération toutes les études éligibles, qu'elles soient publiées ou non. Il est notamment important de ne pas limiter la recherche aux publications sous forme d'articles afin d'éviter le biais de publication. C'est pourquoi la revue systématique doit être réalisée non seulement dans les bases de publications scientifiques (comme PubMed ou Embase) mais aussi dans les registres des essais randomisés (comme les registres de la collaboration Cochrane, clinicaltrials.gov), les actes de congrès (recherche d'essais publiés sous forme de résumé) et les sites internet des agences réglementaires pour les médicaments récents. Et enfin, ces équations de recherche doivent être clairement définies afin qu'elles soient reproductibles et facilement vérifiables.

- Ne pas compter plusieurs fois les études : gérer les doublons

La recherche systématique amène un grand nombre d'études éligibles à l'inclusion dans la méta-analyse. Parmi ces études éligibles, certaines études notamment épidémiologiques sont conduites sur les mêmes bases de données de santé informatisées, sur des périodes d'inclusion identiques. Il s'agit donc de doublons en termes méta-analytiques. Une gestion des doublons est nécessaire afin de ne pas introduire dans la méta-analyse deux fois (voire plus) la même information, ce qui conduirait à un biais de surreprésentation de certaines données. Le choix de la publication à retenir pour la méta-analyse dépendra des critères d'exclusion des doublons. Généralement, l'étude ayant la période de recrutement la plus grande et/ou l'effectif le plus important est pris en considération dans la méta-analyse.

### *C/ Analyse des études trouvées*

La qualité de l'estimation combinée de l'effet étudié va dépendre non seulement de l'identification de l'ensemble des études posant la même question de recherche, mais également de la qualité méthodologique de ces études (présence ou non de biais dans chaque essai). Une fois que les études éligibles sont identifiées par une recherche exhaustive, une analyse de la qualité méthodologique de ces études doit être effectuée. Certaines études réalisées ne respectent pas les contraintes méthodologiques, ce qui réduit la fiabilité de leur résultat. Par exemple, si l'allocation des traitements n'est pas aléatoire ou si les études non randomisées n'utilisent pas des méthodes d'ajustements pour minimiser les biais, les résultats de la méta-analyse seront biaisés. C'est pourquoi il est important d'évaluer la qualité méthodologique des études afin de distinguer les études les plus fiables de celles à haut risque de biais.

#### 2.1.4 Analyse des effets

##### *A/ Estimation ou mesure de l'effet traitement*

La problématique statistique posée par la méta-analyse est double. Il s'agit tout d'abord d'extraire l'estimation de l'effet traitement au sein de chaque étude, puis de combiner en un seul indice ces mesures d'effets issues d'une série d'études<sup>(99,100)</sup>.

Les mesures d'effet varient en fonction de la nature du critère de jugement utilisé (qualitatif ou quantitatif)<sup>(101)</sup>:

- Pour un critère binaire, ces mesures d'effet, comme le risque relatif (RR), le rapport des cotes (RC) ou la différence des risques (DR), se déduisent des fréquences de survenue de l'événement avec et sans traitement. Avec le modèle multiplicatif, deux choix de mesure sont possibles: RR et RC. Ces deux mesures seront  $< 1$  quand le traitement est bénéfique,  $> 1$  si le traitement est délétère et égal à 1 en cas d'absence d'effet du traitement. Concernant la DR, la mesure additive devient: négative en cas de traitement bénéfique, positive en cas de traitement délétère et nulle en cas d'absence d'effet du traitement.
- Pour un critère de jugement continu, la mesure utilisée repose sur l'effet standardisé. L'effet standardisé est la différence des moyennes standardisées sur l'écart type commun aux deux groupes (groupe contrôle et groupe expérimental).

##### *B/ Calcul de l'effet commun*

Les études incluses dans une méta-analyse sont effectuées dans des conditions différentes, et sont issues de groupes de populations différentes. Ceci pose d'emblée la signification d'un effet commun moyen, discutée à travers la notion d'homogénéité/hétérogénéité entre études. Elle amène très schématiquement au choix entre deux méthodes de combinaison des tailles d'effet : la première dite méthode à effets fixes, et la seconde dite méthode à effets aléatoires.

- La méthode à effets fixes fait l'hypothèse d'une homogénéité des résultats entre les études. En effet, la méta-analyse est caractérisée par une hypothèse fondamentale qui est nécessaire pour donner un sens au principe de la méta-analyse : l'hypothèse d'homogénéité. Autrement dit, pour estimer l'efficacité d'un traitement en regroupant plusieurs études, la quantité d'effet du traitement est considéré constante. Ainsi, chaque essai thérapeutique mesure cette même constante ( $\theta$ ). L'effet traitement est le même pour chacune des études. Les seules fluctuations de résultats observées entre les études proviennent de fluctuations aléatoires

autour de cet effet commun moyen  $\theta$  qui correspond à la variabilité intra-étude ( $\sigma_i^2$ ). L'estimation de l'effet traitement commun  $\theta$  est égale à la moyenne des estimations réalisées pour chaque étude, pondérées par un poids  $w_i$  inverse de leur variance estimée ( $w_i = 1/\sigma_i^2$ )<sup>(102)</sup>.

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

L'effet commun est estimé assorti de son intervalle de confiance.

- La méthode à effets aléatoires est utilisée quand les résultats des études sont très hétérogènes sans raison apparente permettant de comprendre cette hétérogénéité. La mesure d'effet global calculée par la méthode à effets fixes sera moins valide et plus difficile à interpréter. La méthode à effets aléatoires propose une autre approche, qui prend en compte une variabilité des différences d'effets observées entre les études (les  $\theta_j$  sont différents d'une étude à l'autre). Ce modèle à effets aléatoires suppose qu'à l'effet commun précédent, se rajoute un effet aléatoire propre à chaque étude. Il intègre une variabilité inter-études, en plus de la variabilité intra-étude. Cette variabilité inter études correspond à l'hétérogénéité.

L'estimateur de l'effet commun dans ce modèle aléatoire est également une moyenne pondérée des effets observés dans chaque étude comme dans le modèle à effets fixes. Mais les  $w_i$  dans ce modèle sont remplacés par des  $w_i^*$ .

L'expression de  $w_i^*$  intègre un terme  $\tau^2$ . Alors  $w_i^* = 1/(\sigma_i^2 + \tau^2)$  où  $\tau^2$  représente la variabilité inter-étude et  $\sigma_i^2$  la variabilité intra-étude qui vient s'ajouter à la variabilité intra-études.

L'estimation de l'effet de traitement  $\theta$  dans ce modèle aléatoire est égale à :  $\hat{\theta}^* = \frac{\sum \hat{\theta}_i w_i^*}{\sum w_i^*}$ .

### *C/ Choix de la méthode statistique de combinaison*

En méta-analyse, plusieurs méthodes statistiques sont disponibles pour une combinaison donnée des critères binaires pour le modèle à effets fixes. Des méthodes classiques ont été proposées dans ce modèle à effets fixes afin d'estimer le RC commun (Peto, le logarithme du rapport des cotes et Mantel-Haensze), le RR commun (méthode du logarithme du risque relatif) et le DR commun (Méthode de la différence de risque). D'autre part, pour le modèle à effets aléatoires, la méthode de DerSimonian & Laird peut s'appliquer à tous types de paramètres (RR, OR, DR).

## D/ Identification des sources d'hétérogénéité et analyse en sous-groupes

L'hétérogénéité entre les études et l'exploration de ses sources est une question très importante lorsque nous interprétons les résultats d'une méta-analyse. L'hétérogénéité est définie par le rejet de l'hypothèse d'homogénéité des effets du traitement. Le test statistique utilisé pour évaluer cette hétérogénéité est le test Q de Cochran. Ce test est calculé en additionnant les écarts au carré de l'estimation de chaque étude par rapport à l'estimation globale de la méta-analyse, en pondérant la contribution de chaque étude de la même manière que dans la méta-analyse.

$$Q = \sum_i w_i (\theta_i - \theta_{Commun})^2$$

où  $\theta_i$  : effet estimé de l'étude i et son poids  $w_i = \frac{1}{var(\theta_i)}$ , et  $\theta_{Commun}$  : l'estimation globale de la méta-analyse.

Cependant, ce test de Cochran a plusieurs limites. Premièrement il ne permet pas de détecter une véritable hétérogénéité entre les études comme étant significative. Deuxièmement, la puissance du test d'hétérogénéité n'est satisfaisante que lorsqu'il existe un nombre important d'études. Face à ces limites, une alternative du test de Cochran a été mise en place : l'indicateur  $I^2$ . Cet indicateur donne une meilleure mesure de la cohérence entre les essais d'une méta-analyse. Il quantifie l'effet de l'hétérogénéité en fournissant une mesure du degré d'incohérence des résultats des études.  $I^2$  peut être facilement calculé à partir des résultats de base obtenus par une méta-analyse typique<sup>(103)</sup>.

$$I^2 = 100 \% \times (Q - df) / Q$$

où Q = statistique d'hétérogénéité de Cochran et df = k-1 les degrés de liberté (où k est le nombre d'études).

L'interprétation des valeurs  $I^2$  est la suivante :

- Une valeur de  $I^2$  de 0% représente une hétérogénéité non observée,
- Des valeurs de  $I^2$  plus élevées indiquent une hétérogénéité croissante.
- Des valeurs de  $I^2$  égales à 25 %, 50 % et 75 % indiquent ainsi une hétérogénéité faible, modéré et élevé<sup>(104)</sup>.
- Une valeur de  $I^2$  proche de 0 signifie que la taille de l'effet est comparable d'une étude à l'autre.

Devant une hétérogénéité, il convient de rechercher la ou les études qui génèrent l'hétérogénéité. Ensuite, il convient d'exclure ces études lors d'une analyse de sensibilité, tout en recherchant la composante de ces études induisant l'hétérogénéité. Il est également intéressant de rechercher systématiquement les interactions entre les résultats observés et une ou plusieurs covariables par une analyse en sous-groupes. Les analyses en sous-groupes sont fréquemment réalisées en complément de l'analyse principale d'une étude, afin de rechercher une interaction entre l'effet du traitement et le facteur définissant ses sous-groupes. Une analyse en sous-groupes consiste à subdiviser la population des études en deux ou plusieurs sous-groupes d'études et à étudier l'efficacité du traitement dans chacun de ces sous-groupes. Les sous-groupes peuvent être définis par le type de traitement, par la méthodologie choisie pour les études, le type de patients inclus... Ainsi, les études sont regroupées par exemple en fonction de la molécule, des classes pharmacologiques ou de la nature de l'intervention (régime, médicaments, chirurgie). Les sous-groupes peuvent aussi être définis par les caractéristiques propres aux patients (l'âge, le sexe, le risque de base ou du pronostic), ou par le type de mesure du critère de jugement, les conditions de réalisation des études et le type d'études (études en ouvert ou double aveugle).

Enfin, si l'hétérogénéité observée ne s'explique pas par une interaction, il est possible de recourir à un modèle à effets aléatoires.

#### 2.1.5 Illustrer les résultats

##### A/ Représentation graphique de l'effet traitement

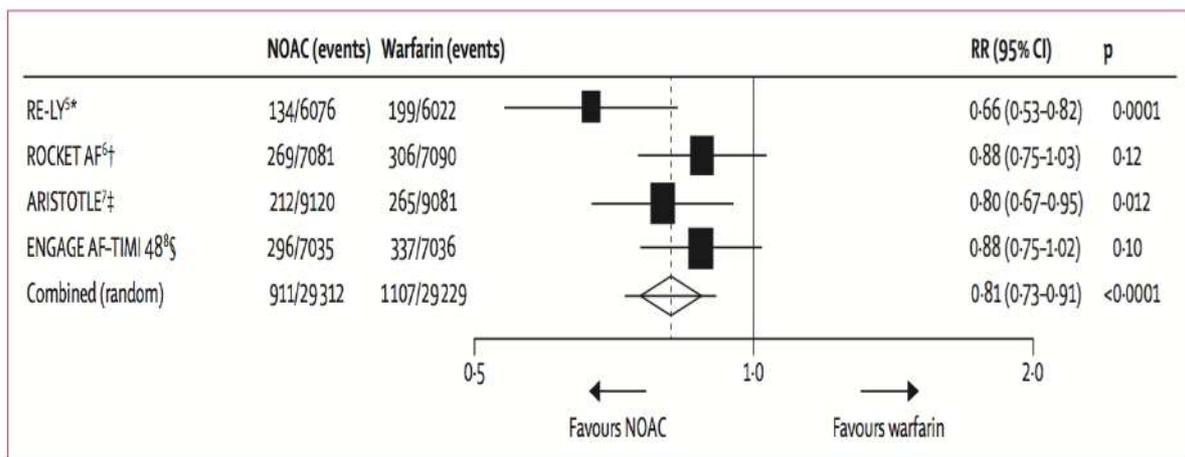
Les résultats d'une méta-analyse sont représentés graphiquement sous la forme d'un graphique en forêt (« forest plot »). Ce graphique montre aussi bien les résultats individuels de chaque étude que le résultat global (l'effet traitement commun) estimé par la méta-analyse.

Les effets traitement de chaque étude sont représentés sous la forme de points entouré par leur intervalle de confiance à 95% (le trait horizontal). L'aire de chaque point est souvent proportionnelle au poids de l'étude dans la méta-analyse. Le résultat global figure au bas du graphique, sous la forme d'un losange, qui représente l'évaluation de l'effet global et l'intervalle de confiance. L'échelle du graphique dépend de la mesure de l'effet traitement retenu : RC, RR, DR.

Sur ce graphique, un trait vertical continu indique l'absence d'effet traitement. Avec le RR ou le RC ce trait a pour abscisse 1. Pour une représentation utilisant la DR ou l'effet standardisé, l'absence d'effet traitement est traduit par la valeur zéro.

La figure 3 ci- dessous représente un exemple de graphique « Forest plot » d'une méta-analyse.

Figure 3: Exemple d'un forest plot



Source : Ruff et al. Lancet. 2014;383(9921):955-962.

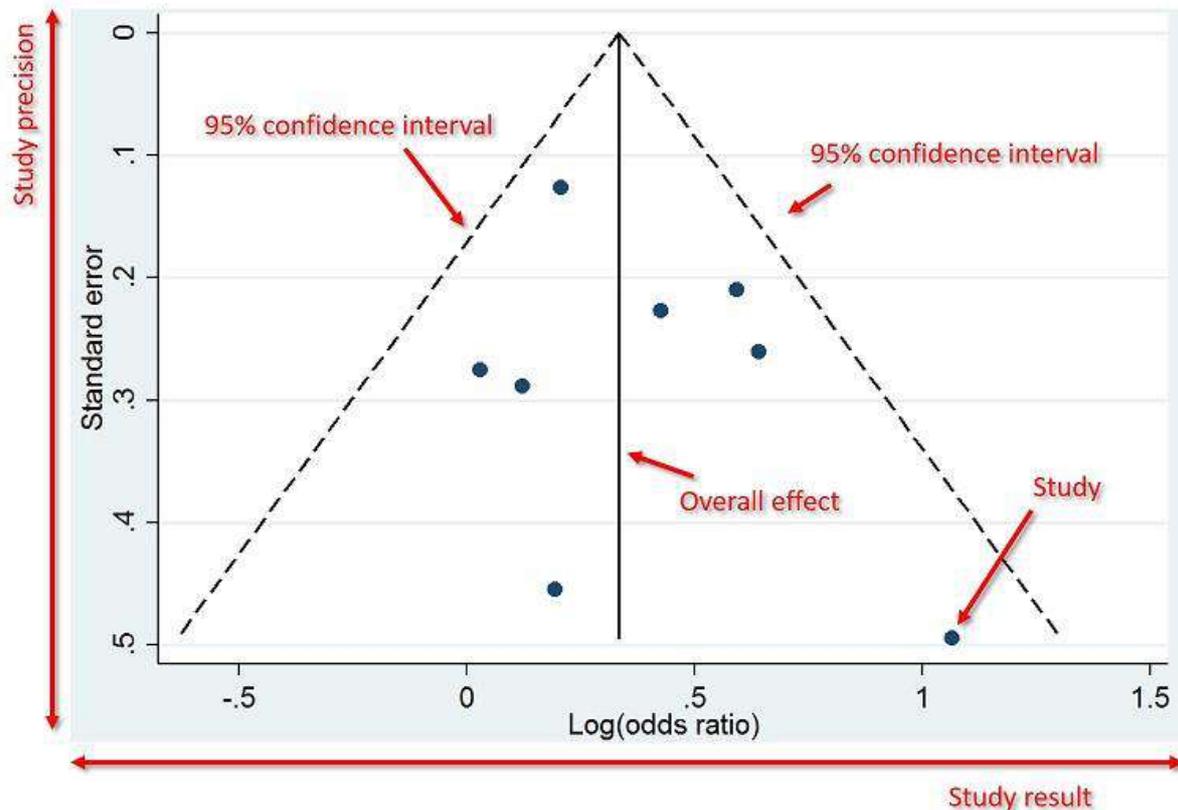
### B/ Représentation graphique pour l'identification de biais ou d'interactions

La méta-analyse doit regrouper toutes les études qui ont été réalisées dans le domaine, sans sélection arbitraire ou involontaire<sup>(105)</sup>. Cependant, l'exhaustivité nécessaire pour la fiabilité des résultats de la méta-analyse est parfois difficile à atteindre, en particulier du fait de l'existence de travaux non publiés. En effet, les études dont les résultats sont statistiquement significatifs (résultats positifs) ont plus de chances d'être publiées par une revue scientifique que les études ne montrant pas de différence significative (résultats négatifs)<sup>(106)</sup>. Il existe ainsi une sélection involontaire des publications ayant des résultats statistiquement significatifs. Ce phénomène est appelé biais de publication. Ce biais va fausser les conclusions que l'on peut obtenir à partir des simples résultats publiés<sup>(107,108)</sup>. Dans une méta-analyse, si aucune recherche poussée des essais non publiés n'est effectuée, le risque encouru est une surestimation de l'efficacité du traitement<sup>(109,110)</sup>.

Afin de détecter un éventuel biais de publication dans une méta-analyse, la méthode la plus simple et la plus utilisée est le graphique en entonnoir (« funnel plot »). Il représente les estimations ponctuelles en fonction de la précision de l'estimation (souvent liée à la taille de l'échantillon pour chaque étude). Un graphique en entonnoir est un nuage de points qui consiste

à représenter pour chaque étude la valeur estimée de l'effet traitement (axe x) par rapport à la taille de son échantillon (axe y). Un exemple est présenté Figure 4.

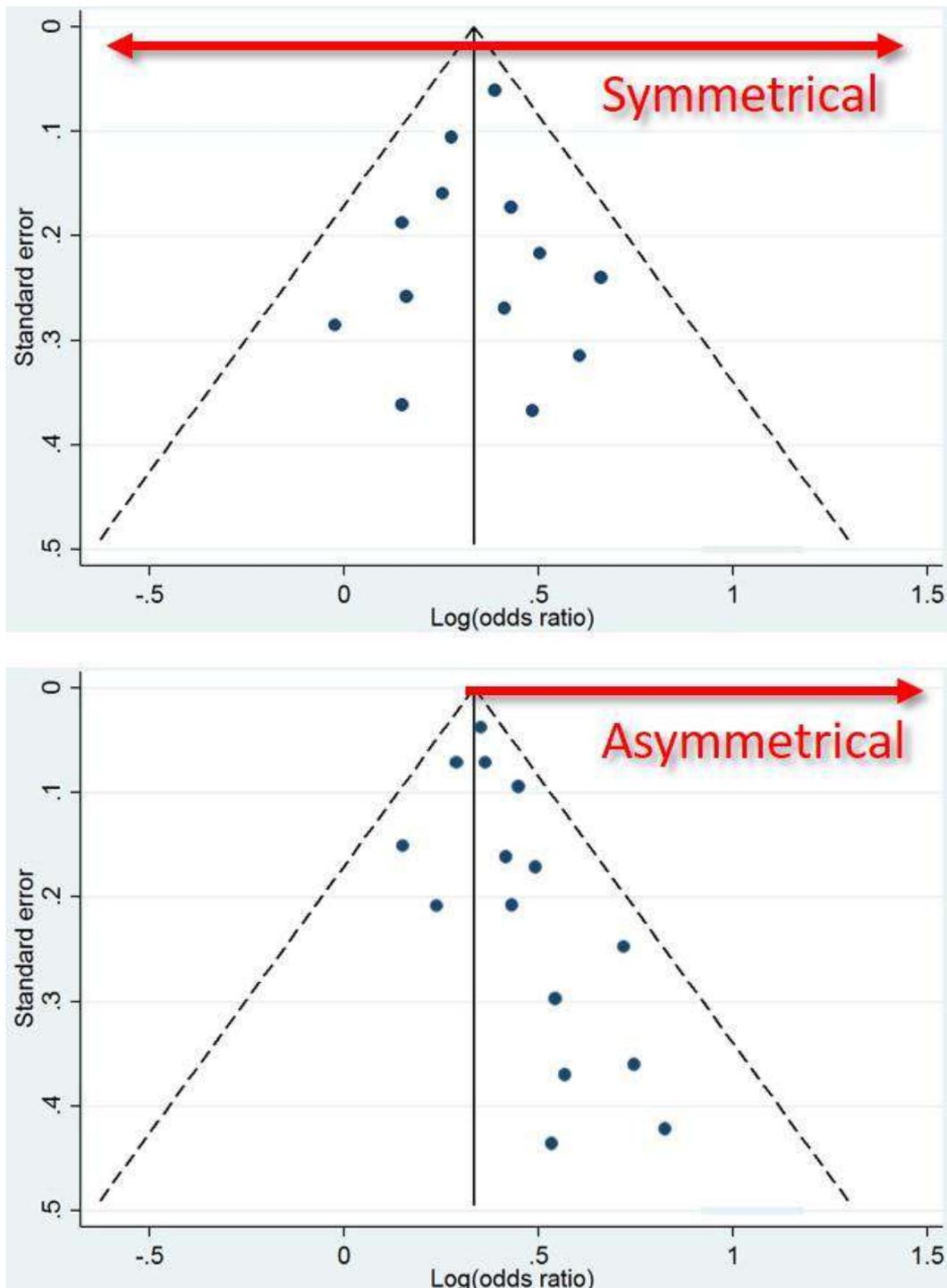
Figure 4: Illustration d'un funnel plot. Une symétrie est attendue puisque chaque étude doit se distribuer de façon aléatoire autour de l'effet commun



En l'absence de biais de publication, les différentes estimations de l'effet du traitement vont se répartir autour de la valeur commune (la moyenne) et donnent un nuage de points évasé. Dans le cas d'un biais de publication, la répartition n'est plus homogène. Un déséquilibre apparaît avec disparition des points situés dans la zone correspondant aux résultats non significatifs et le graphique en entonnoir est faussé : il devient creux. Ce graphique est généralement considéré comme un très bon moyen visuel pour identifier un biais de publication dans une méta-analyse qui se traduit par un entonnoir asymétrique<sup>(111,112)</sup>.

A titre d'exemple pédagogique, la figure 5 présente les deux cas de figure.

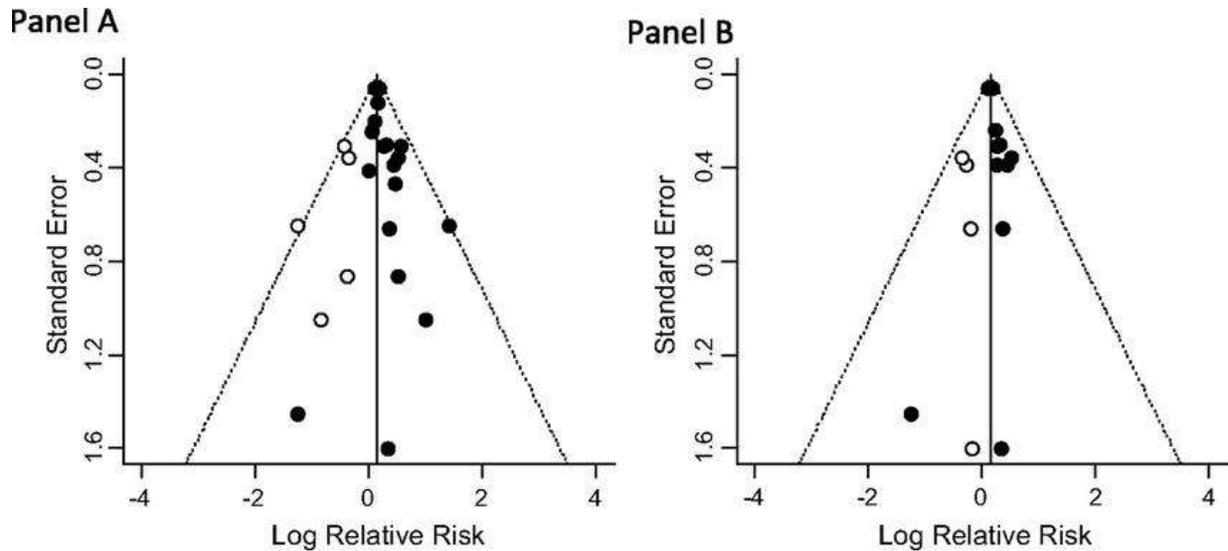
Figure 5: Illustration d'un funnel plot symétrique et asymétrique, faisant suspecter la présence d'un possible biais de publication dans le graphique du bas



Les graphiques en entonnoir peuvent aussi être utilisés pour détecter des interactions avec le type d'étude, voir des biais. Si on représente par exemple les études en double aveugle par des

ronds noirs et les études en ouvert par des ronds blancs sur ce type de graphique, il est aisé de voir si le résultat global est influencé par certains types d'étude, comme c'est le cas sur la Figure 6.

Figure 6: Exemple d'identification d'une interaction avec les caractéristiques des études.

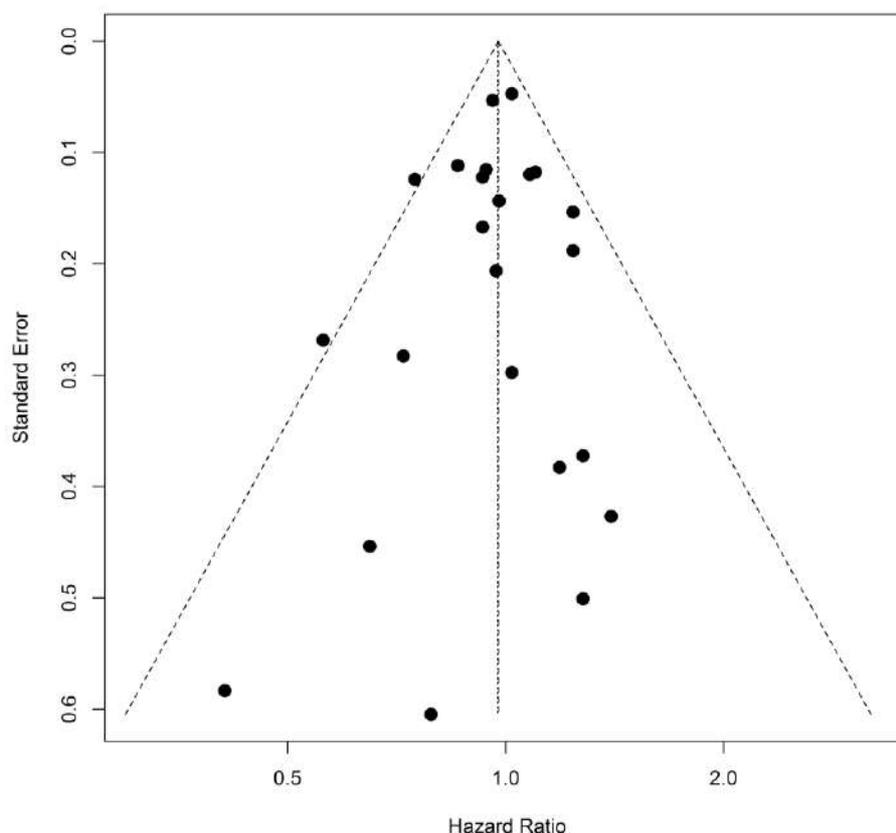


Le résultat est plus favorable lorsque les études présentent la caractéristique « rond blanc » (par exemple étude en ouvert) par rapport aux études « rond noir » (études en double aveugle).

Cet outil peut donc également être utilisé en méta-épidémiologie, comme nous allons le voir dans la partie suivante.

La figure 7 du funnel plot ci-dessous, issue d'un article<sup>(113)</sup> que nous avons publié récemment, montre un bon exemple de mise en œuvre d'une bonne symétrie avec une absence de biais. Dans cet article, mon principal rôle était la sélection des études, la recherche des doublons, et du biais de publication.

Figure 7: Funnel plot symétrique illustrant le biais de publication



En résumé, les méta-analyses sont des outils indispensables pour synthétiser les résultats de plusieurs études disponibles ayant la même question de recherche. Mais elles peuvent également être utilisées afin de développer la méthodologie de recherche en épidémiologie : on parle alors de « la méta-épidémiologie ».

## 2.2 De la méta-analyse à la méta-épidémiologie

### 2.2.1 Principes de la méta-épidémiologie

#### *A/ Objectif de la méta-épidémiologie*

Dans les ECR, il existe un ensemble de processus méthodologiques à respecter, faute de quoi l'essai est exposé au risque de fournir une estimation biaisée de l'effet traitement. Ces processus méthodologiques sont: la génération adéquate de la séquence de randomisation (biais de sélection), le maintien de l'assignation secrète (biais de sélection), l'aveugle des participants et du personnel de l'essai (biais de performance), l'aveugle des évaluateurs des critères de

jugement (biais de détection) et le suivi complet des patients (biais d'attrition). Ces cinq processus sont inclus dans l'outil de Cochrane Rob2<sup>(11)</sup>. Le choix de ces cinq processus a été basé sur des considérations théoriques mais également sur des preuves empiriques fournies par des études méta-épidémiologiques. Des erreurs dans la conception ou dans la réalisation d'un ECR peuvent affecter l'un ou l'autre des processus et donc la validité de ses résultats.

Des études dites de méta-épidémiologie ont permis d'étudier l'influence de caractéristiques associées à la validité interne ou à la validité externe. Elles sont basées sur la combinaison de deux concepts : l'épidémiologie et la méta-analyse. Ces études permettent de comparer l'effet traitement entre les ECR, avec et sans la caractéristique d'intérêt, dans des échantillons de méta-analyses. De façon succincte, la méthode consiste à identifier une collection de méta-analyses. Quels que soient la maladie et les traitements évalués, chacune de ces méta-analyses doit inclure au moins un essai avec la caractéristique d'intérêt (double aveugle par exemple) et un autre essai sans la caractéristique d'intérêt (en ouvert). Pour chaque méta-analyse sélectionnée, on estime, via un modèle à effets fixes ou un modèle à effets aléatoires, l'effet traitement commun aux essais avec la caractéristique et l'effet traitement commun aux essais sans la caractéristique. Cette méthode permet alors d'évaluer l'association entre une ou plusieurs caractéristiques des essais cliniques (par exemple l'absence du double aveugle) et l'estimation de l'effet traitement à partir d'une série de méta-analyses<sup>(114)</sup>.

Initialement, la méta-épidémiologie était définie comme une méthode statistique pour évaluer l'influence des problèmes méthodologiques dans les ECR<sup>(114)</sup>. Par la suite, le principe de la méta-épidémiologie est passé d'une méthode statistique à une nouvelle méthodologie visant à combler les écarts entre les preuves et la pratique. Les méta-analyses sont en fait aussi exploitées à titre pédagogique pour convaincre les cliniciens sur la nécessité de conduire des études cliniques avec de bons principes méthodologiques et alerter sur la lecture de résultats potentiellement biaisés. Le but de l'approche méta-épidémiologique est donc d'améliorer la méthodologie des essais cliniques<sup>(114,115)</sup>.

### *B/ Modèles statistiques sous-jacents*

Les collections de méta-analyses publiées, comme les revues systématiques de la bibliothèque de Cochrane<sup>(116)</sup>, constituent une ressource essentielle pour examiner l'impact des caractéristiques des essais sur l'effet du traitement par une analyse méta-épidémiologique des données collectées. Étudier l'impact des caractéristiques consiste à évaluer les biais associés à ces caractéristiques. Afin d'évaluer ces biais, deux modèles statistiques ont été développés<sup>(114)</sup>.

Pour un échantillon de R revues systématiques comprenant un total de S essais, ces deux

modèles statistiques sont présentés ci-dessous :

- Un modèle de régression logistique standard qui se base sur le nombre des événements dans les deux groupes pour évaluer le biais<sup>(114)</sup>. Ce modèle était l'outil standard de la plupart des recherches méta-épidémiologiques<sup>(117,118)</sup>. Une version multivariée de ce modèle a été développée: elle inclut les caractéristiques de l'essai (P) dans un modèle linéaire. Avec ce modèle, la probabilité  $\pi$  d'un résultat défavorable pour un individu est décrite par l'équation suivante<sup>(119)</sup>:

$$\text{logit}(\pi) = \sum_{k=1}^P \beta_k c_k I_t + \sum_{r=1}^R \gamma_r I_t I_r + \sum_{s=1}^S \tau_s I_s \quad (1)$$

où  $I_t, I_r, I_s$  sont des variables indicatrices pour qu'un individu soit dans le groupe expérimental, dans la revue  $r$  et dans l'essai  $s$  respectivement.  $c_k = c_k(s)$  est un indicateur d'une caractéristique inappropriée pour un  $k$  ième essai.

- Un modèle de régression linéaire multiple pondérée qui évalue le biais associé aux composantes de qualité pour les mesures de synthèses des essais. Ce modèle de régression pondérée qui intègre la variation entre les essais donne cependant des intervalles de confiance plus larges<sup>(119)</sup>. Chaque essai donne une seule observation, la taille de l'effet observé est exprimée en  $\log(\text{OR})$  avec un poids égal à l'inverse de son erreur type. Par conséquent, les résultats de l'essai sont modélisés par l'équation suivante :

$$\log(\text{OR})_s = \sum_{k=1}^P \beta_k c_k + \sum_{r=1}^R \gamma_r I_r + \varepsilon_s \quad (2)$$

Avec  $\varepsilon_s \sim N(0, \psi v_s)$ , où  $v_s$  est la variance de loi Bernoulli des  $\log(\text{OR})_s$  correspondants, et  $\psi$  est un paramètre de sur-dispersion dont la variance du résidu  $\varepsilon_s$  est supposée proportionnelle à la variance de l'estimation de l'essai  $\log(\text{OR})_s$ .

Dans les modèles de régression logistique, les mesures de résultats non binaires et l'utilisation des techniques pour traiter les données manquantes ou censurées peuvent être impossibles. De plus, de nombreuses données cliniques provenant d'autres sources peuvent être manquantes dans ce modèle. Dans les modèles de régression pondérée, il est possible d'inclure des données provenant d'un plus grand nombre de sources. Au total, parmi ces deux modèles statistiques, le modèle de régression pondérée est préféré dans certains cas car il est plus pratique, flexible, plus facilement généralisé et incluant un plus grand nombre d'essais.

Dans ces deux modèles de base, le biais est supposé homogène d'une revue à l'autre<sup>(120)</sup>. Dans les deux modèles, les coefficients de biais  $\beta_k$  sont les mesures de l'influence des biais associés aux caractéristiques de l'essai sur les effets estimés du traitement. Le paramètre  $\gamma_r$  représente le véritable effet du traitement que la revue  $r$  vise à estimer. Dans le modèle de régression

logistique,  $\tau_s$  est le logarithme du rapport de cotes des résultats de base de l'essai s.

Le modèle de régression pondérée permet d'estimer l'effet de traitement dans l'essai. Lorsque cet effet est exprimé en échelle logarithmique (log OR), il correspond à la différence en logit ( $\pi$ ) :

$$E \log(\text{OR}) = \text{logit}(\pi)_{I_{t=1}} - \text{logit}(\pi)_{I_{t=0}} \text{ pour chaque essai}$$

Les deux modèles estiment les mêmes paramètres et le même  $\beta_k$ . Cependant, ils ne sont pas équivalents puisqu'ils étudient les sources de variation d'une façon différente<sup>(119)</sup>.

Le coefficient  $\beta_k$  correspond à la  $k$ ème caractéristique binaire  $c_k$ . Pour chaque essai, l'effet estimé du traitement est calculé en termes de logarithme du rapport de cotes (OR) avec l'erreur-type conventionnelle correspondante à l'OR. Les effets positifs du traitement sont représentés par des faibles valeurs de log(OR). Le coefficient  $\beta_k$  correspond à un changement dans log(OR) lorsqu'on passe de la classe de référence (adéquate) à une autre classe alternative (inadéquate)<sup>(119)</sup>. Le biais associé à une caractéristique de l'essai est exprimé comme le rapport des rapports de cotes (ROR) entre les essais dont les caractéristiques sont inadéquates et ceux dont les caractéristiques sont adéquates. Pour la caractéristique binaire  $k$  :

$$ROR_k = \frac{OR_{inadéquate}}{OR_{adéquate}} = e^{\beta_k}$$

Un  $OR < 1$  indique la supériorité du traitement expérimental par rapport au traitement témoin.

Un  $ROR < 1$  indique que les essais avec la caractéristique inadéquate donnent des estimations de l'effet traitement plus grandes que celles données par les essais avec la caractéristique adéquate<sup>(118, 119)</sup>.

#### *C/ Méthode de comparaison de l'effet traitement via le ratio des odds ratio (ROR)*

Afin de comparer deux quantités estimées telles que les moyennes ou les proportions (a accompagnées de leur dispersion), une méthode de comparaison a été développée en 2003 par Altman et al<sup>(121)</sup>. Cette méthode nécessite que les deux estimations soient indépendantes, issues d'analyses distinctes donc non obtenues auprès des mêmes patients. Pour comparer deux estimations  $E_1$  et  $E_2$  avec les erreurs types  $SE(E_1)$  et  $SE(E_2)$  respectivement, cette méthode consiste à calculer la différence ( $d = E_1 - E_2$ ) entre ces deux estimations avec l'erreur type de la différence. L'erreur type de la différence est égale à la racine carrée de la somme des carrés des différentes erreurs types ( $\sqrt{[SE(E_1)]^2 + [SE(E_2)]^2}$ ). Le test d'interaction  $z = d/SE(d)$  permet de tester l'hypothèse nulle selon laquelle, dans la population, la différence  $d$  est égale à zéro, en

comparant la valeur de  $z$  à la distribution normale standard. L'intervalle de confiance à 95% de la différence est compris entre  $d-1,96SE(d)$  et  $d+1,96SE(d)$ <sup>(121)</sup>.

Cette méthode est ainsi utilisée afin de comparer l'effet traitement entre une méta-analyse avec la caractéristique d'intérêt et une autre sans cette caractéristique. Pour un critère de jugement binaire, cette différence d'effet traitement entre les deux méta-analyses est mesurée par le rapport des risques relatifs ou le rapport des rapports de cotes (en anglais, ratio of odds ratios, ROR). Les mesures de l'effet traitement (RR, OR) sont toujours analysées en échelle logarithmique puisque les distributions des log ratios ont tendance à être plus proches de la distribution normale que les ratios eux-mêmes, ce qui facilite les comparaisons<sup>(121)</sup>. Le tableau 2 ci-dessous explique les étapes à suivre pour calculer le ROR :

- 1/ Calcul des logarithmes des mesures de l'effet (OR) et de leurs intervalles de confiance à 95% (ligne 1).
- 2/ Calcul de l'erreur type(SE) de chaque logarithme de risque relatif : diviser la largeur de son intervalle de confiance par  $2 \times 1,96$  (ligne3)
- 3/ Mesurer la différence estimée entre les logarithmes des rapports de risques ( $d$ ) (ligne 4) et calculer ainsi son erreur type (ligne5).
- 4/ Tester l'interaction ( $z$ ) et chercher le ( $p$ ) correspondant dans le tableau de la distribution de loi normale (ligne 7).
- 5/ Estimer le ratio des odds ratio qui est égal à l'exponentielle de la différence( $d$ ) et calculer aussi son intervalle de confiance (ligne 8,9). Cette valeur (ROR) peut également être calculée directement en divisant  $E_1$  par  $E_2$ .

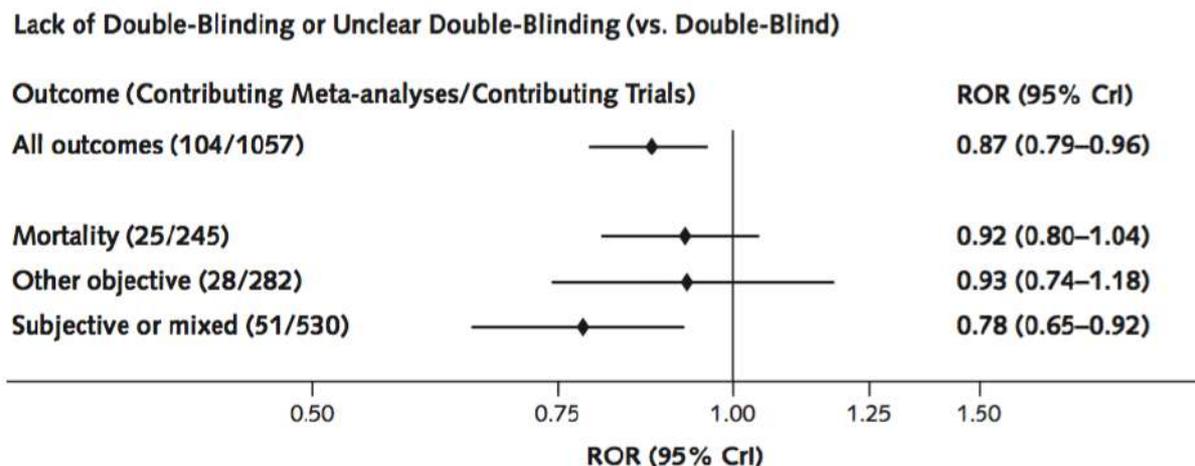
**Tableau 2:** Tableau de calcul permettant de comparer deux rapports de risque

		<b>Meta-analyse A</b>	<b>Meta-analyse B</b>
<b>1</b>	log(OR)	$E_1 = \log(OR_1)$	$E_2 = \log(OR_2)$
<b>2</b>	Largeur IC log(OR)	$\log[b_s(OR_1)] - \log[b_i(OR_1)]$	$\log[b_s(OR_2)] - \log[b_i(OR_2)]$
<b>3</b>	S(E)	$S(E_1) = \frac{\text{largeur de } IC_1}{2 \times 1,96}$	$S(E_2) = \frac{\text{largeur de } IC_2}{2 \times 1,96}$
<b>Calcul du ROR</b>			
<b>4</b>	d	$E_1 - E_2$	
<b>5</b>	SE(d)	$\sqrt{[SE(E_1)]^2 + [SE(E_2)]^2}$	
<b>6</b>	IC à 95% (d)	bi (IC) = $d - 1,96SE(d)$ ; bs(IC) = $d + 1,96SE(d)$ .	
<b>7</b>	Test d'interaction	$z = d / SE(d)$	
<b>8</b>	ROR	exp(d)	
<b>9</b>	IC à 95% (ROR)	bi = $\exp[d - 1,96SE(d)]$ ; bs = $\exp[d + 1,96SE(d)]$	

OR<sub>1</sub>: rapport des cotes combinées issu de la méta-analyse A, OR<sub>2</sub>: rapport des cotes combinées issu de la méta-analyse B, log: logarithme, exp: exponentiel, bi= borne inférieure, bs: borne supérieure, SE erreur type,

La figure 8 représente un exemple de graphique des ROR d'une étude méta-épidémiologique comparant les effets traitement estimés entre les essais en ouvert par rapport aux essais en double aveugle. Le ROR combiné de tous les résultats (ROR = 0.87 [0.79 ; 0.97]) montre que l'absence du double aveugle est associée en moyenne à une exagération de 13% de l'effet du traitement.

Figure 8: ROR estimé des essais sans double aveugle par rapport aux essais en double aveugle



Source : Savovic et al 2012<sup>(122)</sup>

### 2.2.2 Exemples « célèbres » d'études de méta-épidémiologie

Plusieurs études méta-épidémiologiques ont montré que certaines caractéristiques méthodologiques sont associées à une surestimation des effets du traitement dans les essais randomisés<sup>(6,14,15)</sup>.

Parmi les premières études méta-épidémiologiques les plus connues, on se doit de citer la synthèse de Juni et al publiée dans le BMJ en 2001. Elle a évalué l'association entre les caractéristiques méthodologiques de l'essai et l'estimation de l'effet du traitement dans quatre grandes méta-analyses. Les caractéristiques méthodologiques étudiées étaient : la méthode pour générer la séquence de la randomisation, l'assignation secrète, l'absence d'aveugle et l'exclusion de patients de l'analyse.

Concernant le double aveugle, l'étude méta-épidémiologique a synthétisé les résultats de 4 études méta-épidémiologiques distinguant les études en ouvert des études en double aveugle en calculant le ROR. Un ROR < 1 indiquait que les ECR en ouvert donnaient une taille d'effet plus importante que les ECR en double aveugle. Ci-dessous, le détail de ces 4 études:

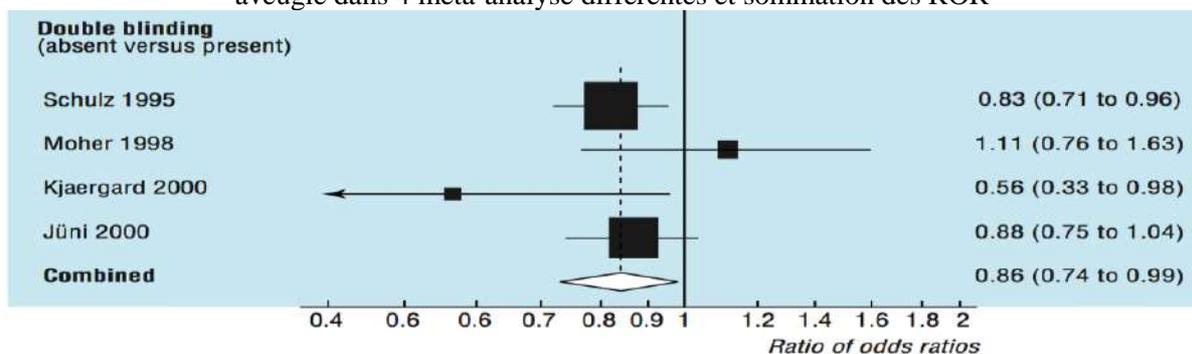
- L'étude Schulz est la première étude méta-épidémiologique publiée dans JAMA en 1995<sup>(8)</sup> qui a évalué la qualité méthodologique de 250 ECR à partir de 33 méta-analyses de la base de données Cochrane sur la grossesse et l'accouchement. Elle a étudié ainsi les associations entre le double aveugle et les effets estimés des traitements en calculant le ROR. Les ECR

en ouvert étaient associées à une surestimation de 17% de l'effet du traitement par rapport aux essais en double aveugle (ROR = 0.83 [0.71; 0.96])

- L'étude Moher en 1998<sup>(117)</sup> a évalué la qualité méthodologique de 127 ECR à partir de 11 méta-analyses ayant étudié l'efficacité des interventions utilisées pour les maladies circulatoires et digestives, la santé mentale, la grossesse et l'accouchement. Dans les essais en ouvert, aucune différence significative de l'effet des interventions n'a été observée par rapport aux essais en double aveugle (ROR = 1.11 [0.76 ; 1.63]).
- L'étude Kjaergard en 2000<sup>(123)</sup> a étudié 14 méta-analyses portant sur 190 ECR. Dans cette étude, les ECR en ouvert étaient associées à une exagération de 44% de l'effet du traitement par rapport aux ECR en double aveugle (ROR = 0.56 [0.33;0.98]). Il existe donc un biais important lié à l'absence de double aveugle.
- Dans l'étude Juni et al en 2000<sup>(124)</sup>, les essais en ouvert ont été associés à une surestimation de 12% de l'effet du traitement par rapport aux essais en double aveugle (ROR = 0.88 [0.75; 1.04])

La synthèse est présentée Figure 9<sup>(125)</sup>. Les essais en ouvert étaient au total associés une surestimation (exagération) de 14% de l'effet de traitement par rapport aux essais réalisés en double aveugle (ROR combiné = 0.86 [0.74; 0.99]). Les études en ouvert ont donné une ampleur d'effet plus importante que celle observée dans les études en double aveugle. Par conséquent, le risque de biais dû à l'absence de double aveugle a pu être quantifié. Cette méta-analyse a souligné l'importance de conduire un essai en double aveugle pour obtenir des résultats fiables.

**Figure 9:** Rapport des odds ratio (ROR) des essais sans double aveugle versus essais en double aveugle dans 4 méta-analyse différentes et sommation des ROR



Meta-analysis of four empirical studies relating key aspects of methodological quality of controlled trials to their effect estimates. Meta-analysis was by random effects model. Size of squares is proportional to inverse of variance of estimate

Source: Juni et al 2001<sup>(125)</sup>

- Concernant les autres caractéristiques méthodologiques de l'essai, l'étude Juni et al a montré aussi que ces caractéristiques peuvent affecter l'estimation de l'effet de traitement et conduire à des résultats biaisés :
  - Les essais ayant une séquence d'attribution inadéquate ont été associés à une surestimation de 19% de l'effet de traitement par rapport aux essais ayant une séquence d'attribution adéquate (ROR combiné = 0.81 [0.60; 1.09]).
  - Les essais ayant une assignation secrète inadéquate ont été associés à une surestimation significative de 30 % de l'effet de traitement par rapport aux essais ayant une assignation secrète adéquate (ROR combiné = 0.70 [0.62; 0.80]).

Au total, l'étude Juni et al a souligné l'importance de l'évaluation de la qualité méthodologique des ECR et la réalisation d'analyses de sensibilité. L'influence possible de la qualité de l'étude sur les estimations de l'effet devrait cependant toujours être examinée dans un ensemble donné d'études incluses. Des mises à jour récentes en 2012 [Savovic et al. *Ann Intern Med* 2012<sup>(122)</sup>] avec 234 méta-analyses incluant 1973 ECR et en 2017 [Savovic J et al. *Am J Epidemiol* 2017<sup>(126)</sup>] avec 228 méta-analyses Cochrane incluant 2443 ECR ont confirmé ces résultats et permis de réaliser des analyses en fonction des critères d'évaluation, et en multivarié.

D'autres études méta-épidémiologiques ont été conduites et se sont intéressées à d'autres sujets :

- L'influence de la taille des études sur la quantification de l'effet traitement [Nuësch et al *BMJ* 2010] : les petites études seraient associées à une majoration de l'estimation du bénéfice des traitements par rapport aux études de grande envergure (traduisant probablement un biais de publication), sur la base de 13 méta-analyses incluant 153 ECR (41 605 patients).
- L'influence du caractère monocentrique ou multicentrique des études [Dechatres et al <sup>(127)</sup>]: une étude méta-épidémiologique a été réalisée sur la base de 48 méta-analyses (421 ECR) publiées dans des journaux avec le facteur d'impact le plus élevé en 2008 et 2010. La méta-analyse, ainsi que les ECR, ont été analysées avec les critères d'évaluation méthodologique de Cochrane RoB2. L'échantillon de patients était plus faible dans les études monocentriques avec une moyenne de 90 vs 243 dans les études multicentriques. Le risque de biais dans les ECR était moins fréquent dans les études multicentriques pour : le secret de l'attribution, la dissimulation des allocations, l'élaboration de la séquence de randomisation, le double insu (patients, des chercheurs/praticiens et des évaluateurs) et la mention sélective des résultats. Les résultats de cette étude ont montré une différence significative de l'effet traitement entre les ECR monocentriques et les ECR multicentriques

(ROR combiné = 0.73 [0.64, 0.83]) soit une majoration significative de l'effet traitement de 27% dans les études monocentriques. Ce risque de biais doit inciter les auteurs de méta-analyses à faire une analyse de sensibilité, suivant le caractère monocentrique ou multicentrique des études incluses, pour montrer comment leurs conclusions pourraient être affectées par l'exclusion des essais monocentriques de l'analyse.

- L'influence de la taille des ECR<sup>(128)</sup>: dans le même ordre d'idée, une étude méta-épidémiologique a été réalisée afin d'évaluer l'influence de l'effectif de l'essai sur l'effet traitement. Cette étude a montré que les essais de petite taille ou de taille modérée ont donné un effet traitement plus important que les essais de grande taille dans une méta-analyse. Par rapport aux plus grands essais, les petits essais étaient associés à une surestimation significative de 32% de l'effet traitement (ROR combiné = 0.68 [0.57, 0.82]). Les essais de taille modérée étaient également associés à une surestimation significative de 17% de l'effet de traitement (ROR combiné = 0.83 [0.75, 0.91]).
- L'influence du mode de financement des ECR<sup>(129)</sup>: la plupart des essais cliniques évaluant les médicaments sont financés par des industries pharmaceutiques. Compte tenu des enjeux, les financements des études industrielles concernent très souvent des études avec des médicaments très « prometteurs » en termes d'efficacité et donc de marché, sans compter les possibles conflits d'intérêt. Plusieurs études ont montré que les essais financés ont donné des résultats plus favorables au traitement testé (produit par le laboratoire) que les essais non sponsorisés/financés par l'industrie pharmaceutique<sup>(130,131)</sup>. Ainsi, une étude méta-épidémiologique a été conduite afin d'évaluer l'influence du financement de l'essai par l'industrie pharmaceutique sur l'estimation de l'effet du traitement. L'analyse a inclus 67 méta-analyses (363 ECR) évaluant différents types de traitements en soins intensifs avec des critères binaires. Elle a montré qu'il n'y a finalement pas de différence significative de l'estimation de l'effet traitement expérimental entre les essais financés par l'industrie et les essais non financés (ROR combiné = 1.10 [0.96, 1.26])
- L'influence du cluster : l'influence de l'utilisation d'un design en cluster sur les estimations de l'effet traitement a été également évaluée dans une étude méta-épidémiologique<sup>(132)</sup>. Pour les critères binaires, les essais en cluster ont fourni des estimations de l'effet traitement similaires aux ECR (ROR combiné = 1.00 [0.93, 1.08]). Cette étude a montré la possibilité de regrouper sans risque les deux types d'essais (essais en cluster et essais randomisés) dans la méta-analyse pour les critères binaires.

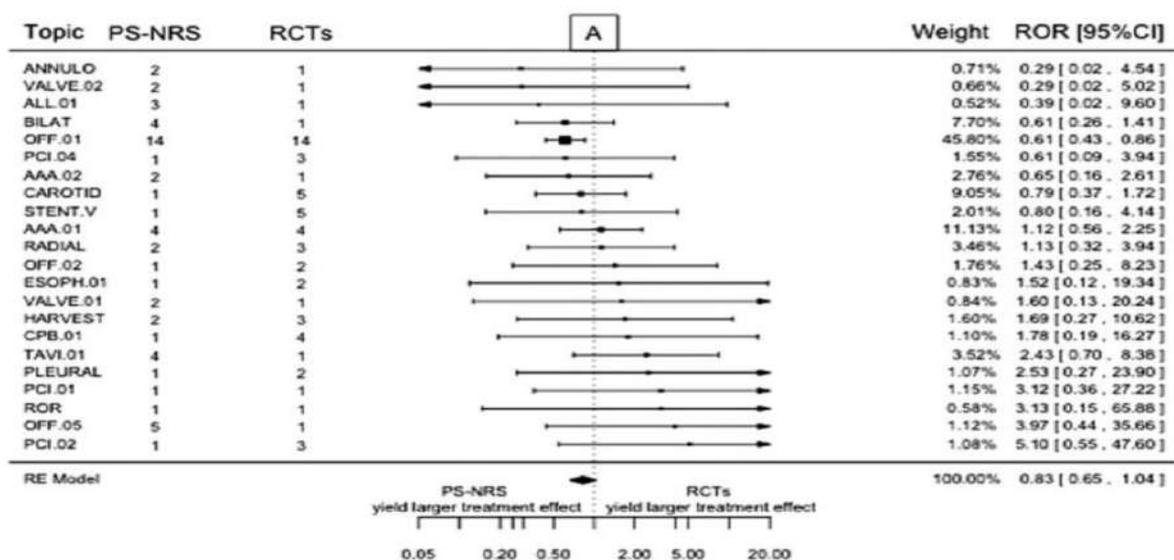
- L'influence du moment de la conduite de l'étude : une étude a été réalisée pour comparer l'estimation de l'effet traitement entre les essais pivots et les essais post-AMM<sup>(133)</sup>. Les effets du traitement étaient 50 % plus élevés (plus bénéfiques) dans les essais pivots que dans les essais post-AMM (ROR combiné= 1.50 [1.01, 2.23]).

Nous allons voir plus en détail l'utilisation de l'outil méta-analytique pour comparer l'estimation des effets traitements dans les ECR et dans les études observationnelles, afin d'évaluer les résultats que nous pouvons attendre avec les AOD dans le traitement de la MTEV et dans la FA.

### 2.2.3 Méta-épidémiologie, études observationnelles et ECR

L'influence du biais de confusion inhérent aux études observationnelles sur l'effet du traitement peut être évalué par une approche méta-épidémiologique. Plusieurs études méta-épidémiologiques ont comparé l'estimation de l'effet du traitement entre les études observationnelles et les ECR. Ces études méta-épidémiologiques ont conclu que l'estimation de l'effet du traitement par les études observationnelles peut être surestimée<sup>(134,135)</sup>, sous-estimée<sup>(136)</sup>, ou être identique<sup>(137,138)</sup> par rapport à celle fournie par les ECR. Par exemple, l'étude Lonjon et al a comparé l'effet traitement entre les études non randomisées utilisant le SP (70 études) et les 94 ECR<sup>(137)</sup>. Toutes ces études incluses ont évalué une technique chirurgicale, rapportant un critère subjectif et un critère objectif (la mortalité toutes causes). Concernant la mortalité toutes causes, aucune différence significative de l'estimation de l'effet de traitement n'a été démontrée entre les ECR et les études non randomisées utilisant un SP (ROR combiné = 0.83 [0.65, 1.04]) (figure 11). De même pour le critère subjectif, aucune différence significative n'a été démontrée (ROR combiné = 1.07 [0.87, 1.33])<sup>(137)</sup>.

Figure 10: ROR combiné entre les études observationnelles et les ECR



Source: Lonjon et al 2014. Ann Surg. 2014;259(1):18-25.

Cette absence de différence dans l'estimation de l'effet du traitement sur cette seule étude ayant comparé les résultats issus d'ECR et les résultats issus d'études observationnelles avec tentative de correction du biais de confusion avec un SP ne permet pas de convaincre que ces dernières fournissent des résultats non biaisés, au même titre que les ECR, de l'estimation de l'effet du traitement.

L'outil méta-analytique est utilisé pour la conduite des méta-analyses, mais aussi pour identifier les sources de biais liées aux caractéristiques des études, et par là-même convaincre de l'importance de la méthodologie dans la conduite des études évaluant les médicaments. Des recommandations ont été développées afin d'aider les auteurs à rédiger de façon standardisée la méthodologie et les résultats de leurs études. Les objectifs de ces lignes directrices sont la valorisation des résultats de recherche et la possibilité pour les lecteurs d'évaluer/critiquer les résultats et d'assurer la reproductibilité des travaux.

#### 2.2.4 Lignes directrices proposées pour les études méta-épidémiologiques

Tout d'abord concernant les méta-analyses, l'écriture d'une revue systématique doit respecter une méthodologie rigoureuse afin de garantir sa qualité et de permettre son évaluation par les lecteurs. Des guides mis à disposition des auteurs ont comme objectif de standardiser le format des rapports de revues systématiques. Dans ce sens :

- un *guide QUOROM* a été mis en place en 1998 pour les auteurs des études méta-analyses des essais randomisés<sup>(139)</sup>. Il a permis d'améliorer la méthode de réalisation des revues systématiques.
- Une mise à jour a été ensuite développée par Moher et al en 2009 : *le PRISMA* statement (lignes directrices pour la rédaction des rapports). Le PRISMA est constitué d'une liste de contrôle de 27 points et d'un diagramme de flux en quatre phases (*flow chart*), permettant de faciliter aussi bien l'écriture que la lecture de la revue systématique<sup>(140)</sup>.
- La collaboration Cochrane a également élaboré son propre guide, *MECIR* (*Methodological Expectations of Cochrane Intervention reviews*).
- PROSPERO est un registre international prospectif des revues systématiques, qui a été élaboré pour améliorer la qualité des revues systématiques mondiales. Ce registre international est la première base d'enregistrement des revues systématiques dans le domaine de la santé<sup>(141)</sup>.

Les études méta-épidémiologiques se basent sur des méta-analyses afin d'examiner l'impact de certaines caractéristiques des études cliniques sur l'effet de traitement. Les études méta-épidémiologiques doivent ainsi répondre à des normes similaires à celles des méta-analyses. Elles exigent ainsi une recherche systématique pour sélectionner les données. Des lignes directrices ont été proposées par Murad et al pour rapporter les études méta-épidémiologiques, les éléments utilisés dans l'énoncé PRISMA<sup>(142)</sup> ont été adaptés au contexte des études méta-épidémiologiques. Les éléments sont classés en fonction du titre et résumé, de l'introduction, des méthodes, des résultats, des conclusions et des financements<sup>(143)</sup>:

- Titre et résumé. Tout d'abord, l'étude doit être bien identifiée comme une «étude méta-épidémiologique» dans le titre et dans le résumé afin de permettre une meilleure classification de l'étude, une facilitation de la recherche et de la sélection de l'étude. Dès lors, elle permet de retrouver rapidement des résultats pertinents. Le résumé des études méta-épidémiologiques doit être également bien structuré puisqu'une partie des chercheurs ne lisent, à tort, que le résumé d'une étude. Il doit contenir des sous-titres qui précisent le contexte, l'hypothèse, les méthodes, les sources de données, les résultats, les limites et les conclusions<sup>(143)</sup>.
- Introduction. Le contexte et les objectifs d'une étude méta-épidémiologique doivent apparaître dans l'introduction. Cette exigence est similaire à celle de PRISMA<sup>(142)</sup>.
- Méthodes. Dans la partie méthodes, il faut préciser si l'étude méta-épidémiologique suit a priori un protocole bien défini. Toutefois, l'enregistrement du protocole n'est pas nécessaire

puisque les bases d'enregistrement des revues systématiques sont plutôt orientées sur des questions cliniques. Il faut également définir dans cette partie les critères d'éligibilité des études, les sources d'information, les méthodes de recherche, le processus de collecte des données et la sélection de l'étude, les données recueillies, les mesures et les méthodes de synthèse y compris les méthodes d'analyses supplémentaires. La stratégie de recherche dans une étude méta-épidémiologique est différente de celle d'une revue systématique parce qu'elle n'est pas conduite pour répondre à une question clinique. La mesure de l'effet du traitement utilisée doit être détaillée dans l'analyse statistique puisque certaines mesures, comme les coefficients de régression et le ROR, ne sont pas habituellement utilisées dans les autres types d'études évaluant l'effet du traitement. La classification des données dans une étude méta-épidémiologique demande beaucoup plus de discussion entre les chercheurs que dans une revue systématique clinique. Le consensus entre les chercheurs nécessite ainsi un temps plus important dans une étude méta-épidémiologique que dans une revue systématique clinique<sup>(143)</sup>.

- Résultats. Les résultats du processus de sélection des études doivent apparaître dans la partie résultats des études méta-épidémiologiques. La description des caractéristiques des études et des patients doit être adaptée à la question méta-épidémiologique abordée. En méta-épidémiologie, il n'est pas nécessaire de connaître ou de décrire les caractéristiques ou les comorbidités des patients. Une évaluation du biais de publication est indispensable comme pour une méta-analyse standard et une présentation de cette évaluation doit être faite dans la partie résultats selon les recommandations de PRISMA<sup>(142)</sup>.
- Conclusions : la partie discussion est constituée d'un résumé des principaux résultats et une comparaison avec les résultats des autres études existantes sur le même sujet. Les limites de la méthodologie de recherche et les points forts de l'étude méta-épidémiologique conduite doivent également être discutés dans cette partie. Les études méta-épidémiologiques ont des conséquences sur les futures recherches et la pratique clinique. Ces implications doivent être décrites aussi dans cette partie discussion<sup>(143)</sup>.
- Financements : les sources doivent être citées.

La suite du travail étant une présentation d'études méta-épidémiologiques, nous nous efforcerons de respecter ces différentes lignes directrices. Nous allons nous intéresser aux données des ECR et des études observationnelles, avec ou sans SP, dans un domaine d'actualité très riche qu'est celui des AOD.

# 3 Quantification du biais d'évaluation dans les études évaluant les anticoagulants oraux dans la maladie thromboembolique veineuse

Dans cette partie, nous allons présenter l'étude méta-épidémiologique qui a été effectuée afin de quantifier les biais d'estimation des effets des traitements dans les études observationnelles évaluant les AOD chez les patients avec MTEV. Pour atteindre cet objectif, nous allons déterminer les différences entre les ECR et les études observationnelles sur :

- La population des patients inclus
- Les résultats en termes d'efficacité et de sécurité.

Ce travail a été soumis pour publication dans *Journal of Clinical Epidemiology* et est under Review. L'article complet est fourni Annexe 3.

Le protocole de cette étude méta-épidémiologique a été rédigé conformément aux lignes directrices de PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) et la liste de contrôle PRISMA est fournie en annexe 2-tableau A2.

The screenshot shows the Editorial Manager interface for the Journal of Clinical Epidemiology. At the top, there is a navigation bar with links for HOME, LOGOUT, HELP, REGISTER, UPDATE MY INFORMATION, JOURNAL OVERVIEW, MAIN MENU, CONTACT US, SUBMIT A MANUSCRIPT, INSTRUCTIONS FOR AUTHORS, and PRIVACY. The user's role is set to 'Author' and the username is 'silvy.laporte@chu-st-etienne.fr'. Below this, a section titled 'Revisions Being Processed for Author Silvy Laporte' displays a table with the following data:

Action	Manuscript Number	Title	Date Submission Began	Status Date	Current Status
<a href="#">Action Links</a>	JCE_2020_554R1	Observational studies versus randomized clinical trials with direct oral anticoagulants: a meta-epidemiological approach	Oct 01, 2020	Oct 05, 2020	Required Reviews Completed

## 3.1 Recherche exhaustive et sélection des études

La première étape de cette étude méta-épidémiologique est la revue systématique de la littérature. Elle a été réalisée afin d'identifier les études randomisées et les études observationnelles évaluant les AOD par rapport au traitement standard chez les patients avec MTEV. Des critères d'éligibilité ont été définis afin de déterminer toutes les études éligibles à l'inclusion.

### 3.1.1 Critères d'éligibilité des études

- Les essais randomisés et les études observationnelles devaient avoir évalué les AOD par rapport aux AVK dans le traitement des ETEV sur le risque de récurrence thromboembolique et les hémorragies majeures.
- Ces études devaient avoir comparé les AOD par rapport aux AVK pendant au moins 3 mois dans le traitement initial des ETEV. Ainsi les études chez les patients avec cancer n'étaient pas retenues, car le traitement comparateur était alors les héparines de bas poids moléculaire HBPM.
- L'objectif des études devait être l'évaluation de l'efficacité et de sécurité des anticoagulants.
- Les comparaisons ont été limitées aux AOD approuvés (ou en cours d'examen réglementaire) pour le traitement des ETEV, notamment le dabigatran, le rivaroxaban, l'apixaban et l'edoxaban par rapport à la warfarine orale ou à un autre AVK.

### 3.1.2 Revue de la littérature

- La recherche exhaustive des essais randomisés et des études observationnelles, réalisée dans les bases de données électroniques PubMed, Cochrane Database, Google scholar, porte sur la période allant du 1er janvier 2009 au 30 août 2020 (la première étude sur les AOD a été publiée en 2009). Une équation de recherche a été construite pour identifier les articles sur PubMed (cf. annexe2-tableau A4), avec les mots clés suivants : "venous thromboembolism, acute venous thromboembolism, symptomatic venous thromboembolism, symptomatic pulmonary embolism, symptomatic deep vein thrombosis, oral anticoagulant, rivaroxaban, dabigatran, apixaban, edoxaban, direct oral anticoagulants, novel oral anticoagulant". Aucune restriction sur la langue de publication n'a été effectuée.
- La recherche systématique amène un grand nombre d'études éligibles à l'inclusion. Parmi ces études éligibles, certaines études sont conduites sur des mêmes bases de données de santé informatisées avec des période d'inclusion identique. Ces études ont été considérées comme des doublons si plus de la moitié de la période de recrutement se chevauchait avec une autre étude ayant un effectif plus grand et ayant évalué les mêmes groupes de traitement chez les mêmes patients. Lorsque des doublons ont été identifiés, l'étude avec l'effectif le plus important a été incluse dans la méta-analyse.

Un comité de pilotage de la méta-analyse était constitué de trois personnes. Les résultats de la recherche bibliographique ont été examinés par ce comité. Ainsi, les essais sélectionnés ont été ensuite vérifiés par le comité. Cette recherche exhaustive a permis d'identifier les études randomisées et les études de cohortes ayant étudié les AOD versus AVK dans le traitement des ETEV sur le risque de récurrences et les hémorragies majeures.

### 3.1.3 Analyses des études éligibles

La qualité d'une méta-analyse dépend avant tout de la qualité des études qui la composent. C'est pourquoi il est important d'évaluer la qualité méthodologique par le risque de biais correspondant à chaque étude. Dans le cas où la qualité d'une étude était insuffisante, son exclusion a été discutée avec le comité de pilotage de la méta-analyse pour décider de son exclusion ou de son maintien dans la méta-analyse, en prévoyant des analyses de sensibilité sans cette étude. Par exemple, des analyses de sensibilité sur les deux critères « récurrences ETEV et hémorragies majeures » ont été réalisées en excluant les études observationnelles n'ayant pas utilisé de méthodes d'ajustement permettant de limiter un potentiel biais de confusion lié à l'absence de randomisation.

### 3.1.4 Critères de jugement

- Le critère de jugement principal en termes d'efficacité était la survenue d'une récurrence d'ETEV. Il était généralement le critère principal de tous les essais randomisés évaluant les AOD dans cette indication. Il s'agit d'un critère composite incluant les TVP proximales et les EP fatales ou non fatales.
- Le critère de jugement en termes de sécurité était la fréquence des hémorragies majeures. Ce critère est également un critère composite défini selon la définition de ISTH<sup>(76)</sup>.

### 3.1.5 Extraction des données

Après avoir identifié les études éligibles (revue de la littérature), les données ont été collectées indépendamment par deux personnes. Pour les études randomisées et observationnelles:

- Les caractéristiques des études étaient : la date de publication, les noms des auteurs, le pays où l'étude a été réalisée, le nombre de patients (total et dans chaque groupe), la conception de l'étude et la période d'inclusion.

- Les caractéristiques des traitements étaient : le type d'AOD, la dose reçue et la durée de traitement. Le traitement par anticoagulant dans ces études était de 3 à 12 mois.
- Les caractéristiques des patients comprenaient : l'âge moyen, le sexe, l'insuffisance rénale, les antécédents de MTEV et le cancer.
- Les estimations des effets du traitement AOD versus AVK ont été recueillis avec le HR ou le RR pour les récurrences ETEV et les hémorragies majeures.

Les études observationnelles ont été classées en trois catégories : 1) les études de cohortes prospectives comprenant des études observationnelles avec recrutement prospectif et une base de données construite spécifiquement pour l'étude, 2) les études sur bases vivantes comprenant des études observationnelles avec recrutement prospectif incluses dans une base vivante clinique (c'est-à-dire un registre) ou une base vivantes médico-administratives et 3) les études de cohortes rétrospectives avec recrutement rétrospectif.

De plus nous avons recueilli :

- Les méthodes d'appariement ou d'ajustement utilisées (aucune, score de propension (SP) ou modèle multivarié)
- Les caractéristiques des patients avant et après SP si applicable.
- Les HR et RR après appariement ou ajustement.

### 3.2 Description des études et des populations incluses

Les caractéristiques des populations des ECR et des études observationnelles ont été comparées.

#### 3.2.1 Identification et sélection des études.

Quatre cent treize références éligibles entre 2009 et 2020 ont été identifiées. Après vérification des titres et des résumés, 366 références ont été exclues. Après évaluation des textes intégraux, 17 autres références ont été exclues. Les raisons des exclusions sont fournies Figure 11. Au total, 30 références ont été considérées comme éligibles dans la méta-analyse.

Plusieurs méta-analyses ont déjà été conduites avec les ECR, l'identification et la sélection de celles-ci n'ont pas posé de problème spécifique.

Parmi les études observationnelles, une attention particulière a été portée à l'analyse des doublons :

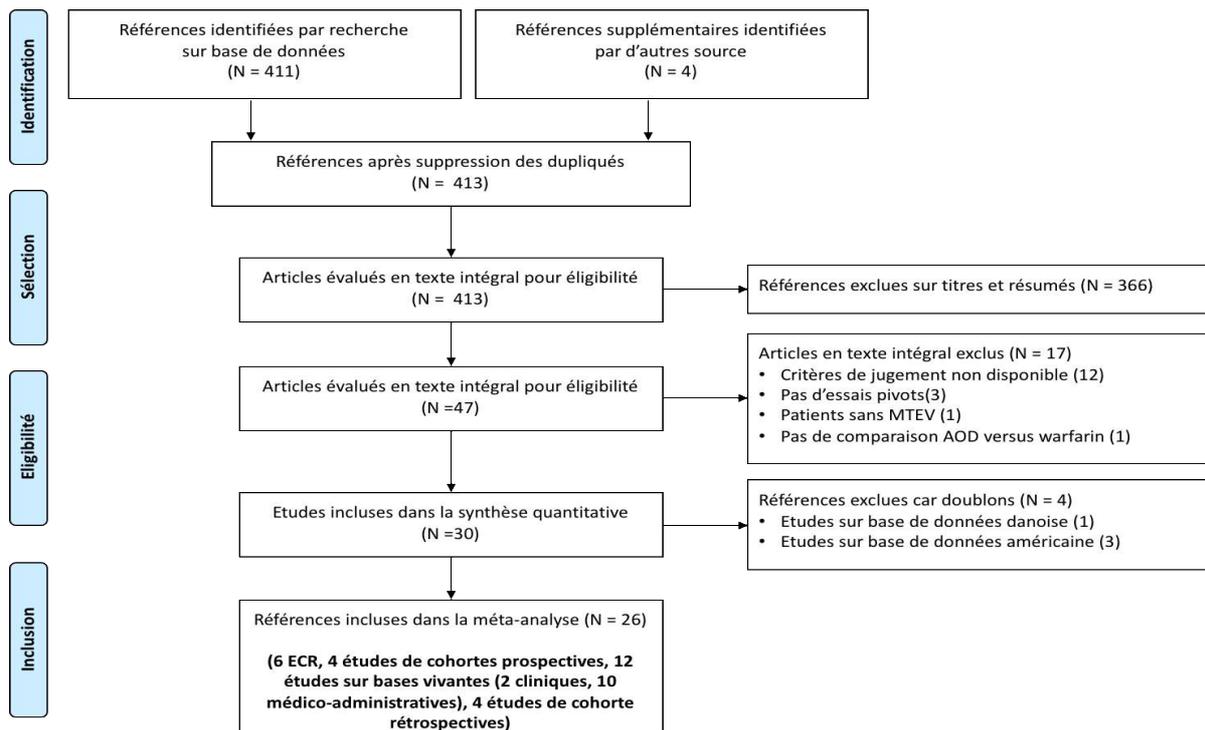
- Deux études ont été réalisées sur la même base de données danoise (Danish national prescription registry) :
  - Sindet-Pedersen et al 2017 incluant 12 318 patients avec MTEV entre février 2012 et septembre 2016<sup>(144)</sup>.
  - Larsen et al incluant 5 004 patients avec MTEV non provoquée entre décembre 2011 et février 2016<sup>(145)</sup>.

Ces deux études ont été considérées comme doublons car plus de la moitié de la période de recrutement de la deuxième étude (Larsen et al) se chevauchait avec la première étude (Sindet-Pedersen et al). Cette dernière étude avait un effectif plus grand et avait évalué les mêmes groupes de traitement chez les mêmes patients. La deuxième étude (Larsen et al) a été considéré comme un sous-groupe de la première publication (Sindet-Pedersen et al) et a donc été exclue.

- Cinq études évaluant le rivaroxaban chez les patients avec MTEV ont été conduites sur la même base américaine (US Truven market scan):
  - Coleman 2017 incluant 45 853 des patients avec MTEV non provoquée entre janvier 2012 et juin 2015<sup>(146)</sup>
  - Coleman et al 2018 incluant 17 618 patients avec MTEV provoquée entre janvier 2012 et décembre 2016<sup>(147)</sup>
  - Coleman et al 2018 incluant 806 patients avec MTEV et présentant un état d'hypercoagulabilité entre Juin 2012 et septembre 2015<sup>(148)</sup>
  - Kohn et al 2018 incluant 36 853 patients avec MTEV non provoquée entre janvier 2012 et décembre 2016<sup>(149)</sup>
  - Coleman et 2018 incluant 47 342 patients avec MTEV entre janvier 2012 et décembre 2016<sup>(150)</sup>

Les trois dernières études ont été considérées comme doublons des deux premières études. En effet, ces trois études ont évalué les mêmes patients avec un effectif plus faible et une période de recrutement plus courte et se chevauchant sur plus de la moitié avec les deux premières études<sup>(146,147)</sup>. Ces trois études ont donc été exclues.

Figure 11: Diagramme de flux du processus de sélection des études



Au total, 26 études (276 092 patients) ont été incluses dans l'analyse globale (Figure 11) :

- Six ECR incluant un total de 27 121 patients : deux essais évaluant le rivaroxaban<sup>(79,80)</sup>, deux essais évaluant le dabigatran<sup>(77,78)</sup>, un essai évaluant l'apixaban<sup>(81)</sup> et un essai évaluant l'edoxaban<sup>(82)</sup>.
- Quatre études de cohortes prospectives incluant 8 704 patients et évaluant toutes le rivaroxaban : 2 études ont utilisé la méthode de SP (1 appariement sur SP et 1 ajustement sur SP) et les 2 autres n'ont pas utilisé des méthodes de correction.
- Douze études sur bases vivantes incluant 227 894 patients ( 2 sur bases vivantes cliniques et 10 sur bases vivantes médixio-administartives): six études évaluant le rivaroxaban<sup>(151,152,146,147,144,153)</sup>, trois études évaluant l'apixaban une étude évaluant 3 AOD (apixaban, dabigatran et rivaroxaban)<sup>(154)</sup>, et une autre étude évaluant 4 AOD (apixaban, dabigatran, edoxaban et rivaroban)<sup>(155)</sup>. Onze études ont utilisé des méthodes d'ajustement afin de minimiser les biais : dix études ont utilisé un SP (6 appariements sur SP, 4 pondérations inverses sur SP) et une seul études a appliqué un modèle de régression multiple.
- Quatre études de cohortes rétrospectives incluant 1 536 patients : deux études ont évalué les 4 AOD (apixaban, rivaroxaban, dabigatran et edoxaban)<sup>(156,157)</sup>, une étude a évalué les 3 AOD (apixaban, rivaroxaban et dabigatran)<sup>(158)</sup> et une autre étude a évalué le

rivaroxaban<sup>(159)</sup>. Ces quatre études n'ont pas appliqué des méthodes d'ajustement. Les caractéristiques des études incluses sont décrites dans l'annexe2- tableau A5.

### 3.2.2 Comparaison des populations ECR et des études observationnelles

Des analyses statistiques descriptives ont été effectuées afin de comparer les populations incluses entre les ECR et les études observationnelles évaluant les AOD dans le traitement de la MTEV. Les caractéristiques des patients incluses dans ces études ont été comparées avant et après l'application du SP. Pour chaque type d'étude, les variables catégorielles extraites sous forme de pourcentages ainsi que les variables continues extraites sous formes de moyennes ont été présentées par une médiane non pondérée. Les caractéristiques des populations des études incluses sont présentées dans l'annexe2-tableau A3.

Les caractéristiques des patients traités par AOD selon le type d'étude ont été décrites dans le tableau 3 . Les caractéristiques des patients traités par AOD dans les études de cohortes prospectives étaient similaires aux patients inclus dans les ECR. A titre d'exemple l'âge médian des patients était de 58 ans dans les cohortes prospectives et de 56 ans dans les ECR et la gravité des maladies (Tableau 3). Néanmoins, les caractéristiques des patients sous AOD dans les études de cohortes rétrospectives et dans les études sur bases vivantes sont différentes de celles incluses dans les ECR. En effet, l'âge médian des patients traités par AOD était plus élevé dans ces études que dans les ECR avec un écart d'âge médian de 7 ans. Les patients sous AOD inclus dans des études de cohorte rétrospectives ou les études sur bases vivantes présentaient également un plus grand nombre de comorbidités, laissant supposer que ces études permettaient d'étudier des patients moins sélectionnés que les études de cohortes prospectives dévolues à l'étude des AOD.

Compte tenu des similitudes et des différences des populations entre les différents types d'études, la comparaison des estimations d'efficacité et de sécurité des AOD entre les ECR et les études de cohortes prospectives ne devrait traduire qu'un biais d'estimation dans les études de cohortes prospectives, dû à l'absence de randomisation, alors que la différence des effets des traitements entre ECR et études de cohortes rétrospectives et études sur bases de données vivantes pourrait traduire un biais d'estimation du fait de l'absence de randomisation ou d'une interaction de l'effet traitement avec les caractéristiques des patients.

Quels sont les résultats des estimations? C'est ce que nous allons décrire dans la partie suivante.

**Tableau 3:** Caractéristiques des patients des groupes AOD selon le type d'étude avant tout ajustement ou appariement

	<b>Essais contrôlés randomisés (6 études)</b>	<b>Études de cohortes prospectives (4 études)</b>	<b>Études sur bases vivantes (registre et BMA) (12 études)</b>	<b>Études de cohortes rétrospectives ( 4 études)</b>
<b>Nombre des études</b>	6	4	6	4
<b>Age (années)</b>	56 (55-57)	58 (57-59)	63 (60-64)	63 (63-65)
<b>Sexe masculin (%)</b>	58 (57-58)	52 (45-55)	53 (50-55)	50 (39-69)
<b>Insuffisance rénale (%)</b>	7 (7-7)	8 (6-9)	12 (9-20)	12 (12-20)
<b>Antécédents d'ETEV (%)</b>	19 (19-19)	17 (10-23)	17 (15-23)	10 (6-14)
<b>Cancer (%)</b>	5 (4-6)	8 (6-11)	11 (4-14)	17 (8-25)
<b>EP à l'inclusion (%)</b>	33 (31-39)	23 (13-33)	45 (41-90)	35 (30-45)

### 3.3 Approche méta-analytique : comparaison des estimations des effets du traitement entre les études randomisées et les études observationnelles

#### 3.3.1 Méthodes

Des méthodes méta-analytiques ont été utilisées afin de comparer l'effet du traitement des AOD par rapport aux AVK entre les ECR et les études observationnelles (études de cohortes prospectives, études utilisant des bases de données de santé informatisées ou études de cohortes rétrospectives) dans le traitement des ETEV. Pour chaque type d'étude, un Hazard ratio (HR) commun avec son intervalle de confiance (IC) à 95% a été calculé pour le risque de récurrences thromboemboliques et des hémorragies majeures. Les méta-analyses ont été réalisées avec la méthode du logarithme des HR en pondérant les études par l'inverse de la variance. Lorsque la valeur du HR n'était pas disponible dans l'article, celui-ci a été approximé par le RR.

Pour chaque méta-analyse :

- L'hétérogénéité entre les études a été évaluée à l'aide du test statistique Q de Cochrane
- La cohérence entre les études a été mesurée à l'aide de l'indicateur  $I^2$ <sup>(160)</sup>.

La méta-analyse était basée sur un modèle à effets aléatoires<sup>(161,162)</sup>.

Les résultats obtenus ont été considérés statistiquement significatifs si la valeur de  $P$  était  $<0.05$ . Les résultats de la méta-analyse sont représentés par un «Forest plot». Un HR égal à 1 indique l'absence de différence entre les deux traitements (AOD et AVK), un HR  $<1$  indique que les AOD sont plus efficaces que les AVK, un HR  $>1$  indique que les AVK sont plus efficaces que les AOD.

Des graphiques en entonnoir de Begg ont été utilisés d'une part pour détecter les biais potentiels des études incluses dans la méta-analyse par la présence d'une asymétrie des graphiques et d'autre part pour montrer l'effet de l'étude de petite taille par rapport à l'étude de grande taille selon le type d'étude<sup>(163)</sup>. En outre, pour quantifier la différence de l'effet du traitement estimé entre les études observationnelles et les ECR, le ratio des HR (RHR) a été calculé en divisant les valeurs des HR obtenues dans les études observationnelles par celles obtenues dans les ECR. Le RHR correspond à l'association entre le type d'étude et la taille de l'effet du traitement. Un RHR  $< 1$  signifie que les estimations de l'effet du traitement sont plus élevées dans les études

observationnelles que dans les ECR, un RHR > 1 signifie que les estimations de l'effet du traitement sont plus faibles dans les études observationnelles. Des IC de 95% ont été calculés en utilisant des erreurs types qui prennent en compte l'hétérogénéité des études incluses<sup>(126,164,165)</sup>. Les méta-analyses ont été réalisées avec le logiciel R version 3.5.1 (meta package, téléchargé sur [www.r-project.org](http://www.r-project.org))<sup>(166)</sup>.

### 3.3.2 Résultats

Pour rappel, les ECR ont montré que les AOD étaient non inférieurs aux AVK si on se fie aux bornes de non-infériorité utilisées dans les différentes études (HR = 0.89 [0.76 ; 1.05]) du risque de récurrence de ETEV, sans différence significative entre les deux classes pharmacologiques. On observe par contre une réduction significative de 39% du risque d'hémorragie majeure (HR = 0.61 [0.45 ; 0.83]).

La comparaison des études de cohortes prospectives aux ECR permet de faire les observations suivantes :

- Dans les études de cohortes prospectives, les AOD ont été associés à une diminution non significative de 40% (HR = 0.60, [0.35 ; 1.02]) du risque de récurrence d'ETEV et une diminution significative de 43% du risque d'hémorragies majeures (HR = 0.57, [0.37 ; 0.89]) (Figures 14 et 15).
- Par rapport aux ECR, les études de cohorte prospectives sont associées à une surestimation non significative de 33 % de l'effet des AOD (RHR = 0.67 [0.39 ; 1.18]) en termes d'efficacité. Nous n'avons pas observé une association entre le type de l'étude et l'effet du traitement en termes de sécurité (RHR = 0.93 [0.55; 1.60]) (Figure 16).
- L'analyse de sensibilité excluant les études n'ayant pas utilisé d'ajustement ou d'appariement pour limiter les biais dus à l'absence de randomisation<sup>(167,168)</sup> fournit des résultats similaires: RHR = 0.93 [0.57 ; 1.53] pour le critère d'efficacité, RHR = 1.30, [0.64 ; 2.61] pour le critère de sécurité.

Figure 12: Graphique en entonnoir des résultats sur les récidives d'EVEV en fonction des types d'études

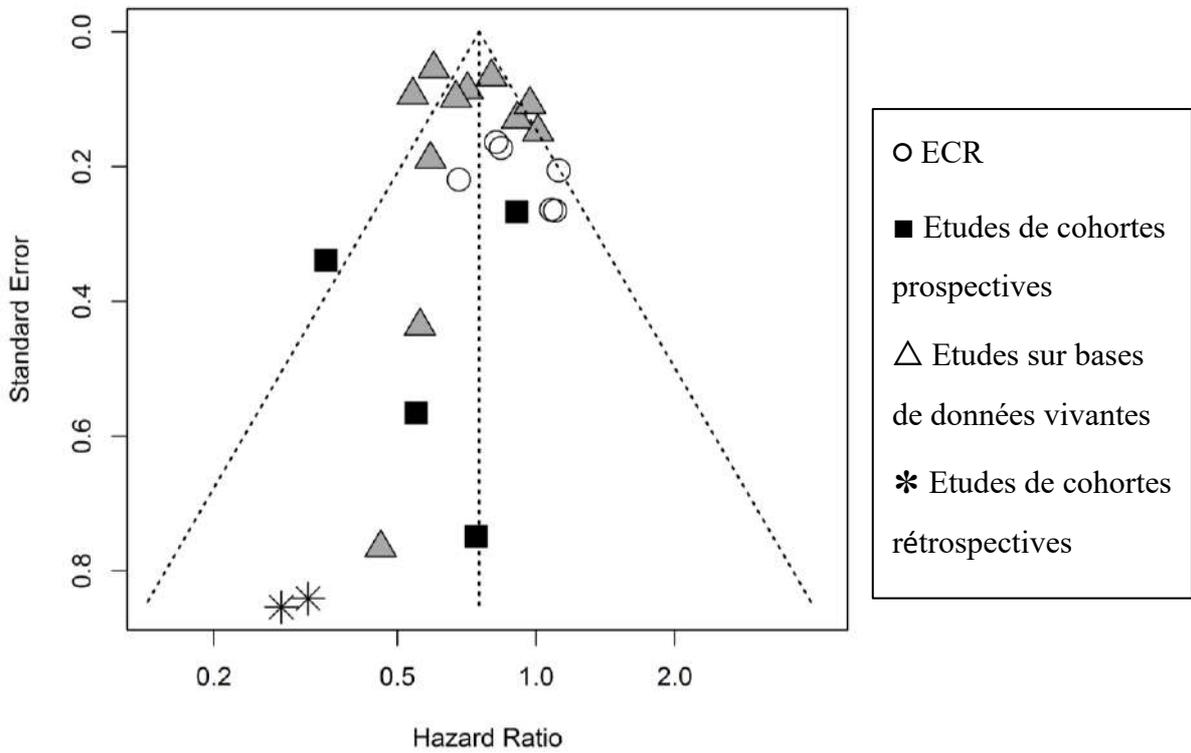


Figure 13: Graphique en entonnoir des résultats sur hémorragies majeures en fonction des types d'études

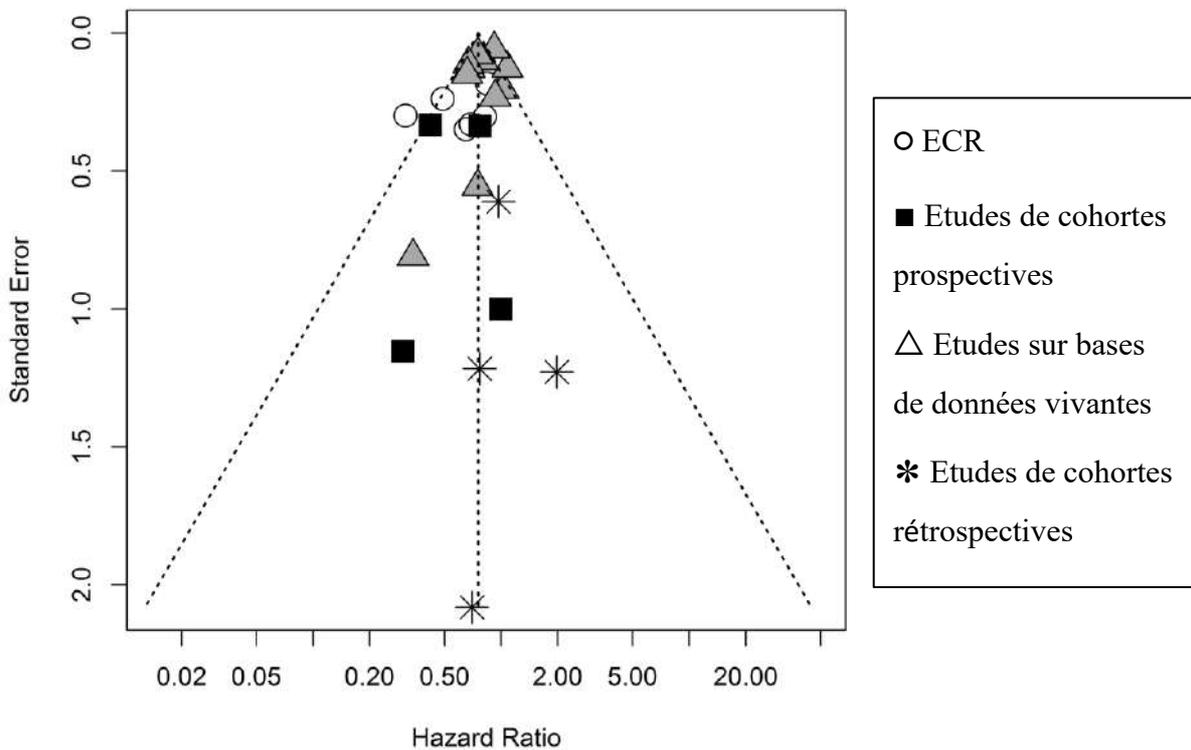


Figure 14: Comparaison des estimations des effets du traitement en fonction des types d'études sur les récurrences d'ETEVE

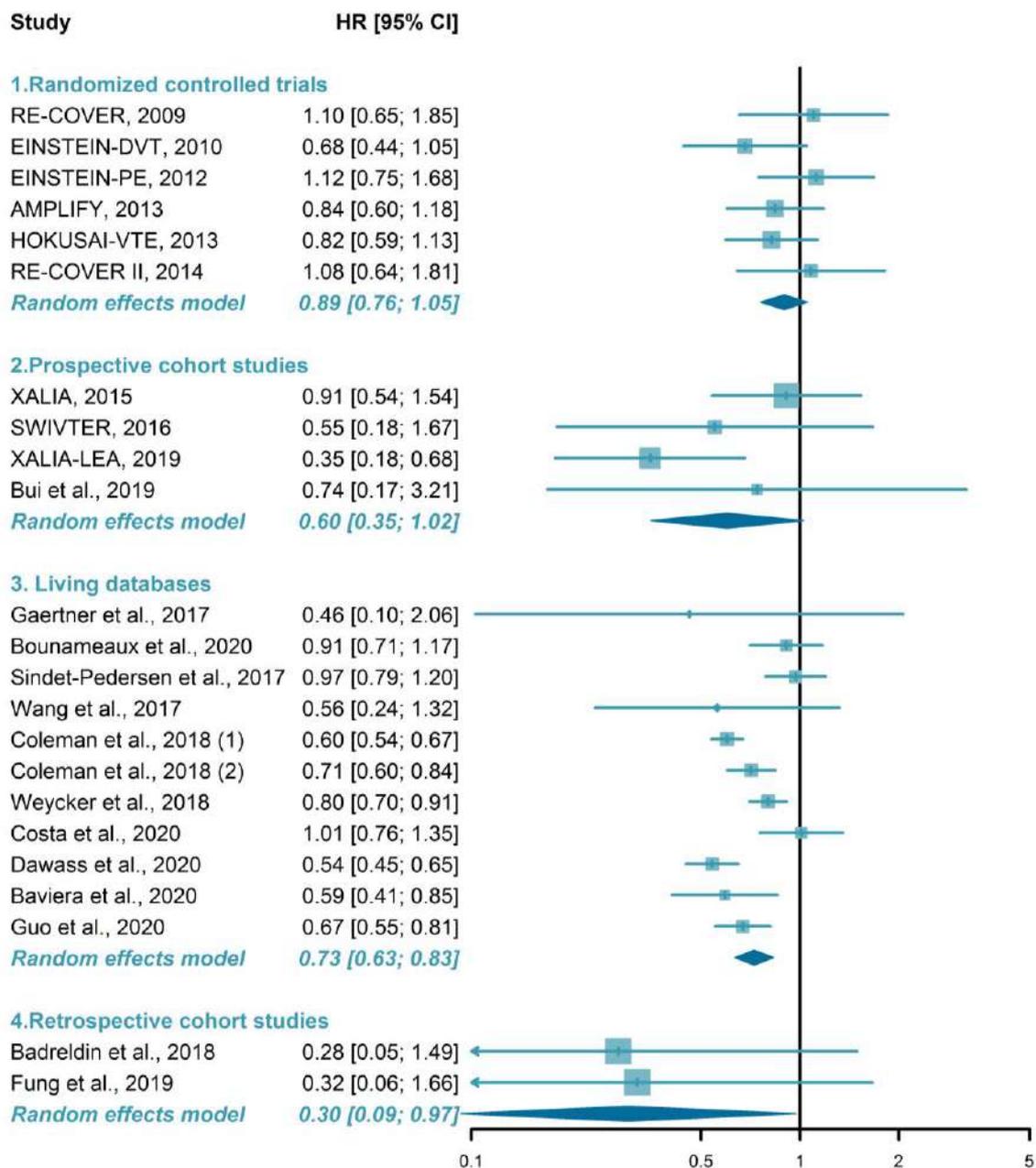
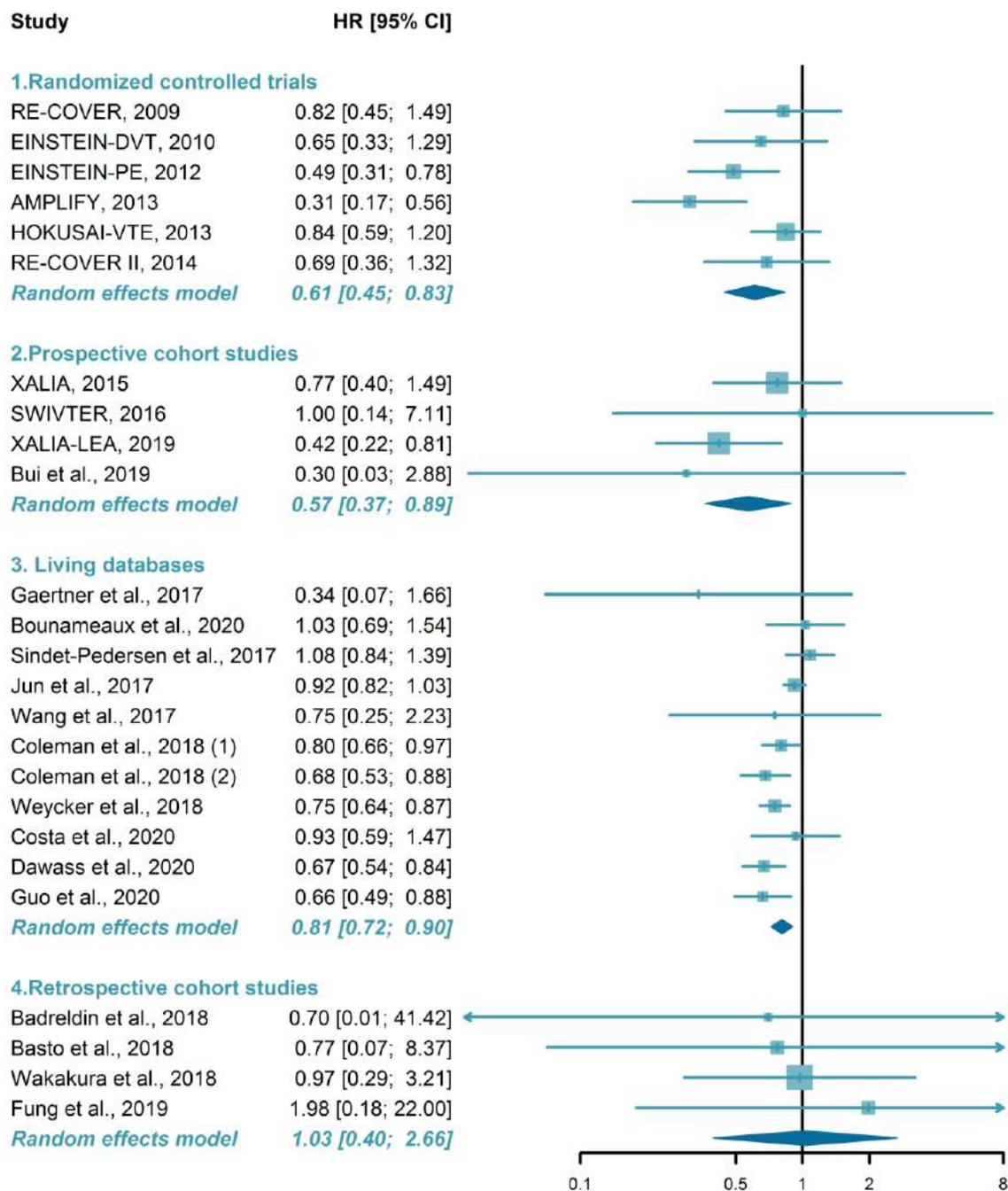


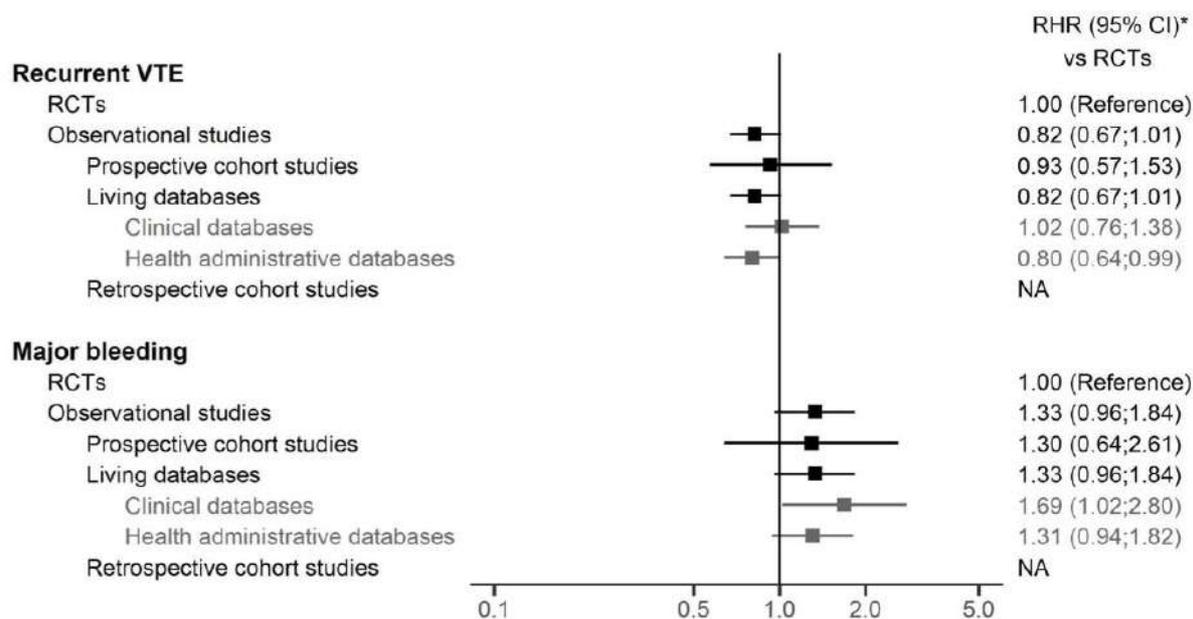
Figure 15: Comparaison des estimations des effets du traitement en fonction des types d'études sur les hémorragies majeures



Les faits marquants concernant la comparaison des études sur bases de données vivantes aux ECR sont les suivants :

- Dans les études sur bases de données vivantes, les AOD étaient associés à une réduction significative de 27% (HR= 0.73 [0.63 ; 0.83]) du risque de récurrence d'ETEV et de 19% (HR = 0.81 [0.72 ; 0.90]) du risque d'hémorragies majeures (Figures 14 et 15). Par rapport aux ECR, les études sur bases de données vivantes étaient associées à une surestimation non significative de 18 % de l'effet du traitement en termes d'efficacité (RHR = 0.82 [0.66 ; 1.01]), cette augmentation non significative était similaire après exclusion des études qui n'utilisaient aucune méthode d'ajustement des facteurs de confusion (RHR 0,82, [IC 95%, 0,67 à 1,01]). A l'inverse, l'estimation de l'effet des AOD était 33% plus faible (RHR = 1.33 [0.96 ; 1.84]) en termes de sécurité (Figure 16), ce résultat était similaire même après l'exclusion des études qui n'utilisaient aucune méthode d'ajustement des facteurs de confusion.

**Figure 16:** RHR et IC à 95% des RHR pour quantifier la différence de l'effet traitement entre les études observationnelles et les ECR



Lorsque nous comparons les résultats des études de cohortes rétrospectives aux ECR, nous disposons d'un faible nombre d'études.

- Les AOD étaient associés à une réduction significative de 70% du risque de récurrence d'ETEVE par rapport aux AVK (HR =0.30 [0.03 ; 0.97]), avec aucune différence du risque d'hémorragies majeures (HR =1.03 [0.40 ; 2.66]) (Figures 14 et 15).
- Par rapport aux ECR, les études de cohortes rétrospectives étaient associées à une surestimation non significative de l'effet du traitement en termes d'efficacité (RHR = 0.34 [0.10 ; 1.12]). A l'inverse, l'estimation de l'effet des AOD était 69% plus élevé dans les ECR en termes de sécurité, malgré que cette différence soit non significative (RHR = 1.69 [0.62 ; 4.57]) (Figure 16).
- Aucune des études de cohortes rétrospectives n'a utilisé de méthode d'ajustement des facteurs de confusion.

#### *Identification du biais de publication et de l'effet des petites études*

Les graphiques en entonnoir illustrant l'analyse des récurrences de ETEVE et des hémorragies majeures sont présentés respectivement dans les figures 13 et 14. L'asymétrie des graphiques présentés en fonction du type d'étude montre une surestimation apparente de l'effet du traitement en ce qui concerne les critères d'efficacité et de sécurité. Cependant, mais le test d'Egger n'était pas significatif.

### 3.4 Discussion des principaux résultats et limites

Cette étude méta-épidémiologique a comparé les estimations de l'effet du traitement des AOD par rapport aux AVK chez les patients avec MTEV entre les études observationnelles et les ECR. Il a été montré que les caractéristiques des patients avec MTEV inclus dans les ECR et dans les études observationnelles peuvent être différentes. Cette différence entre les caractéristiques peut modifier l'estimation des effets du traitement si une interaction existe entre caractéristiques des patients et effet du traitement. Ainsi elle peut entraîner des différences dans les estimations des effets du traitement entre les ECR et les études observationnelles sur les risques de la récurrence d'ETE ou d'hémorragies majeures.

Dans les études de cohortes prospectives, les caractéristiques des patients traités par AOD étaient similaires à celles incluses dans les ECR. Cela indique que l'inclusion des patients dans les études de cohortes prospectives n'était pas moins sélective que dans les ECR. Souvent, les mêmes centres de recherche clinique ont accepté de participer à la fois aux ECR, puis par la suite aux études de cohortes prospectives. Par contre, les caractéristiques initiales des patients sous AVK étaient différentes de celles des patients traités par AOD, traduisant ainsi le biais de sélection classique des études observationnelles. Ici, les patients sous AVK étaient plus âgés avec un pourcentage plus élevé de cancer et d'insuffisance rénale. L'ajustement avec un SP a permis de présenter une comparabilité apparente des caractéristiques des patients entre les deux groupes (AOD et AVK). Par conséquent, les populations incluses dans les études de cohortes prospectives et les ECR étaient apparemment plus proches. Malgré l'utilisation des méthodes d'ajustement dans ces études de cohortes prospectives, cette étude méta-épidémiologique a montré que ces cohortes prospectives étaient associées à une surestimation non significative de 33% l'effet du traitement en termes d'efficacité par rapport aux ECR. Cette surestimation de l'effet de traitement peut être expliquée par l'existence de biais de confusion résiduelle qui persiste dans ces études observationnelles même avec l'utilisation d'un SP.

En résumé, ces études de cohortes prospectives n'avaient pas inclus des patients moins sélectionnés que les ECR. Par conséquent, la mise en place de ces études de cohortes de phase IV spécifiques devrait être réalisée à distance des ECR afin d'être plus représentative de la population traitée "dans la vie réelle" et d'apporter des informations utiles et intéressantes, notamment les dérives de prescription dans le temps.

Les caractéristiques initiales des patients avec MTEV inclus dans les études sur bases de données de santé informatisées étaient différentes de celles des patients inclus dans les ECR avec une différence d'âge médian de 10 ans. Cette constatation montre que les patients sont moins sélectionnés dans la "vraie vie", puisqu'ils ne sont pas sélectionnés selon des critères d'inclusion et d'exclusion stricts comme dans les ECR. Des études méta-épidémiologiques précédentes ayant comparé les ECR et les études observationnelles ont également constaté des différences dans les caractéristiques initiales des patients inclus. Ces études ont montré également que les patients inclus dans les ECR étaient plus jeunes avec une durée de maladie plus courte que celle dans les études observationnelles<sup>(169)</sup>. Plusieurs méta-analyses ont comparé les estimations de l'effet du traitement entre ECR et études observationnelles utilisant un SP dans différents contextes cliniques. Les estimations de ces dernières études étaient en accord avec celles des ECR<sup>(170-172)</sup>.

Cependant, les résultats présentés dans ce travail ont montré une surestimation non significative de l'effet du traitement en termes d'efficacité dans les études observationnelles par rapport aux ECR. Néanmoins, les résultats en termes de sécurité étaient différents puisque les études sur bases de données de santé informatisées étaient associées à une diminution, certes non significative, de 33 % du bénéfice des AOD sur le risque des hémorragies majeures (RHR = 1.33 [0.96 ; 1.84]). Le constat est inversé en termes de tendance pour l'efficacité. Plusieurs explications peuvent être évoquées. Tout d'abord, il peut exister une interaction avec les caractéristiques des patients et l'effet des AOD : le bénéfice des AOD serait moindre en termes de sécurité si les patients étaient moins sélectionnés. Deuxièmement, il serait lié à un biais de sélection : les patients sous AOD seraient plus à risque car le médicament semblant prometteur, il a été préféré dans la population dite fragile. Mais aucun argument ne permet de l'affirmer, ni même la tendance inverse sur l'efficacité, et ce résultat est observé même après ajustement sur le SP. Certes, l'ajustement à un SP est insuffisant pour contrôler un biais de confusion car il dépend de la disponibilité des variables dans la base de données utilisée. Le calcul d'un SP fiable et complet est difficile vue l'indisponibilité de données pour des variables spécifiques de ces études. Enfin, l'absence de double aveugle, ou d'évaluation à l'aveugle peut aussi expliquer ce résultat. A notre connaissance, il n'existe pas d'études méta-épidémiologiques ayant évalué les différences d'estimation sur bases de données de santé informatisées et les ECR.

Les études utilisant des bases de données de santé informatisées semblent inclure des patients plus proches que les patients traités dans la vraie vie avec des résultats moins optimistes que

les ECR concernant la sécurité. Néanmoins, les facteurs de confusion inhérents dans ces études seront toujours présents, et il est impossible de savoir s'il s'agit d'une sous-estimation de l'effet traitement ou d'une interaction avec celui-ci. Les données observées dans la FA peuvent peut-être permettre d'avancer sur cette question.

Et enfin, les patients dans étude de cohortes rétrospectifs étaient moins sélectionnés et plus proches des patients traités dans la vraie vie. Néanmoins, il n'y avait qu'un petit nombre d'études ayant un recrutement rétrospectif et non exhaustif. Ces études ont donc un faible niveau de preuves pour être considérées comme des études pharmaco-épidémiologiques.

# 4 Quantification du biais d'évaluation dans les études évaluant les anticoagulants oraux dans la fibrillation auriculaire

## 4.1 Recherche exhaustive et sélection des études

Une revue systématique de la littérature a été réalisée afin d'identifier les études randomisées et les études observationnelles évaluant les AOD versus AVK chez les patients avec une FA non valvulaire. Des critères d'éligibilité ont été définis afin de déterminer toutes les études éligibles à l'inclusion.

### 4.1.1 Critères d'éligibilité des études

Par rapport à la question posée, les critères d'éligibilité étaient les suivants :

- Les essais randomisés et les études observationnelles devaient avoir évalué l'effet des AOD par rapport à celui des AVK (en traitement initial d'une FA non valvulaire) sur le risque de survenue des TEA (AVC ou embolie systémique) et des hémorragies majeures.
- L'objectif des études devait être l'évaluation de l'efficacité et de la sécurité des anticoagulants.
- Les comparaisons ont été limitées aux effets des AOD approuvés (ou en cours d'examen réglementaire) sur la prévention des TEA chez les patients avec FA non valvulaire, notamment le dabigatran, le rivaroxaban, l'apixaban et l'edoxaban par rapport à la warfarine orale ou à un autre AVK.
- Les études ne devaient pas inclure des populations spécifiques de FA. Les populations spécifiques sont :
  - Les patients souffrant à la fois d'une FA et d'une autre maladie (comme le cancer, l'insuffisance rénale, un infarctus aigu du myocarde) ou patients atteints de FA sous dialyse.
  - Les patients avec FA subissant une ablation de la FA (par cathéter ou chirurgie), une cardioversion électrique, une chirurgie cardiaque ou une intervention coronarienne percutanée (ICP).

#### 4.1.2 Revue de la littérature

La recherche exhaustive des essais randomisés et des études observationnelles couvre la période allant du 1er janvier 2009 au 14 mars 2019. Elle a été réalisée dans les bases de données électroniques PubMed, Cochrane Database et Google scholar (la première étude sur les AOD a été publiée en 2009). Une équation de recherche a été construite pour identifier les articles sur PubMed (cf. annexe 2- tableau A7), avec les mots clés suivants : " atrial fibrillation, anticoagulants, rivaroxaban, dabigatran, warfarin, apixaban, edoxaban, thromboembolic events, stroke, hemorrhage, myocardial ischemia, gastrointestinal hemorrhage". Aucune restriction sur la langue de publication n'a été utilisée.

La recherche systématique a permis de constater que certaines études étaient conduites sur les mêmes bases de données de santé informatisées avec des périodes d'inclusion identiques. Ces études pouvant être considérées comme des doublons, nous avons rajouté des critères de sélection : ont été considérées comme des doublons les études dont plus de la moitié de la période de recrutement se chevauchait avec celle d'une autre étude ayant un effectif plus grand et ayant évalué les mêmes groupes de traitement chez les mêmes patients. Lorsque des doublons ont été identifiés, l'étude avec l'effectif le plus important a été incluse dans la méta-analyse.

En termes de conséquences de la gestion des études éligibles à l'inclusion :

- Une étude ayant inclus des patients naïfs n'est pas considérée comme le doublon d'une étude réalisée sur la même base de données et sur la même période si cette étude concernait des patients switchers ;
- Une étude ayant évalué des doses réduites n'est pas considérée comme le doublon d'une étude réalisée sur la même base de données et sur la même période si cette étude concernait des pleines doses de la molécule ;
- Une étude peut être considérée comme le doublon d'une autre étude pour la comparaison A vs AVK, mais pas forcément pour la comparaison B vs AVK si cette comparaison n'était pas disponible dans l'étude la plus importante.

Un comité de pilotage de la méta-analyse a été constitué afin d'examiner les résultats de la recherche bibliographique, vérifier ensuite les études sélectionnées et évaluer la qualité méthodologique des études éligibles. Ce comité anime l'intégralité du processus d'évaluation. Cette recherche exhaustive a permis d'identifier les études randomisées et les études de cohortes

ayant étudié les effets des AOD versus AVK, chez des patients avec une FA non valvulaire, sur le risque de TEA, les hémorragies majeures, les hémorragies intracrâniennes et les hémorragies gastro-intestinales.

#### 4.1.3 Analyses des études éligibles

Afin d'évaluer la qualité méthodologique de chaque étude par le risque de biais, le comité de pilotage de la méta-analyse a évalué la qualité des études pour décider de son exclusion ou de son maintien dans la méta-analyse, en prévoyant dans ce dernier cas des analyses de sensibilité sans cette étude. Des analyses de sensibilité ont été réalisées sur deux critères (survenue de TEA et d'hémorragies majeures) en excluant les études observationnelles n'ayant pas appliqué de méthodes d'ajustement.

#### 4.1.4 Critères de jugement

Le critère de jugement principal en termes d'efficacité était la survenue des événements TEA. Il était le critère principal de tous les essais randomisés évaluant les AOD dans cette indication. Il s'agit d'un critère composite incluant les AVC ou les embolies systémiques.

Le critère de jugement principal en termes de sécurité était la fréquence des hémorragies majeures. Ce critère est également un critère composite défini selon la définition de ISTH<sup>(76)</sup>.

Nous avons également retenu en critère de jugement secondaire pour la sécurité les hémorragies gastro-intestinales et les hémorragies intracrâniennes. Les hémorragies intracrâniennes regroupaient les AVC hémorragique et les hémorragies sous-durales ou sous-arachnoïdiennes.

#### 4.1.5 Extraction des données

Après avoir identifié les études éligibles (revue de la littérature), les données ont été collectées indépendamment par deux personnes. Pour les études randomisées et observationnelles, nous avons recueilli :

- Les caractéristiques des études : la date de publication, les noms des auteurs, le pays où l'étude a été réalisée, le nombre de patients (total et dans chaque groupe), la conception de l'étude et la période d'inclusion.
- Les caractéristiques des traitements : le type d'AOD, la dose reçue et la durée de traitement.

La durée de traitement par anticoagulant dans ces études était de 3 à 12 mois.

- Les catégories de patients par rapport à leur exposition au traitement :
  - Patient naïf au traitement : se réfère à un patient qui n'a encore jamais pris de traitement
  - Patient switch (switcher) : c'est le patient qui change son traitement contre un autre traitement ayant la même efficacité (switcher = basculer sur un autre traitement).
  - Patient expérimenté (experienced) : patient ayant déjà été exposé à un traitement.
- Les caractéristiques des patients : l'âge moyen, le sexe, le score CHADS2 (ce score permet d'évaluer le risque d'événements thrombo-emboliques de patients ayant une FA), l'hypertension, les antécédents d'AVC, le diabète et l'insuffisance cardiaque congestive.
- Les estimations des effets du traitement (AOD vs AVK), résumées par le HR ou le RR pour les événements TEA et les hémorragies (majeures, gastro-intestinales et intracrâniennes).

De plus, les études observationnelles ont été classées en trois catégories : 1/ les études de cohortes prospectives, 2/ les études de cohortes rétrospectives et 3/ les études sur bases de données de santé informatisées ou des registres de soins de santé avec un recueil prospectif d'informations sur les patients. De plus, nous avons recueilli :

- Les méthodes d'appariement ou d'ajustement utilisées (aucune, score de propension SP ou modèle multivarié).
- Les caractéristiques des patients avant et après SP si applicable.
- Les HR et RR après appariement ou ajustement.

## 4.2 Description des études et des populations incluses

Les caractéristiques des populations des ECR et des études observationnelles ont été comparées.

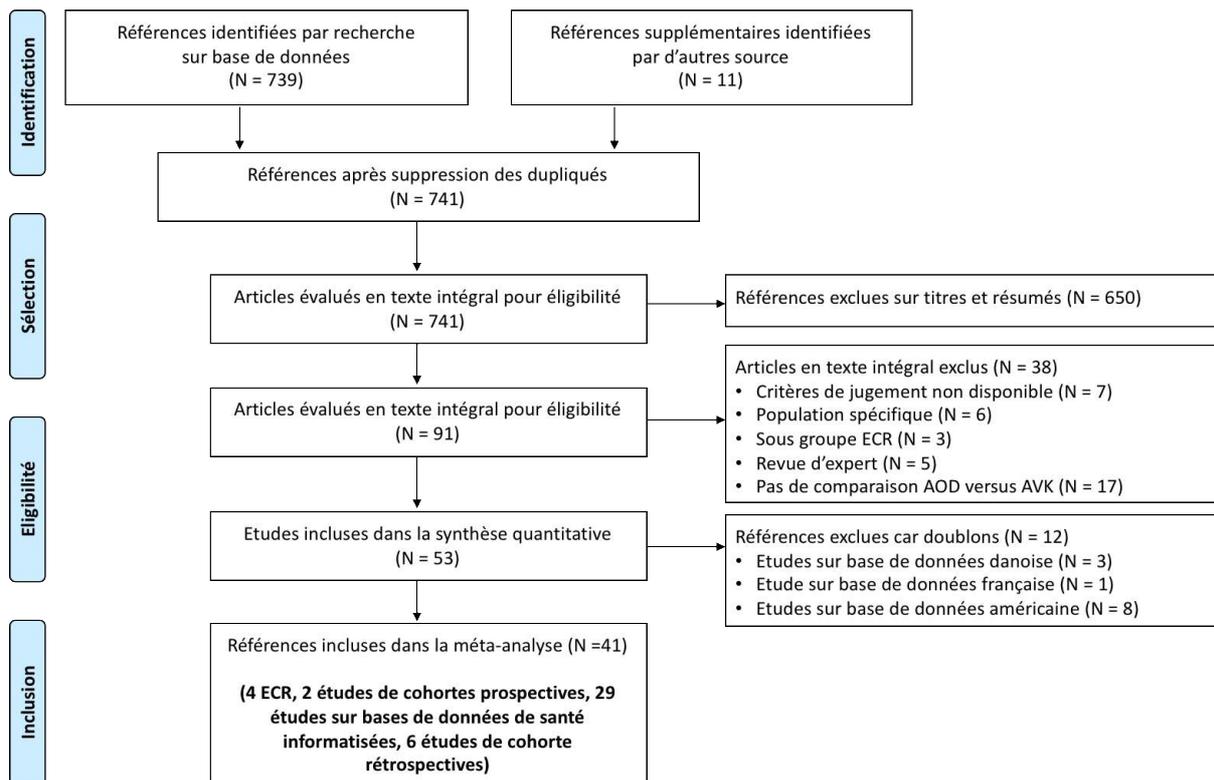
### 4.2.1 Identification et sélection des études

741 références d'études éligibles entre 2009 et 2019 ont été identifiées. Après vérification des titres et des résumés, 650 références ont été exclues. Après évaluation des textes en intégrité, 38 autres références ont été exclues. Les raisons des exclusions sont fournies dans la figure 18. Au total, 53 études ont été considérées comme éligibles dans la méta-analyse : 4 ECR et 49 études observationnelles.

Plusieurs méta-analyses ont déjà été conduites avec les ECR. L'identification et la sélection de celles-ci n'ont pas posé de problème spécifique.

Parmi les études observationnelles, une attention particulière a été portée à l'analyse des doublons.

Figure 17: Diagramme de flux du processus de sélection des études



#### 4.2.2 Gestion des doublons

Le nombre d'études observationnelles éligibles à l'inclusion était plus important dans la FA que dans la MTEV. La majorité de ces études observationnelles a été réalisée sur des bases de données de santé informatisées. La gestion de ces études conduites sur ces bases de données a nécessité un travail de recueil/analyse plus important que dans la MTEV, notamment du fait de l'analyse des doublons.

*A/ Les études sur bases de données de santé américaines*

Dix-neuf études évaluant les AOD chez les patients avec FA ont été conduites sur les bases de données de santé américaines. Ces études sont résumées dans le tableau 4.

**Tableau 4:** Liste des études conduites sur des bases de données américaines

<b>Base de données</b>	<b>Références</b>	<b>Patients</b>	<b>Période</b>	<b>AOD</b>
Optum labs data warehouse and Medicare Advantage	Yao et al 2016 <sup>(220)</sup>	76 354 patients naïfs et expérimentés	octobre 2010 à juin 2015	dabigatran 150 mg rivaroxaban 20 mg apixaban 5 mg
	Abraham et al 2015 <sup>(174)</sup>	7 749 patients naïfs	novembre 2010 à septembre 2013	dabigatran
Medicare/Medicaid	Graham et al 2015 <sup>(175)</sup>	134 414 patients naïfs	octobre 2010 à décembre 2012	dabigatran (75, 150) mg
	Hernandez et al 2015 <sup>(176)</sup>	9 404 patients naïfs	octobre 2010 à octobre 2011	dabigatran (75, 150) mg
Market scan	Bengtson et al 2016 <sup>(177)</sup>	134 085 patients 11 581 patients	octobre 2010 à décembre 2012	dabigatran (75, 150) mg rivaroxaban (10, 15, 20) mg
	Lauffenburger et al 2015 <sup>(178)</sup>	64 935 patients	octobre 2010 à décembre 2012	dabigatran (75, 150) mg
	Norby et al 2017 <sup>(179)</sup>	77 991 patients	janvier 2010 à décembre 2014	rivaroxaban (10, 15, 20) mg
Medicare/Medicaid, Market Scan, PharMetrics, Optum et Humana	Li et al 2017 <sup>(180)</sup>	76 940 patients	janvier 2012 à septembre 2015	apixaban 5 mg
	Lip et al 2016 <sup>(181)</sup>	29 338 patients	janvier 2013 à décembre 2014	dabigatran 150 mg rivaroxaban 20 mg apixaban 5 mg
	Coleman et al 2016 <sup>(173)</sup>	22 822 patients naïfs	janvier 2012 à octobre 2014	rivaroxaban 20 mg apixaban 5 mg
	Amin et al 2017 <sup>(182)</sup>	186 132 patients naïfs	janvier 2012 à décembre 2014	dabigatran 150 mg rivaroxaban 20 mg apixaban 5 mg
	Seeger et al 2015 <sup>(183)</sup>	59 663 patients naïfs	octobre 2010 à décembre 2012	dabigatran 150 mg

Base de données	Références	Patients	Période	AOD
Medicare/ Medicaid, Market Scan, PharMetrics, Optum et Humana	Lip et al 2018 <sup>(184)</sup>	321 182 patients naïfs	janvier 2013 à septembre 2015	dabigatran (75, 150) mg rivaroxaban, (10, 20) mg apixaban (2.5, 5) mg
	Coleman et al 2017 <sup>(185)</sup>	9 684 patients naïfs	janvier 2012 à juin 2015	dabigatran 150 mg rivaroxaban 20 mg apixaban 5 mg
US Department of Defense claims	Villiness et al 2015 <sup>(190)</sup>	19 853 patients naïfs	octobre 2009 à juillet 2013	dabigatran 150, 75mg
Veterans administrative encounter and pharmacy	Sarrazin et al 2014 <sup>(189)</sup>	85 344 patients switchers	2010 à 2012	dabigatran 150 mg
HealthCore Integrated Research Environment	Adeboyeje et al, 2017 <sup>(192)</sup>	44 075 patients naïfs	novembre 2010 à février 2016	dabigatran 150 mg rivaroxaban 15 mg, apixaban 2.5 mg
Sentinel distributed database	Go et al 2017 <sup>(191)</sup>	50 578 patients naïfs	novembre 2010 à mai 2014	dabigatran
Symphony Health Solutions Datasets	Laliberté et al 2014 <sup>(193)</sup>	18 270 patients naïfs	mai 2011 à juillet 2012	rivaroxaban

- Deux études ont été réalisées sur la même base de données «Optum labs data warehouse and Medicare Advantage» (Yao et al, Abraham et al 2015). Ces deux études ont été considérées comme doublons car toute la période de recrutement de l'étude Abraham et al se chevauchait avec l'étude Yao et al. Cette dernière étude avait un effectif plus grand et avait évalué les mêmes groupes de traitement chez les mêmes patients. L'étude Abraham et al a été considérée comme un sous-groupe de la publication Yao et al et a donc été exclue.
- Deux études ont été conduites sur la même base de données américaine «Medicare/Medicaid». Ces deux études ont été considérées comme doublons puisque toute la période

de recrutement de l'étude Hernandez et al se chevauchait avec l'étude Graham et al ayant un effectif plus important. Les deux études avaient évalué le dabigatran chez les mêmes patients naïfs. Ainsi, l'étude Hernandez et al a été exclue puisqu'elle était considérée comme un sous-groupe de l'étude Graham et al.

- Trois études ont été conduites sur la même base (Market scan) et ont évalué le dabigatran chez des patients naïfs et switchers. Pour l'évaluation du dabigatran, les deux études Bengtson et al et Lauffenburger et al ont été considérées comme doublons car elles ont évalué le dabigatran chez les mêmes patients et toute la période de l'étude Lauffenburger et al se chevauchait avec l'étude Bengtson et al qui avait un effectif plus grand. L'étude Lauffenburger et al a donc été exclue de l'analyse. Les patients de l'étude Bengtson et al (dabigatran) ont été inclus dans l'analyse. De même pour l'évaluation du rivaroxaban, les études Bengtson et al et Norby et al ont été considérées comme doublons puisque toute la période de recrutement de l'étude Bengtson et al se chevauchait avec l'étude Norby et al ayant un plus grand nombre de patients sous rivaroxaban. Ainsi, les patients sous rivaroxaban de l'étude Bengtson et al ont été exclus de l'analyse. L'étude Norby et al a été retenue dans l'analyse.
- Sept études ont été réalisées sur l'ensemble ou une partie de ces bases de données américaines : Medicare/Medicaid, Market Scan, PharMetrics, Optum et Humana (défini et détaillé dans l'annexe1-Figure A1). Cinq études (Coleman et al 2017, Li et al 2016, Lip et al 2016, Coleman et al 2016 et Amin et al 2017) ont été considérées comme doublons de l'étude Lip et al de 2018. En effet, ces cinq études ont évalué les mêmes patients avec un effectif plus faible et une période de recrutement se chevauchant sur plus de la moitié avec l'étude Lip et al de 2018. Ces cinq études ont donc été exclues. L'étude Lip et al de 2018 ayant l'effectif le plus important et réalisée sur un regroupement des grandes bases de données a été retenue. Enfin, l'étude Seeger et al a été incluse dans l'analyse puisque sa période de recrutement ne se chevauchait pas avec d'autres études conduites sur les mêmes bases (Market scan et Optum).
- Cinq autres études ont été réalisées sur des bases de données différentes des précédentes (Villiness et al 2015, Sarrazin et al 2014, Laliberté et al 2014, Adeboyeje et al 2017, Go et al 2017)<sup>(186-190)</sup>. Ces cinq études ont été incluses dans l'analyse.

En résumé, 8 études ont été considérées comme des doublons sur les bases américaines et ont donc été exclues de l'analyse. Au total, 11 études utilisant des bases de données américaines ont été incluses dans l'analyse.

*B/ Les études sur bases de données de santé danoises*

**Tableau 5:** Liste des études conduites sur la base danoise

Références	Patients	Période	AOD
<b>Nielsen et al 2017<sup>(191)</sup></b>	55 644 patients naïfs	août 2011 à fév 2016 fév 2012 à fév 2016 déc 2012 à fév 2016	dabigatran 110 mg rivaroxaban 15 mg apixaban 2.5 mg
<b>Larsen et al 2016<sup>(192)</sup></b>	61 678 patients naïfs	août 2011 à nov 2015 fév 2012 à nov 2015 déc 2012 à nov 2015	dabigatran 110 mg rivaroxaban 15 mg apixaban 2.5 mg
<b>Gorst-Rasmussen et al 2016<sup>(193)</sup></b>	22 358 patients naïfs	oct 2010 à déc 2012 nov 2011 à déc 2012	dabigatran 110 et 150 mg rivaroxaban 15 et 20 mg
<b>Larsen et al 2013<sup>(194)</sup></b>	14 267 patients	août 2011 à déc 2012	dabigatran 110 et 150 mg
<b>Sorensen et al 2013<sup>(195)</sup></b>	52 366 patients	août 2011 à déc 2011	dabigatran 110 et 150 mg
<b>Staerk et al 2015<sup>(196)</sup></b>	11 045 patients	août 2011 à déc 2012	dabigatran 110 et 150 mg
<b>Larsen et al 2014 A (Bleeding events)<sup>(197)</sup></b>	33 945 patients naïfs et switchers	août 2011 à mai 2013	dabigatran 110 et 150 mg
<b>Larsen et al 2014 B (stroke)<sup>(198)</sup></b>	6 141 patients naïfs et expérimentés	août 2011 à mai 2013	dabigatran 110 et 150 mg

Huit études ont été réalisées sur la base de données danoise « Danish national prescription registry ». Ces études sont résumées Tableau 5.

Quatre études (Gorst-Rasmussen et al 2016, Larsen et al 2013, Sorensen et al 2013, Staerk et al 2015) ont été considérées comme doublons des deux premières (Nielsen et al 2017, Larsen et al 2016). En effet, ces quatre études ont évalué les mêmes patients sous dabigatran avec un effectif plus faible et une période de recrutement plus courte se chevauchant sur plus de la moitié avec les deux premières études<sup>(195,196)</sup>. Les patients sous dabigatran de l'étude Gorst-

Rasmussen et al 2016 ont donc été exclus de l'analyse ainsi que les 3 autres doublons. Les études Nielsen et al 2017 et Larsen et al 2016 ainsi que les patients sous rivaroxaban de l'étude Gorst-Rasmussen et al 2016 ont été inclus dans l'analyse. Deux autres études ont évalué les différentes catégories de patients : naïfs/switchers (Larsen et al, 2014 A) et naïfs/expérimentés (Larsen et al, 2014 B). Certes, ces deux études se chevauchaient avec d'autres études sur la période de recrutement, mais elles n'ont pas analysé les mêmes patients. Nous avons donc fait le choix avec le comité de pilotage d'inclure ces deux études dans l'analyse.

### *C/ Les études sur bases de données de santé françaises*

Trois études ont été conduites sur le même regroupement de bases de données françaises (SNIIRAM et PMSI) :

- Nacora et al 2014 incluant 71 589 patients naïfs sous dabigatran et rivaroxaban entre juin 2012 et novembre 2012<sup>(199)</sup>
- Maura et al 2015 incluant 38 409 patients naïfs sous dabigatran et rivaroxaban entre juillet 2012 et novembre 2012<sup>(200)</sup>
- Bouillon et al 2015 incluant 17410 patients naïfs et switchers sous dabigatran et rivaroxaban entre janvier 2011 et novembre 2012<sup>(201)</sup>

Deux études ont été considérées comme doublons car toute la période de recrutement de l'étude Maura et al se chevauchait avec l'étude Nacora et al. Cette dernière étude avait un effectif plus grand et avait évalué les mêmes groupes de traitement (dabigatran et rivaroxaban) chez les mêmes patients. Ainsi, l'étude Maura et al a été exclue de l'analyse et l'étude Nacora et al a été retenue. L'étude Bouillon et al a également été retenue puisqu'elle incluait différentes catégories de patients, contrairement aux deux autres études conduites sur la même base.

Au total, 41 études ont été incluses dans l'analyse globale (Figure 17). Quatre ECR incluant un total de 27 096 patients<sup>(86,89,202,203)</sup>. Chacun de ces essais a évalué un des 4 AOD : le dabigatran<sup>(86)</sup>, le rivaroxaban<sup>(202)</sup>, l'apixaban<sup>(88)</sup> ou l'edoxaban<sup>(89)</sup>.

- Deux études de cohortes prospectives incluant 5 172 patients : une étude a évalué le dabigatran<sup>(204)</sup> et une autre étude a évalué les 3 autres AOD (apixaban, rivaroxaban et edoxaban)<sup>(205)</sup>. Parmi ces deux études, une seule étude a utilisé une méthode de correction de biais : le modèle de régression multiple.

- Six études de cohortes rétrospectives incluant 8 647 patients : 4 études ont évalué le dabigatran<sup>(206–209)</sup>, 1 étude a évalué le rivaroxaban<sup>(210)</sup> et une étude a évalué 3 AOD (apixaban, dabigatran et rivaroxaban)<sup>(211)</sup>. Trois études ont utilisé l'appariement sur SP pour minimiser les biais et les 3 autres études n'ont pas utilisé de méthodes de correction.
- Vingt-neuf études utilisant des bases de données de santé informatisées incluant 1 763 551 patients : 4 études ont évalué le rivaroxaban<sup>(190,193,212,213)</sup>, 11 études ont évalué le dabigatran<sup>(183,186–188,206,212,214–218)</sup>, 4 études ont évalué 2 AOD (dabigatran et rivaroxaban)<sup>(177,219)</sup> et 10 autres études ont évalué 3 AOD (dabigatran, rovaroxaban et apixaban)<sup>(181,189,191,192,199,201,220–225)</sup>. Toutes ces études ont utilisé des méthodes d'ajustement : 22 études ont utilisé un score de propension (14 appariement sur SP, 6 pondération inverse sur SP et 2 ajustement sur SP) et 7 études ont appliqué un modèle de régression multiple.

Les caractéristiques des études incluses sont décrites dans l'annexe2-tableau A6.

#### 4.2.3 Comparaison des populations ECR et des études observationnelles

Des analyses statistiques descriptives ont été effectuées afin de comparer les populations incluses dans les ECR et dans les études observationnelles évaluant les AOD chez les patients avec FA non valvulaire. Les caractéristiques des patients inclus dans ces études ont été comparées avant et après l'application du SP. Pour chaque type d'étude, les variables catégorielles extraites sous forme de pourcentages ainsi que les variables continues extraites sous formes de moyennes ont été présentées par une médiane non pondérée.

Les caractéristiques des patients traités par AOD dans les études observationnelles et dans les ECR sont très différentes. A titre d'exemple, les patients sous AOD dans les études de cohortes prospectives sont grosso modo plus jeunes, plus souvent des femmes, moins hypertendus, avec une fréquence d'insuffisance cardiaque moindre et un score CHADS2 plus faible, donc plutôt un profil moins à risque de complications. Pour les études de cohorte rétrospective, la comparaison est moins évidente : l'âge est comparable, mais les patients sont plus souvent des femmes, moins hypertendus, avec une fréquence d'insuffisance cardiaque moindre par rapport aux ECR et pour autant plus souvent diabétiques et avec des antécédents d'AVC plus importants (tableau 6). La comparaison des caractéristiques des patients sous AOD dans les études sur bases de données de santé informatisées est encore différente : les patients sont plus âgés que dans les ECR, mais la fréquence des comorbidités est plus faible que dans les ECR, ce qui ne permet pas de déterminer les populations les plus à risque entre ECR et études

observationnelles. Ainsi, la comparaison des estimations d'efficacité et de sécurité des AOD entre les ECR et les études de cohortes rétrospectives ou sur bases de données de santé informatisées devrait principalement quantifier un biais d'estimation dû à l'absence de randomisation dans les études observationnelles.

Après appariement ou ajustement, les caractéristiques des populations semblent apparemment comparables entre ECR et études sur bases de données de santé informatisées ou études de cohortes rétrospectives (tableau 7).

Tableau 6: Caractéristiques des patients selon le type d'étude avant ajustement ou appariement

	Essais contrôlés randomisés		Études de cohortes prospectives		Études sur bases de données de santé informatisées		Études de cohortes rétrospectives	
<b>Nombre des études</b>	4		2		29		6	
<b>Nombre total des patients</b>	27 096		5 172		1 763 551		8 647	
<b>Méthode de correction du biais de confusion</b>	/		1 modèle de régression multiple 1 pas de correction		14 appariement sur SP 6 pondération inverse sur SP  2 ajustement sur PS 7 modèles de régression multiple		3 appariement sur SP 3 pas de correction	
<b>Groupe de traitement</b>	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK
<b>Age (années)</b>	72 (71-72)	72 (71-72)	67 (67-71)	69 (69-70)	75 (72-77)	73 (71-75)	70 (65-71)	69 (68-71)
<b>Sexe masculin (%)</b>	63 (61-64)	63 (62-64)	41 (38-57)	54 (45-62)	56 (53-61)	56 (56-59)	56 (52-60)	55 (51-59)
<b>Hypertension (%)</b>	89 (81-93)	89 (85-92)	75 (67-78)	62 (61-62)	60 (49-67)	62 (55-75)	69 (68-70)	69 (62-77)
<b>Diabète</b>	24 (23-29)	25 (24-32)	24 (21-27)	26 (22-30)	18 (16-26)	22 (18-30)	30 (29-30)	33 (30-37)
<b>Score CHADS2*</b>	2,5 (2,1-2,8)	2,5 (2,1-3)	2 (1,7-2)	2,3 (2-2,7)	2 (1,8-2,3)	2,1 (2-2,3)	2,5°	2,3°
<b>Atcd d'AVC (%)</b>	24 (20-28)	24 (20-35)	21 (17-23)	18 (16-19)	20 (16-23)	19 (14-21)	35 (11-39)	11 (9-16)
<b>Insuffisance cardiaque congestive (ICC)</b>	46 (33-58)	46 (35-59)	22 (21-25)	26 (24-28)	16 (11-20)	20 (14-24)	25 (20-28)	25 (24-27)

°une seule étude , Atcd: antécédent

**Tableau 7:** Caractéristiques des patients selon le type d'étude après ajustement ou appariement (si disponible)

	<b>Essais contrôlés randomisés</b>		<b>Études de cohortes prospectives**</b>		<b>Études sur bases de données de santé informatisées</b>		<b>Études de cohortes rétrospectives</b>	
<b>Nombre des études</b>	4		2		19		2	
<b>Nombre total des patients</b>	27 096		5 172		1 239 838		1 710	
<b>Groupe de traitement</b>	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK
<b>Age (années)</b>	72 (71-72)	72 (71-72)	67 (67-71)	69 (69-70)	72 (70-75)	74 (71-75)	72 (72-73)	73 (72-73)
<b>Sexe masculin (%)</b>	63 (61-64)	63 (62-64)	41 (38-57)	54 (45-62)	59 (53-62)	59 (55-61)	58 (56-59)	57 (54-60)
<b>Hypertension (%)</b>	89 (81-93)	89 (85-92)	75 (67-78)	62 (61-62)	70 (52-85)	73 (54-84)	70 (67-73)	68 (67-70)
<b>Diabète</b>	24 (23-29)	25 (24-32)	24 (21-27)	26 (22-30)	29 (18-34)	30 (23-35)	21 (19-24)	22 (20-23)
<b>Score CHADS2</b>	2,5(2,1-2,8)	2,5 (2,1-3)	2 (1,7-2)	2,3 (2-2,7)	2 (1,9-2,4)	2,2 (1,9-2,4)	1,8 (1,7-1,8)	1,8 (1,7-1,8)
<b>Atcd d'AVC (%)</b>	24 (20-28)	24 (20-35)	21 (17-23)	18 (16-19)	15 (11-27)	16 (12-27)	12 (11-13)	13 (12-14)
<b>Insuffisance cardiaque congestive (ICC)</b>	46 (33-58)	46 (35-59)	22 (21-25)	26 (24-28)	28 (18-31)	28 (18-37)	15 (15-16)	16 (16-16)

°une seule étude, Atcd: antécédent.

\*\* Caractéristiques des patients présentées dans ce tableau sont avant ajustement pour les études de cohortes prospectives.

### 4.3 Approche méta-analytique : comparaison des estimations des effets des traitements entre les études randomisées et les études observationnelles

#### 4.3.1 Méthodes

Des méthodes méta-analytiques ont été utilisées afin de comparer l'estimation de l'effet traitement des AOD par rapport aux AVK entre les ECR et les études observationnelles (études de cohortes prospectives, études utilisant des bases de données de santé informatisées ou études de cohortes rétrospectives) pour la prévention des TEA chez les patients avec FA non valvulaire. Pour chaque type d'étude, un HR commun avec son intervalle de confiance (IC) à 95% a été calculé pour le risque d'AVC ou d'embolie systémique, d'hémorragies majeures, d'hémorragies intracrâniennes et d'hémorragies gastro-intestinales. Les méta-analyses ont été réalisées avec la méthode du logarithme des HR en pondérant les études par l'inverse de la variance. Lorsque la valeur du HR n'était pas disponible dans la publication, celle-ci a été approximée par le RR.

La méta-analyse était basée sur un modèle à effets aléatoires<sup>(161,162)</sup>. Pour chaque méta-analyse :

- L'hétérogénéité entre les études a été évaluée à l'aide du test statistique Q de Cochrane.
- La cohérence entre les études a été mesurée à l'aide de l'indicateur  $I^2$ <sup>(160)</sup>.

Les résultats obtenus ont été considérés statistiquement significatifs si la valeur de  $P$  était  $<0.05$ . Les résultats de la méta-analyse sont représentés par un « Forest plot ». Un HR égal à 1 indique l'absence de différence entre les deux traitements (AOD et AVK), HR  $<1$  indique que les AOD sont plus efficaces que les AVK, HR  $>1$  indique que les AVK sont plus efficaces que les AOD.

Des graphiques en entonnoir de Begg ont été utilisés afin de détecter les biais potentiels des études incluses dans la méta-analyse par la présence d'une asymétrie des graphiques<sup>(163)</sup>. En outre, pour quantifier la différence de l'effet du traitement estimé entre les études observationnelles et les ECR, le ratio des HR (RHR) a été calculé en divisant les valeurs des HR obtenues dans les études observationnelles par celles obtenues dans les ECR. Le RHR correspond à l'association entre le type d'étude et la taille de l'effet du traitement. Un RHR  $<1$  signifie que les estimations de l'effet du traitement sont plus élevées dans les études observationnelles que dans les ECR, un RHR  $>1$  signifie que les estimations de l'effet du

traitement sont plus faibles dans les études observationnelles. Des IC de 95% ont été calculés en utilisant des erreurs types qui prennent en compte l'hétérogénéité des études incluses<sup>(126,164,165)</sup>.

Les méta-analyses ont été réalisées avec le logiciel R version 3.5.1 (meta package, téléchargé sur [www.r-project.org](http://www.r-project.org))<sup>(166)</sup>.

#### 4.3.2 Résultats

Une hétérogénéité élevée a été observée entre les études utilisant des bases de données de santé informatisées en termes d'efficacité ( $I^2= 86\%$ ) et en termes de sécurité ( $I^2= 91\%$ ). D'où le choix du modèle à effets aléatoires.

Les ECR ont montré que les AOD ont une balance bénéfique/risque favorable par rapport aux AVK. Les AOD ont été associés à une réduction significative du risque d'AVC ou d'embolie systémique (HR = 0.85 [0.74; 0.98]), d'hémorragies majeures (HR = 0.77 [0.62; 0.95]) et d'hémorragies intracrâniennes (HR = 0.42 [0.33; 0.53]). Néanmoins, on n'observe pas de différence significative du risque d'hémorragie gastro-intestinale entre les AOD et les AVK (HR = 1.04 [0.79; 1.37]).

Lorsque l'on compare les estimations des études de cohortes prospectives à celles obtenues dans les ECR, nous disposons d'un faible nombre d'études. Les principaux résultats sont cependant présentés ci-dessous.

- Les graphiques en entonnoir n'apportent pas d'information marquante pour les études de cohortes prospectives quant à la suspicion d'une quelconque interaction compte tenu du faible nombre d'études (carrés noirs Figures 18 et 19).
- Les AOD sont associés à une réduction significative de 54% du risque d'AVC ou d'embolie systémique (HR = 0.46 [0.30 ; 0.71], Figure 20) et à une réduction non significative du risque d'hémorragies majeures par rapport aux AVK (HR = 0.43 [0.16 ; 1.16], Figure 21).
- Par rapport aux ECR, les études de cohortes prospectives étaient associées à une surestimation significative de 46% de l'effet traitement (RHR= 0.54 [0.34 ; 0.85] en termes d'efficacité. Nous n'avons pas observé de surestimation significative en termes d'hémorragies majeures (RHR = 0.56, [0.20 ; 1.54]) même si la tendance semble identique (Figure 22).

- L'analyse de sensibilité, excluant l'étude<sup>(204)</sup> n'ayant pas utilisé d'ajustement ou d'appariement, fournit des résultats similaires : RHR = 0.49, [0.30 ; 0.81] pour le critère d'efficacité et RHR = 0.53, [0.303 ; 0.86] pour le critère de sécurité.

Figure 18: Graphique en entonnoir des résultats sur les AVC ou les embolies systémiques en fonction du type d'études

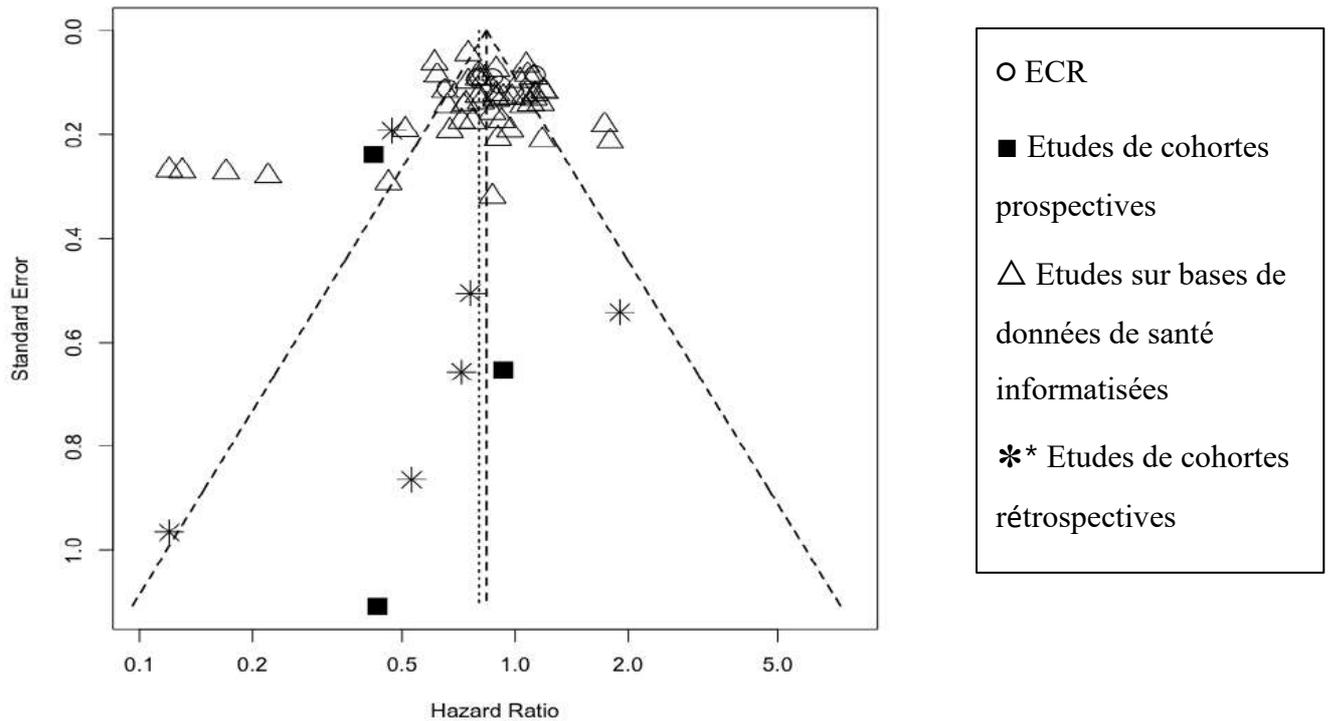
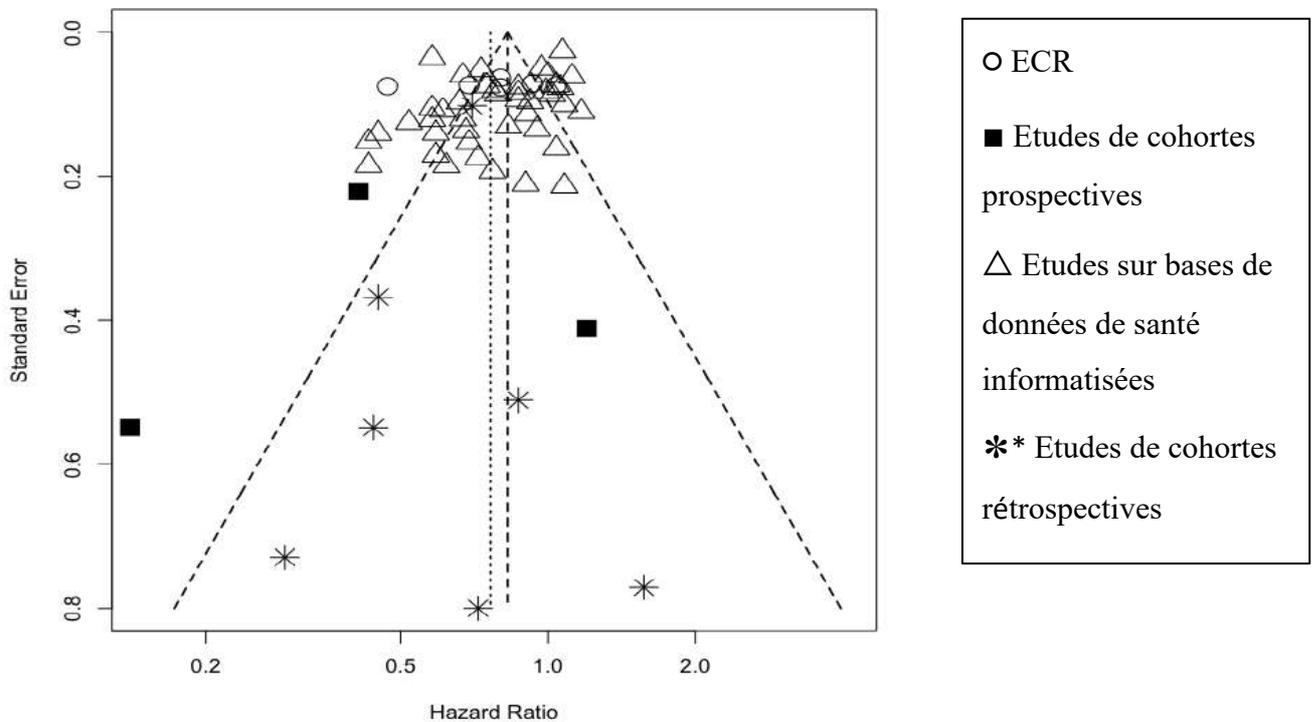


Figure 19: Graphique en entonnoir des résultats sur les hémorragies majeures en fonction du type d'études



Les points essentiels concernant la comparaison des estimations issues des ECR et celles issues des bases de données de santé informatisées sont les suivants :

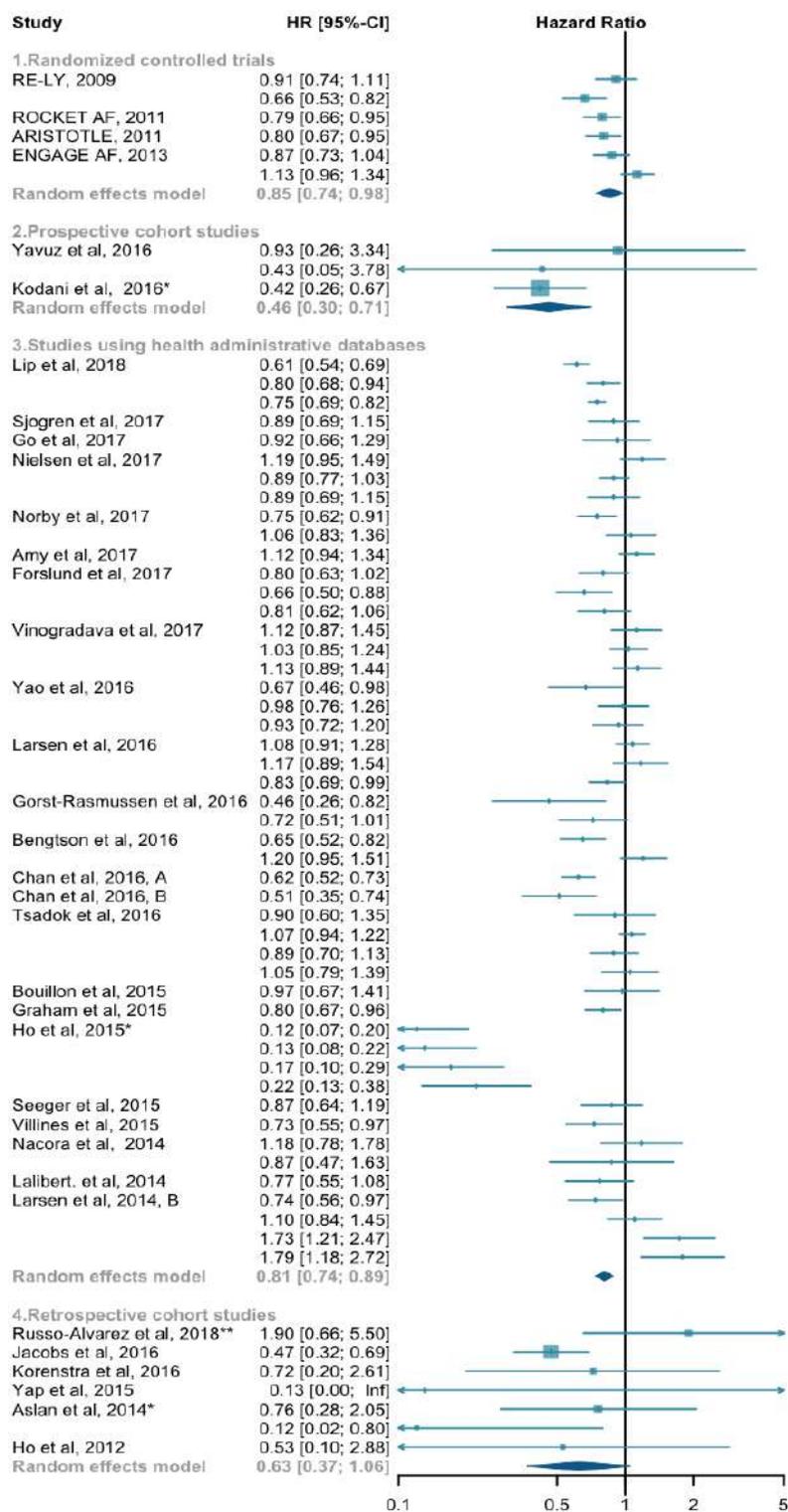
- Les graphiques en entonnoir illustrant les résultats de la survenue d'AVC, d'embolie systémique et d'hémorragies majeures montrent que ces études sont de taille importante, sans asymétrie particulière, que ce soit en termes d'efficacité ou en termes de sécurité (triangles, Figures 18, 19). Les résultats d'efficacité et de sécurité de ces études sont superposés aux résultats observés dans les ECR (cercles, Figures 18 et 19). Ainsi, les estimations issues des études sur bases de données de santé informatisées devraient être apparemment comparables à celles des ECR en termes d'efficacité et de sécurité.
- Dans les études sur bases de données de santé informatisées, les AOD étaient associés à une réduction significative de 19% du risque d'AVC ou d'embolie systémique ( $HR = 0.81 [0.74 ; 0.89]$ ) et de 21% du risque d'hémorragies majeures ( $HR = 0.78 [0.72 ; 0.85]$ ) (Figures 20 et 21).
- Par rapport aux ECR, nous n'avons pas observé de surestimation ou sous-estimation de l'effet des AOD dans les études sur bases de données de santé informatisées en termes

d'efficacité (RHR = 0.95 [0.81 ; 1.13]) ou de sécurité (RHR = 1.01 [0.81 ; 1.27]). Pour les deux critères d'efficacité et de sécurité, l'estimation de l'effet des AOD était similaire entre les ECR et les études sur bases de données de santé informatisées (Figure 22).

La comparaison des études de cohortes rétrospectives aux ECR permet de faire les observations suivantes :

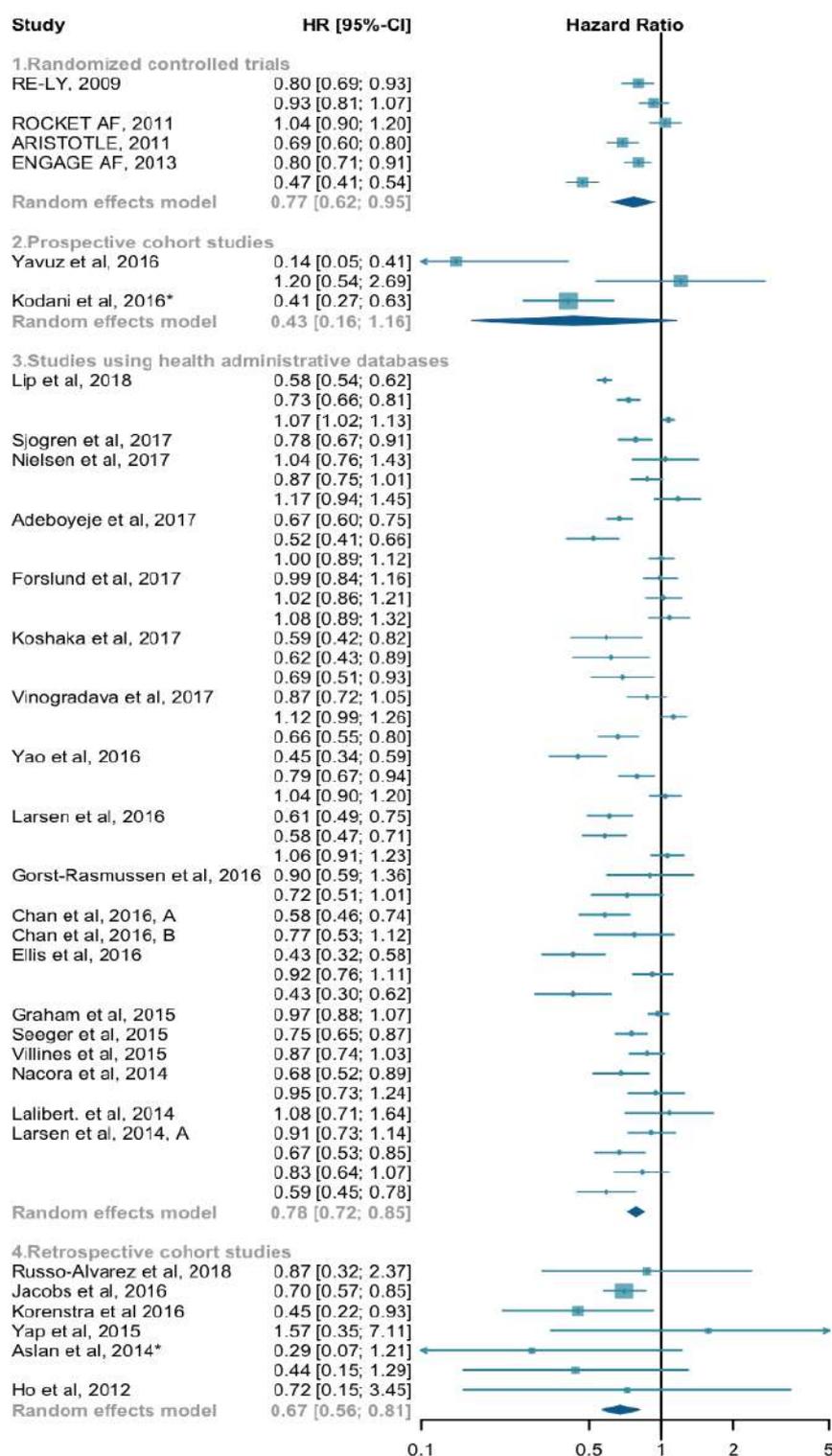
- Les graphiques en entonnoir en fonction du type d'étude (Figures 18 et 19) montrent que les études de cohorte rétrospectives (étoiles) sont de petites études et sont associées à une estimation plus favorable de l'effet du traitement en termes d'efficacité et de sécurité par rapport aux ECR (cercles).
- Dans les études de cohortes rétrospectives, les AOD ont été associés à une diminution non significative de 37% du risque d'AVC ou d'embolie systémique (HR = 0.63, [0.37 ; 1.06]) et une réduction significative de 33% du risque d'hémorragies majeures (HR = 0.67, [0.56 ; 0.81]) (Figures 20 et 21).
- Par rapport aux ECR, les études de cohorte rétrospectives sont associées à une surestimation non significative de l'effet des AOD de 26 % en termes d'efficacité (RHR = 0.74 [0.43 ; 1.28]) et de 13% en termes de sécurité (RHR = 0.87 [0.66; 1.15]) (Figure 22).
- L'analyse de sensibilité, excluant les études<sup>(209-211)</sup> n'ayant pas utilisé d'ajustement ou d'appariement, fournit des résultats similaires : RHR = 0.93 [0.38 ; 2.30] pour le critère d'efficacité, RHR = 0.88 [0.66 ; 1.17] pour le critère de sécurité.

Figure 20: Comparaison des estimations des effets du traitement en fonction du type d'études sur les AVC ou les embolies systémiques



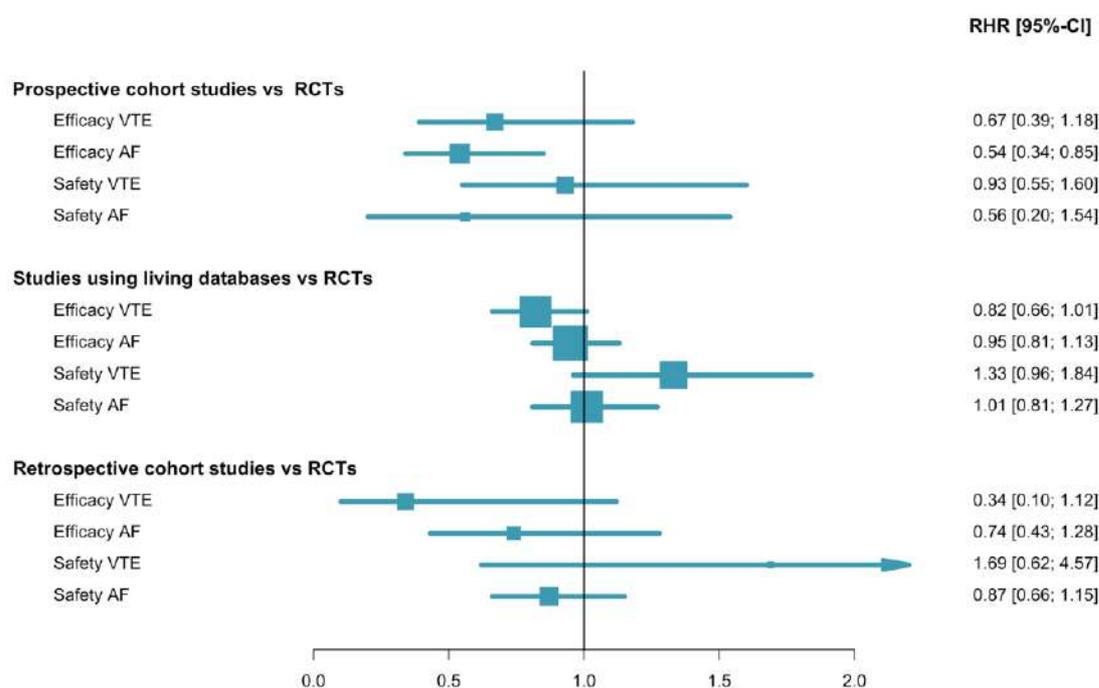
\* : RR estimé, \*\* : AVC ischémique et ETEV. Une ligne correspond à l'effet d'une molécule d'AOD par rapport aux AVK. Une même étude peut être représentée par n lignes en fonction de n molécules de AOD.

**Figure 21:** Comparaison des estimations des effets du traitement en fonction du type d'études sur les hémorragies majeures



\* : RR estimé. Une ligne correspond à l'effet d'une molécule d'AOD par rapport aux AVK. Une même étude peut être représentée par n lignes en fonction de n molécules de AOD.

**Figure 22:** Ratio des HR (RHR) et intervalles de confiance à 95% des RHR pour quantifier la différence d'estimation de l'effet traitement entre les études observationnelles et les ECR



- Les résultats des critères secondaires concernant la comparaison des estimations issues des ECR et celles issues des études observationnelles sont présentés ci-dessous.
  - Dans les études sur bases de données de santé informatisées, les AOD étaient associés à une réduction significative de 48% du risque d'hémorragie intracrânienne (HR = 0.52 [0.47; 0.58]). Néanmoins, il n'y avait pas de différence entre les AOD et les AVK (HR = 1.02 [0.93 ; 1.12]) en termes d'hémorragie gastro-intestinale (annexe2- figures A2 et A3). Dans les études sur bases de données de santé informatisées, l'estimation de l'effet des AOD était 24% plus faible (RHR = 1.24 [0.96 ; 1.60]) en termes d'hémorragie intracrânienne. Néanmoins, nous n'avons pas observé de surestimation ou sous-estimation de l'effet des AOD dans les études sur bases de données de santé informatisées en termes d'hémorragie gastro-intestinale (RHR = 0.98 [0.73 ; 1.13]) (annexe2-figure A4).
  - Dans les études de cohortes prospectives, les AOD étaient associés à une réduction significative de 64% du risque d'hémorragie intracrânienne (HR = 0.36 [0.26 ; 0.49]). Néanmoins, les AOD n'étaient pas associés à une réduction significative du risque d'hémorragie gastro-intestinale (HR = 0.60 [0.23 ; 1.60]) (annexe2- figures A2 et A3).

Par rapport aux ECR, nous n'avons pas observé de surestimation significative dans les études de cohortes prospectives en termes d'hémorragie intracrânienne (RHR = 0.86, [0.58 ; 1.27]) ni en termes d'hémorragie gastro-intestinale (RHR = 1.24 [0.96 ; 1.60]) (annexe2-figure A4).

- La comparaison des études de cohortes rétrospectives aux ECR permet de faire les observations suivantes : les AOD n'étaient pas associés à une réduction significative du risque d'hémorragie intracrânienne par rapport aux AVK (HR =0.93, [0.24; 3.51]) ni du risque d'hémorragie gastro-intestinale (HR =0.57, [0.24; 1.35]) (annexe2- figures A2 et A3). Il n'a pas été possible de quantifier une différence d'estimation des effets des traitements par rapport aux ECR (RHR = 2.21 [0.57 ; 8.65] pour les hémorragies intracrâniennes, RHR = 0.51 [0.22 ; 1.36] pour les hémorragies gastro-intestinales) (annexe2-figure A4).

#### 4.3.3 Analyse de sensibilité

Une analyse de sensibilité a été réalisée sur les critères d'efficacité et de sécurité en excluant les études observationnelles n'ayant pas appliqué de méthodes d'ajustement ou d'appariement. Dans cette analyse, 1 étude de cohorte prospective et 3 études de cohortes rétrospectives ont été exclues. Pour les études de cohortes prospectives, les résultats étaient similaires aux résultats obtenus dans l'analyse principale pour le critère d'efficacité (RHR = 0.49, [0.30 ; 0.81]) et pour le critère de sécurité (RHR = 0.53, [0.303 ; 0.86]). Dans les études de cohortes rétrospectives, les résultats étaient également similaires aux résultats obtenus de l'analyse principale (RHR = 0.93 [0.38 ; 2.30] et RHR = 0.88 [0.66 ; 1.17], respectivement).

#### 4.4 Discussion des principaux résultats et limites

Dans cette étude méta-épidémiologique, nous avons comparé les estimations des effets du traitement avec AOD par rapport aux AVK dans la prévention des TEA chez les patients avec FA non valvulaire dans les études observationnelles par rapport aux ECR. Il a été montré que les caractéristiques des patients avec FA inclus dans les ECR peuvent être différentes des celles des études observationnelles. Cette différence dans les caractéristiques peut modifier l'estimation des effets du traitement en cas de présence d'une interaction entre les caractéristiques des patients et l'effet du traitement. Ainsi, elle peut entraîner des différences

dans les estimations des effets du traitement entre les ECR et les études observationnelles sur les risques de survenue d'AVC, d'embolie systémique ou d'hémorragie majeure. Compte tenu des différences observées, il n'est cependant pas possible d'anticiper dans quel sens les résultats pourraient différer entre études observationnelles et ECR.

Avec seulement deux études de cohortes prospectives disponibles pour cette étude méta-épidémiologique, il n'est pas possible de quantifier un biais d'estimation de façon fiable, ces études n'ayant par ailleurs pas utilisé la méthode de SP pour tenter de rechercher une comparabilité apparente entre les deux groupes, AOD et AVK. Une seule étude a utilisé la méthode de régression multiple afin de minimiser les biais dans ces études. Du fait d'une forte suspicion d'une confusion résiduelle, ces cohortes prospectives étaient associées à une surestimation significative de 46% de l'effet du traitement en termes d'efficacité par rapport aux ECR (RHR= 0.54 [0.34 ; 0.85]). Malgré le faible nombre d'études de cohortes prospectives incluses dans cette étude méta-épidémiologique, cette dernière a pu confirmer les données observées dans la MTEV.

Pour les études de cohortes rétrospectives, les caractéristiques initiales des patients sous AOD ne permettent pas de savoir si la sélection des patients était modifiée par rapport aux ECR. Pour 3 études sur 6, l'ajustement à un SP a permis une comparabilité apparente des caractéristiques des patients entre les deux groupes, AOD et AVK. Par conséquent, les populations incluses dans les études de cohortes prospectives et les ECR étaient apparemment assez proches. Cette étude méta-épidémiologique permet de suspecter que ces cohortes rétrospectives peuvent être associées à une surestimation, certes non significative, mais de 26% en termes d'efficacité (RHR = 0.74 [0.43 ; 1.28]) et de 13% en termes de sécurité (RHR =0.87 [0.66; 1.15]). Un manque de puissance est à déplorer dans le cadre de cette estimation des RHR, mais la surestimation de l'effet de traitement est probablement en lien avec l'existence de biais de confusion résiduelle qui persiste dans ces études observationnelles, même avec l'utilisation d'un SP.

Dans les études sur bases de données de santé informatisées, les caractéristiques initiales des patients avec FA inclus étaient différentes de celles des patients inclus dans les ECR avec un pourcentage de comorbidités plus faible tels que l'hypertension, le diabète et l'insuffisance cardiaque mais avec un âge plus avancé. Ces observations ne permettent pas de savoir si les patients sélectionnés dans les ECR sont très différents des patients traités en pratique clinique, alors même que les patients des études sur bases de données de santé informatisées ne sont pas

sélectionnés selon des critères d'inclusion et d'exclusion stricts comme dans les ECR. Les résultats en termes d'efficacité et de sécurité des études utilisant des bases de données de santé informatisées étaient en accord avec celles des ECR avec  $RHR = 0.95$  [0.81 ; 1.13] pour le critère d'efficacité et  $RHR = 1.01$  [0.81 ; 1.27] pour le critère de sécurité. Nous n'avons pas observé de différences significatives dans les estimations de l'effet du traitement (sous-estimation ou surestimation de l'effet des AOD) entre les études utilisant des bases de données de santé informatisées et les ECR. Plusieurs explications peuvent être évoquées. Tout d'abord, l'inclusion d'un grand nombre d'études utilisant des bases de données de santé informatisées a permis d'améliorer la précision et la fiabilité des résultats mis en commun de ces études et de les rapprocher des résultats obtenus par les ECR. Deuxièmement, l'absence d'une confusion résiduelle après ajustement ou appariement dans cette population particulière de FA a permis de retrouver les estimations des ECR.

En résumé, les études utilisant des bases de données de santé informatisées semblent inclure des patients plus proches des patients traités dans la vraie vie. De plus, les résultats de ces études semblent aussi être proches de ceux obtenus par les ECR. Néanmoins, on ne peut pas éliminer la présence des facteurs de confusion inhérents à ces études observationnelles.

## 5 Discussion générale

Il est tout d'abord important de rappeler le cadre de cette thèse. Les deux études méta-épidémiologiques présentées dans cette thèse ont pour but de quantifier les biais de confusion inhérents aux études observationnelles. Elles illustrent la comparaison entre les ECR et les études observationnelles sur la population des patients inclus et les résultats en termes d'efficacité et de sécurité. Ces deux études méta-épidémiologiques portent sur des études ayant évalué les AOD par rapport aux AVK dans le traitement des ETEV et dans la FA.

### 5.1 Une sélection moindre des patients dans les études observationnelles ?

Ces deux études méta-épidémiologiques ont permis de comparer les populations incluses dans les ECR et dans les études observationnelles en distinguant les différences et les similitudes entre ces deux types d'études. Globalement, ces études méta-épidémiologiques ont montré que les études de cohortes (prospectives et rétrospectives) évaluant les AOD sélectionnent les patients aussi bien que les ECR. Ainsi, ces études de « vrai vie » ne remplissent pas leur rôle de « sélection de moins de patients avec un effectif plus important possible ». Elles ne permettent donc pas de documenter plus l'intérêt d'un traitement chez des patients ayant un tableau clinique complexe. Ces études ne sont donc probablement pas si utiles que ça en pratique. En parallèle, Les études sur bases de données de santé informatisée sélectionnent les patients moins que les ECR. Néanmoins, en utilisant des méthodes d'ajustement afin de rendre les groupes de patients comparables, les populations des études sur bases de données de santé informatisée deviennent similaires aux ceux des ECR.

### 5.2 Principaux résultats des études méta-épidémiologiques dans le traitement des ETEV et dans la FA

Les deux études méta-épidémiologiques présentées dans ce travail de thèse ont permis de comparer, entre les études observationnelles et les ECR, les estimations de l'effet du traitement des AOD par rapport aux AVK dans le traitement des ETEV et dans la FA. Les études observationnelles ont été classées en trois grandes catégories selon leur schéma expérimental : études de cohortes prospectives, études de cohortes rétrospectives et études sur bases de

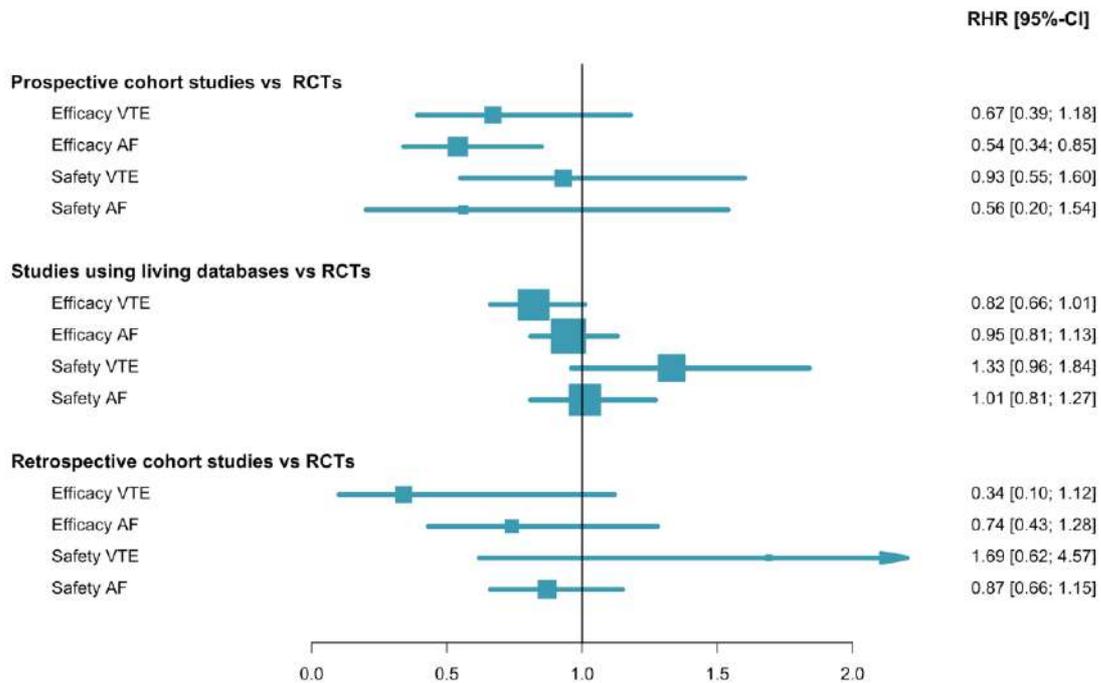
données de santé informatisées.

Par rapport aux ECR, et en termes d'efficacité, les études de cohortes prospectives étaient associées à une surestimation non significative de 33% (RHR = 0.67 [0.39 ; 1.18]) de l'effet du traitement chez les patients avec MTEV et à une surestimation significative de 46% de l'effet du traitement (RHR= 0.54 [0.34 ; 0.85]) chez les patients avec FA. Ces surestimations de l'effet de traitement peuvent être expliquées par l'existence de biais de confusion résiduelle qui persiste dans ces études observationnelles même avec l'utilisation des méthodes de correction du biais. Par conséquent, ces méthodes statistiques d'ajustement ont corrigé uniquement les facteurs de confusion observables et les variables mesurées.

Néanmoins, dans ces études de cohortes prospectives, et en termes de sécurité, nous n'avons pas observé de surestimation significative même si la tendance semble identique avec RHR = 0.93 [0.55; 1.60] chez les patients avec MTEV et RHR = 0.56, [0.20 ; 1.54] chez les patients avec FA (Figure 23).

En termes d'efficacité, les résultats des études utilisant des bases de données de santé informatisées étaient proches de ceux obtenus par les ECR avec RHR = 0.82 [0.66 ; 1.01] chez les patients avec MTEV et avec RHR = 0.95 [0.81 ; 1.13] chez les patients avec FA. En termes de sécurité, les résultats de ces études étaient associés à une sous-estimation non significative de 33 % du bénéfice des AOD (RHR = 1.33 [0.96 ; 1.84]) chez les patients avec MTEV alors que les résultats chez les patients avec FA étaient en accord avec celles des ECR (RHR = 1.01 [0.81 ; 1.27]).

**Figure 23:** Ratio des HR (RHR) et intervalles de confiance à 95% des RHR pour quantifier la différence d'estimation des effets traitement entre les études observationnelles et les ECR



Cet accord entre les résultats (estimations de l'effet du traitement) des études utilisant des bases de données de santé informatisées et les ECR peut être expliqué par l'inclusion d'un grand nombre d'études utilisant des bases de données de santé informatisées dans ces études méta-épidémiologiques. En effet, ce grand nombre a permis d'améliorer la précision et la fiabilité des résultats mis en commun de ces études et de les rapprocher des résultats obtenus par les ECR.

Chez les patients avec MTEV, les cohortes rétrospectives n'ont pas apporté des informations marquantes vu le faible nombre de ces études. En effet, il n'a pas été possible de quantifier une différence d'estimation des effets traitement par rapport aux ECR pour le critère d'efficacité (RHR= 0.34 [0.10 ; 1.12]) ni pour le critère de sécurité (RHR =1.69, [0.62 ; 4.57]). Chez les patients avec FA, on note également un possible manque de puissance avec des résultats non significatifs en termes d'efficacité (RHR = 0.74 [0.43 ; 1.28]) et en termes de sécurité (RHR = 0.87 [0.66; 1.15]).

En résumé, ces deux études méta-épidémiologiques ont permis de quantifier la différence d'estimation des effets de traitement des études de cohortes (prospectives et rétrospectives) par

rapport aux ECR et d'estimer une partie du biais inhérent à ces études de cohortes. Elles ont également montré que les résultats des études sur bases de données de santé informatisées et des ECR ne différaient pas. Pour autant, l'absence de biais n'est pas la règle. Premièrement, ces études sont conduites en ouvert, ce qui entraîne une sur-estimation de l'efficacité de traitement et une sous-estimation de l'avantage de la sécurité d'un traitement. Deuxièmement, ces études sont sujettes au biais de sélection dû à l'absence de randomisation. En effet, les patients aux pathologies les moins graves sont souvent sous AOD, ce qui entraîne une tendance à la sur-estimation du bénéfice et du risque. Troisièmement, la population incluse dans ces bases de données est généralement plus à risque que la population incluse dans les ECR. Par conséquent, les résultats attendus de ces études peuvent être : 1/ soit identiques aux ECR en l'absence d'une interaction entre les caractéristiques des patients et l'effet du traitement 2/ soit sur-estimés en termes d'efficacité (plus le patient est grave plus le bénéfice du traitement est apparent).

D'autres limites se rajoutent à la précédente limite. Certaines bases de données ne disposent pas d'information pour des variables cliniques importantes : indication d'une prescription, traitements reçus en cours d'hospitalisation, et résultats des explorations paracliniques. La description détaillée des données, l'enrichissement des données administratives par des données médicales et le soin apporté au protocole d'étude des données issues de ces bases sont donc autant de prérequis indispensables à leur utilisation optimale.

L'analyse de sensibilité réalisée dans les études de la FA excluant les études observationnelles n'ayant pas appliqué de méthodes d'ajustement ou d'appariement n'a pas modifié les résultats de l'analyse principale. Les surestimations des effets observés peuvent donc laisser penser que les méthodes d'ajustement des autres études n'ont pas réussi à corriger le biais de confusion inhérent aux études observationnelles.

### 5.3 Limites de ce travail de pharmaco-épidémiologie

La première limite de ce travail est la présence d'une hétérogénéité élevée entre les études observationnelles. En effet, un écart important a été observé entre les estimations des effets de traitements de ces études. Il fallait donc savoir la cause de ces variabilités des estimations d'effet du traitement avant de l'imputer au design utilisé dans ces études. Cette difficulté d'interprétation des résultats peut être due à la multiplicité des molécules d'AOD étudiés et donc à la comparaison des molécules certes d'AOD mais différentes. Pour mieux comprendre

les résultats observés, il fallait donc réaliser des analyses complémentaires: des analyses stratifiées par molécule d'AOD et des analyses en fonction des scores de qualité méthodologique des études observationnelles selon les échelles Robins et Robins-I. Une autre limite est la manque de puissance pour les études de cohorte prospectives et rétrospectives.

#### 5.4 Perspectives

Premièrement, dans ces deux études méta-épidémiologiques, on note un manque de puissance notamment pour les études de cohorte. Un travail similaire et complémentaire pourrait être réalisés avec les antiagrégants plaquettaires. La littérature est très riche sur les antiagrégants plaquettaires, que ce soit pour les ECR et les cohortes. Du fait de l'ancienneté de ces molécules par rapports aux AOD, les études sur les antiagrégants plaquettaires sont plus nombreuses, ce qui rendra la puissance plus importante. Ceci est d'autant plus vrai qu'il existe une certaine similitude entre les caractéristiques des patients sous AOD et celles des patients sous antiagrégants plaquettaires.

Deuxièmement, il est important aussi de réaliser dans la partie FA une analyse des patients incidents (naïfs de traitement) par rapport aux patients prévalent (déjà exposés au traitement). En effet, les patients naïfs de traitement permettent d'éviter le biais d'immortalité dans les études de cohortes. Ce biais de temps immortel a généralement tendance à attribuer à tort un bénéfice important au traitement étudié (ou à exagérer un bénéfice réel).

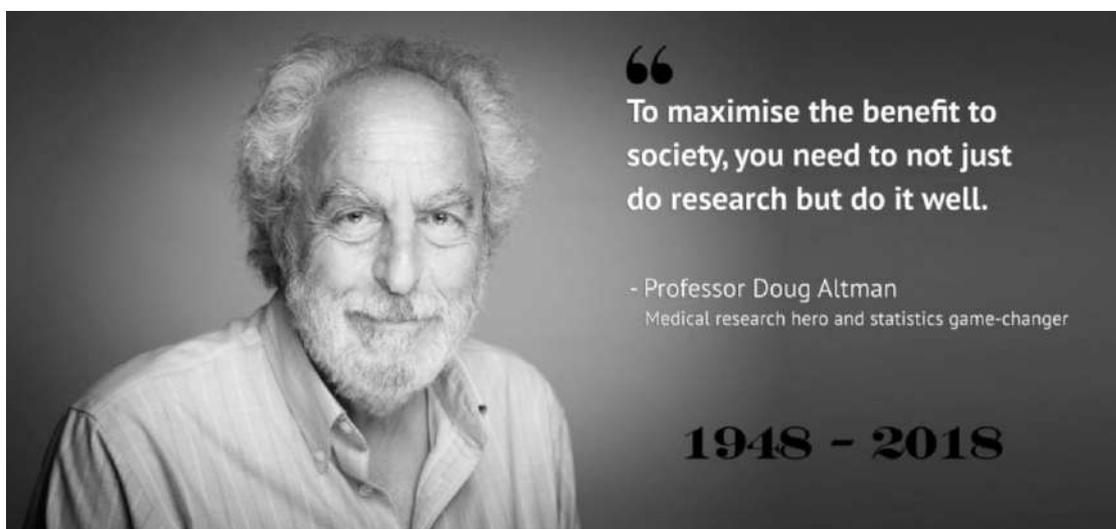
Troisièmement, les études méta-épidémiologiques effectuées dans le cadre de cette thèse donnent un aperçu du futur de l'utilisation et de l'utilité des différents types d'études : observationnelles et des ECR.

#### 5.5 Les leçons actuelles de ce travail

Les ECR restent irremplaçables. La sur-estimation ou la sous-estimation de l'effet de traitement observés dans les études observationnelles, et malgré l'application des méthodes de correction et d'ajustement de biais, peut conduire les cliniciens à déduire à tort le bénéfice d'un traitement. Des ajustements statistiques peuvent tenir compte de certains facteurs de confusion. Cependant, de nombreux facteurs de confusion ne peuvent pas être ajustés parce qu'ils ne sont pas mesurés

ou sont inconnus. Le problème des facteurs de confusion est éliminé par les ECR dans lesquelles le processus de randomisation équilibre efficacement tous les facteurs de confusion (connus ou inconnus), créant ainsi des groupes ayant des caractéristiques essentiellement identiques qui ne varient qu'en fonction du traitement qui leur est attribué. D'où le caractère indispensable du schéma des ECR qui offre le meilleur niveau de preuve scientifique pour estimer l'efficacité et la sécurité d'un nouveau traitement. Les ECR sont donc irremplaçables et, de fait, la multiplicité des ECR à grande échelle réalisées au cours des dix dernières années a considérablement transformé la prise en charge des patients.

Nous proposons donc d'arrêter la conduite des cohortes et de réaliser des ECR pragmatiques en vie réelle. Les ECR doivent également évoluer pour devenir moins lourds, moins coûteux et plus généralisables en s'intégrant dans le monde réel. Ces ECR en situation réelle peuvent être réalisés avec une collecte prospective pour de l'évaluation de traitement, à un coût réduit et avec une durée de suivi plus longue. A titre d'exemple, une étude analysant, après une identification dans des registres au Royaume-Uni<sup>(226)</sup>, 15 480 personnes atteintes de diabète ont été randomisées et suivies pendant plus de 7 ans. Ces approches innovantes peuvent exploiter la puissance des données du monde réel tout en gardant la capacité de la randomisation afin de garantir une véritable évaluation de l'effet d'un traitement. Elles sont donc solides afin d'évaluer l'efficacité d'un traitement, et pour donner des preuves claires et impartiales dans les soins cliniques. En l'absence de randomisation, les analyses de la plupart des données d'observation du monde réel, quelle que soit leur sophistication, ne peuvent être considérées que comme génératrices d'hypothèses. La déclaration célèbre du Prof. Douglas Altman est toujours d'actualité, notamment au vu de la recherche clinique précipitée que nous avons observée dans le contexte COVID-19.



Parallèlement aux efforts de recherche s'appuyant sur les ECR, compte tenu du volume important d'informations disponibles dans les bases de données de santé informatisées, nous proposons en parallèle de les optimiser. En effet, certaines informations médicales sont manquantes dans ces bases car elles ne rentrent pas en considération pour le remboursement des molécules. L'amélioration de ces bases consisteraient à inclure des informations cliniques exhaustives comme la taille, le poids, la tension artérielle du patient, ses habitudes alimentaires dont sa consommation d'alcool, son niveau d'exercice physique ou d'exposition exacte au tabac ou aux polluants atmosphériques, ou encore ses prédispositions génétiques. Il faut aussi rajouter les résultats des examens médicaux cliniques et paracliniques comme les résultats biologiques ou radiologiques. Mais cela passe par un chaînage de l'information et un désir politique en termes de santé publique qui dépasse le cadre de cette thèse.

## 6 Références

1. Vray M, Hamelin B, Jaillon P, Andréjak M, Benamouzig R, Boissel J-P, et al. Place relative des essais cliniques comparatifs et des suivis de cohorte dans l'évaluation préet post-AMM des médicaments. *Thérapies* [Internet]. 2005 Jul 1 [cited 2020 May 29];60(4):339–44. Available from: <http://www.sciencedirect.com/science/article/pii/S0040595716303079>
2. Begaud (b ), BEGAUD (B.), Lennep (f.) V, POLTON (D.). Les données de vie réelle, un enjeu majeur pour la qualité des soins et la régulation du système de santé : L'exemple du médicament. - Résultats de votre recherche - Banque de données en santé publique. In 2017 [cited 2020 May 30]. Available from: <http://bdsp-ehesp.inist.fr/vibad/index.php?action=getRecordDetail&idt=517785>
3. Stamenkovic S, Solesse A, Zanetti L, Zagury P, Vray M. Guide de la Haute autorité de santé (HAS) : les études post-inscription sur les technologies de santé (médicaments, dispositifs médicaux et actes) : principes et méthodes. /data/revues/00405957/v67i5/S0040595716308745/ [Internet]. 2016 Jul 8 [cited 2020 May 30]; Available from: <https://www.em-consulte.com/en/article/1067630>
4. Baumgarten M, Olsen C. Confounding in Epidemiology. :36.
5. Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M. Randomized Controlled Trials. *Dtsch Arztebl Int* [Internet]. 2011 Sep [cited 2020 May 18];108(39):663–8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3196997/>
6. Aptel F, Cucherat M, Blumen-Ohana E, Denis P. L'interprétation des essais cliniques. /data/revues/01815512/v34i10/S0181551211002968/ [Internet]. 2011 Nov 30 [cited 2020 May 30]; Available from: <https://www.em-consulte.com/en/article/676180>
7. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995 Feb;16(1):62–73.
8. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995 Feb 1;273(5):408–12.
9. Redmond CK, Colton T. *Biostatistics in Clinical Trials*. John Wiley & Sons; 2001. 530 p.
10. Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions : Cochrane Book Series*. :674.
11. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* [Internet]. 2016 Oct 12 [cited 2020 Jun 12];355. Available from: <https://www.bmj.com/content/355/bmj.i4919>
12. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann*

Intern Med. 2011 Oct 18;155(8):529–36.

13. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* [Internet]. 2019 Aug 28 [cited 2020 Jun 11];l4898. Available from: <http://www.bmj.com/lookup/doi/10.1136/bmj.l4898>
14. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* [Internet]. 2010 Mar 24 [cited 2020 May 20];340. Available from: <https://www.bmj.com/content/340/bmj.c332>
15. Porta M. *A Dictionary of Epidemiology*. Sixth Edition. Oxford, New York: Oxford University Press; 2014. 376 p.
16. Makady A, de Boer A, Hillege H, Klungel O, Goettsch W, (on behalf of GetReal Work Package 1). What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews. *Value Health*. 2017 Aug;20(7):858–65.
17. Klebanoff MA, Snowden JM. Historical (retrospective) cohort studies and other epidemiologic study designs in perinatal research. *American Journal of Obstetrics and Gynecology* [Internet]. 2018 Nov 1 [cited 2020 May 26];219(5):447–50. Available from: <http://www.sciencedirect.com/science/article/pii/S0002937818307233>
18. Goldberg M, Coeuret-Pellicer M, Ribet C, Zins M. Cohortes épidémiologiques et bases de données d'origine administrative - Un rapprochement potentiellement fructueux. *Med Sci (Paris)* [Internet]. 2012 Apr 1 [cited 2020 May 30];28(4):430–4. Available from: <https://www.medecinesciences.org/articles/medsci/abs/2012/05/medsci2012284p430/medsci2012284p430.html>
19. Trifirò G, Sultana J, Bate A. From Big Data to Smart Data for Pharmacovigilance: The Role of Healthcare Databases and Other Emerging Sources. *Drug Saf*. 2018 Feb;41(2):143–9.
20. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: new opportunities, familiar challenges. *Clin Epidemiol*. 2017;9:245–50.
21. Lau EC, Mowat FS, Kelsh MA, Legg JC, Engel-Nitz NM, Watson HN, et al. Use of electronic medical records (EMR) for oncology outcomes research: assessing the comparability of EMR information to patient registry and health claims data. *Clin Epidemiol*. 2011;3:259–72.
22. Fife CE, Walker D, Thomson B. Electronic Health Records, Registries, and Quality Measures: What? Why? How? *Adv Wound Care (New Rochelle)*. 2013 Dec;2(10):598–604.
23. Rowan CG, Flory J, Gerhard T, Cuddeback JK, Stempniewicz N, Lewis JD, et al. Agreement and validity of electronic health record prescribing data relative to pharmacy claims data: A validation study from a US electronic health record database. *Pharmacoepidemiology and Drug Safety* [Internet]. 2017 [cited 2020 May 29];26(8):963–72. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/pds.4234>
24. Pottgård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M. Data Resource Profile: The Danish National Prescription Registry. *Int J Epidemiol*. 2017 01;46(3):798–798f.
25. Mitchell JB, Bubolz T, Paul JE, Pashos CL, Escarce JJ, Muhlbaier LH, et al. Using

- Medicare Claims for Outcomes Research. *Medical Care* [Internet]. 1994 [cited 2020 May 29];32(7):JS38–51. Available from: <https://www.jstor.org/stable/3766446>
26. Riley GF. Administrative and claims records as sources of health care cost data. *Med Care*. 2009 Jul;47(7 Suppl 1):S51-55.
  27. Wettermark B, Zoëga H, Furu K, Korhonen M, Hallas J, Nørgaard M, et al. The Nordic prescription databases as a resource for pharmacoepidemiological research--a literature review. *Pharmacoepidemiol Drug Saf*. 2013 Jul;22(7):691–9.
  28. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol*. 2015 Jun;44(3):827–36.
  29. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System--a national resource for evidence development. *N Engl J Med*. 2011 Feb 10;364(6):498–9.
  30. Forrow S, Campion DM, Herrinton LJ, Nair VP, Robb MA, Wilson M, et al. The organizational structure and governing principles of the Food and Drug Administration's Mini-Sentinel pilot program. *Pharmacoepidemiol Drug Saf*. 2012 Jan;21 Suppl 1:12–7.
  31. Coloma PM, Schuemie MJ, Trifirò G, Gini R, Herings R, Hippisley-Cox J, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. *Pharmacoepidemiol Drug Saf*. 2011 Jan;20(1):1–11.
  32. AsPEN collaborators, Andersen M, Bergman U, Choi N-K, Gerhard T, Huang C, et al. The Asian Pharmacoepidemiology Network (AsPEN): promoting multi-national collaboration for pharmacoepidemiologic research in Asia. *Pharmacoepidemiol Drug Saf*. 2013 Jul;22(7):700–4.
  33. Suissa S, Henry D, Caetano P, Dormuth CR, Ernst P, Hemmelgarn B, et al. CNODES: the Canadian Network for Observational Drug Effect Studies. *Open Med*. 2012;6(4):e134-140.
  34. Lewis NJ, Patwell JT, Briesacher BA. The role of insurance claims databases in drug therapy outcomes research. *Pharmacoeconomics*. 1993 Nov;4(5):323–30.
  35. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. *J Clin Epidemiol*. 2012 Feb;65(2):126–31.
  36. Lin KJ, Schneeweiss S. Considerations for the analysis of longitudinal electronic health records linked to claims data to study the effectiveness and safety of drugs. *Clin Pharmacol Ther*. 2016;100(2):147–59.
  37. Saczynski JS, Andrade SE, Harrold LR, Tjia J, Cutrona SL, Dodd KS, et al. A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiol Drug Saf*. 2012 Jan;21 Suppl 1:129–40.
  38. Jones N, Schneider G, Kachroo S, Rotella P, Avetisyan R, Reynolds MW. A systematic review of validated methods for identifying acute respiratory failure using administrative and claims data. *Pharmacoepidemiol Drug Saf*. 2012 Jan;21 Suppl 1:261–4.

39. Tamariz L, Harkins T, Nair V. A systematic review of validated methods for identifying ventricular arrhythmias using administrative and claims data. *Pharmacoepidemiol Drug Saf.* 2012 Jan;21 Suppl 1:148–53.
40. Franklin JM, Schneeweiss S. When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials? *Clin Pharmacol Ther.* 2017 Dec;102(6):924–33.
41. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Forum on Drug Discovery, Development, and Translation. Real-World Evidence Generation and Evaluation of Therapeutics: Proceedings of a Workshop [Internet]. Washington (DC): National Academies Press (US); 2017 [cited 2020 May 31]. (The National Academies Collection: Reports funded by National Institutes of Health). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK424402/>
42. Rush CJ, Campbell RT, Jhund PS, Petrie MC, McMurray JJV. Association is not causation: treatment effects cannot be estimated from observational data in heart failure. *Eur Heart J.* 2018 01;39(37):3417–38.
43. Launois R, Trouiller J-B, Cabout E. Comment mesurer l'efficacité en vie réelle ? *Annales Pharmaceutiques Françaises* [Internet]. 2018 Nov 1 [cited 2020 May 31];76(6):421–35. Available from: <http://www.sciencedirect.com/science/article/pii/S0003450918301135>
44. Kwiatkowski F, Laporte S, Karem S. Le score de propension Partie I - Présentation de la méthode. :2.
45. Quilici S. Score de propension [Internet]. [cited 2020 May 31]. Available from: <https://docplayer.fr/61108732-Score-de-propension-sibilia-quilici-sanofi-pasteur-msd-seminaire-jem-sfes-31-janvier-2013.html>
46. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med.* 2007 Feb 20;26(4):734–53.
47. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol.* 2006 Jun 15;163(12):1149–56.
48. Kuss O, Blettner M, Börgermann J. Propensity Score: an Alternative Method of Analyzing Treatment Effects. *Dtsch Arztebl Int.* 2016 Sep 5;113(35–36):597–603.
49. Naess IA, Christiansen SC, Romundstad P, Cannegieter SC, Rosendaal FR, Hammerstrøm J. Incidence and mortality of venous thrombosis: a population-based study. *J Thromb Haemost.* 2007 Apr;5(4):692–9.
50. Spencer FA, Gore JM, Lessard D, Emery C, Pacifico L, Reed G, et al. Venous thromboembolism in the elderly. A community-based perspective. *Thromb Haemost.* 2008 Nov;100(5):780–8.
51. Sanchez O, Benhamou Y, Bertoletti L, Constant J, Couturaud F, Delluc A, et al. Recommandations de bonne pratique pour la prise en charge de la maladie veineuse thromboembolique chez l'adulte. Version courte. *Revue des Maladies Respiratoires* [Internet]. 2019 Feb [cited 2020 Jul 15];36(2):249–83. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0761842519300051>

52. Kannel W, Benjamin E. Final Draft Status of the Epidemiology of Atrial Fibrillation. *Med Clin North Am* [Internet]. 2008 Jan [cited 2020 May 3];92(1):17–ix. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2245891/>
53. Lip GYH, Tse HF, Lane DA. Atrial fibrillation. *Lancet*. 2012 Feb 18;379(9816):648–61.
54. Kirchhof P, Benussi S, Kotecha D, Ahlsson A, Atar D, Casadei B, et al. 2016 ESC Guidelines for the management of atrial fibrillation developed in collaboration with EACTS. *Eur Heart J* [Internet]. 2016 Oct 7 [cited 2020 Apr 27];37(38):2893–962. Available from: <https://academic.oup.com/eurheartj/article-lookup/doi/10.1093/eurheartj/ehw210>
55. Huxley RR, Lopez FL, Folsom AR, Agarwal SK, Loehr LR, Soliman EZ, et al. Absolute and Attributable Risks of Atrial Fibrillation in Relation to Optimal and Borderline Risk Factors: the Atherosclerosis Risk in Communities (ARIC) Study. *Circulation* [Internet]. 2011 Apr 12 [cited 2020 Apr 29];123(14):1501–8. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181498/>
56. Rahman F, Kwan GF, Benjamin EJ. Global epidemiology of atrial fibrillation. *Nat Rev Cardiol*. 2014 Nov;11(11):639–54.
57. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge. *Int J Stroke*. 2020 Jan 19;1747493019897870.
58. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *JAMA*. 2001 May 9;285(18):2370–5.
59. Haute Autorité de Santé. Guide parcours de soins Fibrillation atriale [Internet]. [cited 2020 May 3]. Available from: [https://www.has-sante.fr/jcms/c\\_1741768/fr/guide-parcours-de-soins-fibrillation-atriale](https://www.has-sante.fr/jcms/c_1741768/fr/guide-parcours-de-soins-fibrillation-atriale)
60. Krijthe BP, Kunst A, Benjamin EJ, Lip GYH, Franco OH, Hofman A, et al. Projections on the number of individuals with atrial fibrillation in the European Union, from 2000 to 2060. *Eur Heart J*. 2013 Sep;34(35):2746–51.
61. Miyasaka Y, Barnes ME, Gersh BJ, Cha SS, Bailey KR, Abhayaratna WP, et al. Secular trends in incidence of atrial fibrillation in Olmsted County, Minnesota, 1980 to 2000, and implications on the projections for future prevalence. *Circulation*. 2006 Jul 11;114(2):119–25.
62. Schnabel RB, Yin X, Gona P, Larson MG, Beiser AS, McManus DD, et al. 50 year trends in atrial fibrillation prevalence, incidence, risk factors, and mortality in the Framingham Heart Study: a cohort study. *Lancet*. 2015 Jul 11;386(9989):154–62.
63. Hanchate AD, Schwamm LH, Huang W, Hylek EM. Comparison of ischemic stroke outcomes and patient and hospital characteristics by race/ethnicity and socioeconomic status. *Stroke*. 2013 Feb;44(2):469–76.
64. Rizos T, Wagner A, Jenetzky E, Ringleb PA, Becker R, Hacke W, et al. Paroxysmal atrial fibrillation is more prevalent than persistent atrial fibrillation in acute stroke and transient ischemic attack patients. *Cerebrovasc Dis*. 2011;32(3):276–82.

65. Marini C, De Santis F, Sacco S, Russo T, Olivieri L, Totaro R, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study. *Stroke*. 2005 Jun;36(6):1115–9.
66. Yang Q. Vital Signs: Recent Trends in Stroke Death Rates — United States, 2000–2015. *MMWR Morb Mortal Wkly Rep* [Internet]. 2017 [cited 2020 May 3];66. Available from: <https://www.cdc.gov/mmwr/volumes/66/wr/mm6635e1.htm>
67. Kim MH, Johnston SS, Chu B-C, Dalal MR, Schulman KL. Estimation of total incremental health care costs in patients with atrial fibrillation in the United States. *Circ Cardiovasc Qual Outcomes*. 2011 May;4(3):313–20.
68. Stewart S, Murphy NF, Murphy N, Walker A, McGuire A, McMurray JJV. Cost of an emerging epidemic: an economic analysis of atrial fibrillation in the UK. *Heart*. 2004 Mar;90(3):286–92.
69. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, et al. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation*. 2017 07;135(10):e146–603.
70. Cotté F-E, Chaize G, Kachaner I, Gaudin A-F, Vainchtock A, Durand-Zaleski I. Incidence and cost of stroke and hemorrhage in patients diagnosed with atrial fibrillation in France. *J Stroke Cerebrovasc Dis*. 2014 Feb;23(2):e73-83.
71. Li T, Chang C-Y, Jin D-Y, Lin P-J, Khvorova A, Stafford DW. Identification of the gene for vitamin K epoxide reductase. *Nature*. 2004 Feb 5;427(6974):541–4.
72. Holford NHG. Clinical Pharmacokinetics and Pharmacodynamics of Warfarin. *Clin-Pharmacokinet* [Internet]. 1986 Dec 1 [cited 2020 May 31];11(6):483–504. Available from: <https://doi.org/10.2165/00003088-198611060-00005>
73. Landefeld CS, Beyth RJ. Anticoagulant-related bleeding: clinical epidemiology, prediction, and prevention. *Am J Med*. 1993 Sep;95(3):315–28.
74. Lang D, Freudenberger C, Weinz C. In vitro metabolism of rivaroxaban, an oral, direct factor Xa inhibitor, in liver microsomes and hepatocytes of rats, dogs, and humans. *Drug Metab Dispos*. 2009 May;37(5):1046–55.
75. Pollack CV, Reilly PA, van Ryn J, Eikelboom JW, Glund S, Bernstein RA, et al. Idarucizumab for Dabigatran Reversal — Full Cohort Analysis. *New England Journal of Medicine* [Internet]. 2017 Aug 3 [cited 2020 May 31];377(5):431–41. Available from: <https://doi.org/10.1056/NEJMoa1707278>
76. Schulman S, Kearon C, Subcommittee on Control of Anticoagulation of the Scientific and Standardization Committee of the International Society on Thrombosis and Haemostasis. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. *J Thromb Haemost*. 2005 Apr;3(4):692–4.
77. Schulman S, Kearon C, Kakkar AK, Mismetti P, Schellong S, Eriksson H, et al. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. *N Engl J Med*. 2009 Dec 10;361(24):2342–52.

78. Schulman S, Kakkar AK, Goldhaber SZ, Schellong S, Eriksson H, Mismetti P, et al. Treatment of acute venous thromboembolism with dabigatran or warfarin and pooled analysis. *Circulation*. 2014 Feb 18;129(7):764–72.
79. EINSTEIN Investigators, Bauersachs R, Berkowitz SD, Brenner B, Buller HR, Decousus H, et al. Oral rivaroxaban for symptomatic venous thromboembolism. *N Engl J Med*. 2010 Dec 23;363(26):2499–510.
80. EINSTEIN-PE Investigators, Büller HR, Prins MH, Lensin AWA, Decousus H, Jacobson BF, et al. Oral rivaroxaban for the treatment of symptomatic pulmonary embolism. *N Engl J Med*. 2012 Apr 5;366(14):1287–97.
81. Agnelli G, Buller HR, Cohen A, Curto M, Gallus AS, Johnson M, et al. Oral apixaban for the treatment of acute venous thromboembolism. *N Engl J Med*. 2013 Aug 29;369(9):799–808.
82. Hokusai-VTE Investigators, Büller HR, Décousus H, Grosso MA, Mercuri M, Middeldorp S, et al. Edoxaban versus warfarin for the treatment of symptomatic venous thromboembolism. *N Engl J Med*. 2013 Oct 10;369(15):1406–15.
83. Gómez-Outes A, Terleira-Fernández AI, Lecumberri R, Suárez-Gea ML, Vargas-Castrillón E. Direct oral anticoagulants in the treatment of acute venous thromboembolism: A systematic review and meta-analysis. *Thrombosis Research* [Internet]. 2014 Oct 1 [cited 2020 Jun 17];134(4):774–82. Available from: <http://www.sciencedirect.com/science/article/pii/S0049384814003466>
84. Hulle T van der, Kooiman J, Exter PL den, Dekkers OM, Klok FA, Huisman MV. Effectiveness and safety of novel oral anticoagulants as compared with vitamin K antagonists in the treatment of acute symptomatic venous thromboembolism: a systematic review and meta-analysis. *Journal of Thrombosis and Haemostasis* [Internet]. 2014 [cited 2020 Jun 18];12(3):320–8. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/jth.12485>
85. Di Minno MND, Ambrosino P, Lupoli R, Di Minno A, Dentali F. Direct oral anticoagulants for the treatment of unprovoked venous thromboembolism: a meta-analysis of randomised controlled trials. *Blood Transfus* [Internet]. 2015 Jul [cited 2020 Jun 18];13(3):391–5. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4614289/>
86. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, et al. Dabigatran versus Warfarin in Patients with Atrial Fibrillation. *N Engl J Med* [Internet]. 2009 Sep 17 [cited 2019 Oct 17];361(12):1139–51. Available from: <http://www.nejm.org/doi/abs/10.1056/NEJMoa0905561>
87. Hankey GJ, Patel MR, Stevens SR, Becker RC, Breithardt G, Carolei A, et al. Rivaroxaban compared with warfarin in patients with atrial fibrillation and previous stroke or transient ischaemic attack: a subgroup analysis of ROCKET AF. *Lancet Neurol*. 2012 Apr;11(4):315–22.
88. Granger CB, Alexander JH, McMurray JJV, Lopes RD, Hylek EM, Hanna M, et al. Apixaban versus Warfarin in Patients with Atrial Fibrillation. *N Engl J Med* [Internet]. 2011 Sep 15 [cited 2019 Oct 17];365(11):981–92. Available from: <http://www.nejm.org/doi/10.1056/NEJMoa1107039>

89. Giugliano RP, Ruff CT, Braunwald E, Murphy SA, Wiviott SD, Halperin JL, et al. Edoxaban versus Warfarin in Patients with Atrial Fibrillation. *New England Journal of Medicine* [Internet]. 2013 Nov 28 [cited 2020 Jun 18];369(22):2093–104. Available from: <https://doi.org/10.1056/NEJMoa1310907>
90. Ruff CT, Giugliano RP, Braunwald E, Hoffman EB, Deenadayalu N, Ezekowitz MD, et al. Comparison of the efficacy and safety of new oral anticoagulants with warfarin in patients with atrial fibrillation: a meta-analysis of randomised trials. *The Lancet* [Internet]. 2014 Mar 15 [cited 2020 Jun 21];383(9921):955–62. Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(13\)62343-0/abstract](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(13)62343-0/abstract)
91. Lega J-C, Bertolotti L, Gremillet C, Chapelle C, Mismetti P, Cucherat M, et al. Consistency of safety and efficacy of new oral anticoagulants across subgroups of patients with atrial fibrillation. *PLoS ONE*. 2014;9(3):e91398.
92. Nunes JPL, Rodrigues RP, Gonçalves FR. Comparative analysis and meta-analysis of major clinical trials with oral factor Xa inhibitors versus warfarin in atrial fibrillation. *Open Heart*. 2014;1(1):e000080.
93. Hicks T, Stewart F, Eisinga A. NOACs versus warfarin for stroke prevention in patients with AF: a systematic review and meta-analysis. *Open Heart*. 2016;3(1):e000279.
94. Potpara TS, Lip GYH. Postapproval Observational Studies of Non-Vitamin K Antagonist Oral Anticoagulants in Atrial Fibrillation. *JAMA* [Internet]. 2017 [cited 2020 Jun 22];317(11):1115–6. Available from: <https://vbn.aau.dk/en/publications/postapproval-observational-studies-of-non-vitamin-k-antagonist-or>
95. Camm AJ, Fox KAA. Strengths and weaknesses of “real-world” studies involving non-vitamin K antagonist oral anticoagulants. *Open Heart*. 2018;5(1):e000788.
96. Woolf B. On estimating the relation between blood group and disease. *Ann Hum Genet*. 1955 Jun;19(4):251–3.
97. Cucherat M. La méta-analyse des essais thérapeutiques. 2000;3.
98. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ, et al. Users’ Guides to the Medical Literature: IX. A Method for Grading Health Care Recommendations. *JAMA* [Internet]. 1995 Dec 13 [cited 2020 Mar 6];274(22):1800–4. Available from: <https://jamanetwork.com/journals/jama/fullarticle/392650>
99. Laird NM, Mosteller F. Some Statistical Methods for Combining Experimental Results. *International Journal of Technology Assessment in Health Care* [Internet]. 1990 Jan [cited 2020 Jun 30];6(1):5–30. Available from: <https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/some-statistical-methods-for-combining-experimental-results/3167EA380C6BE7F19C8DD298E7403C7E>
100. Borenstein M, Higgins JPT. Meta-Analysis and Subgroups. *Prev Sci* [Internet]. 2013 Apr 1 [cited 2020 Mar 9];14(2):134–43. Available from: <https://doi.org/10.1007/s11121-013-0377-7>
101. Fleiss JL. The statistical basis of meta-analysis. *Stat Methods Med Res*. 1993;2(2):121–45.

102. Jones DR. Meta-analysis: weighing the evidence. *Stat Med*. 1995 Jan 30;14(2):137–49.
103. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* [Internet]. 2002 [cited 2020 Mar 9];21(11):1539–58. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.1186>
104. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* [Internet]. 2003 Sep 6 [cited 2020 Mar 9];327(7414):557–60. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC192859/>
105. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Statistics in Medicine* [Internet]. 2008 [cited 2020 Feb 19];27(5):625–50. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2934>
106. Rosenthal R. The file drawer problem and tolerance for null results. *Psychological Bulletin* [Internet]. 1979 [cited 2020 Mar 13];86(3):638–41. Available from: <http://content.apa.org/journals/bul/86/3/638>
107. Kicinski M, Springate DA, Kontopantelis E. Publication bias in meta-analyses from the Cochrane Database of Systematic Reviews: M. KICINSKI, D. A. SPRINGATE AND E. KONTOPANTELLIS. *Statist Med* [Internet]. 2015 Sep 10 [cited 2020 Mar 13];34(20):2781–93. Available from: <http://doi.wiley.com/10.1002/sim.6525>
108. Begg CB, Berlin JA. Publication Bias: A Problem in Interpreting Medical Data. *Journal of the Royal Statistical Society Series A (Statistics in Society)* [Internet]. 1988 [cited 2020 Mar 13];151(3):419. Available from: <https://www.jstor.org/stable/10.2307/2982993?origin=crossref>
109. Sutton AJ, Song F, Gilbody SM, Abrams KR. Modelling publication bias in meta-analysis: a review. *Stat Methods Med Res*. 2000 Oct;9(5):421–45.
110. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet*. 2000 Oct 7;356(9237):1228–31.
111. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011 Jul 22;343:d4002.
112. Sterne JAC, Egger M. Regression Methods to Detect Publication and Other Bias in Meta-Analysis. In: Rothstein HR, Sutton AJ, Borenstein M, editors. *Publication Bias in Meta-Analysis* [Internet]. Chichester, UK: John Wiley & Sons, Ltd; 2006 [cited 2020 Mar 13]. p. 99–110. Available from: <http://doi.wiley.com/10.1002/0470870168.ch6>
113. Laporte S, Chapelle C, Trone J-C, Bertoletti L, Girard P, Meyer G, et al. Early detection of the existence or absence of the treatment effect: A cumulative meta-analysis. *Journal of Clinical Epidemiology* [Internet]. 2020 Aug 1 [cited 2020 Jun 30];124:24–33. Available from: <http://www.sciencedirect.com/science/article/pii/S089543561931056X>
114. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in ‘meta-epidemiological’ research: TREATMENT EFFECTS IN ‘META-EPIDEMIOLOGICAL’

RESEARCH. *Statist Med* [Internet]. 2002 Jun 15 [cited 2019 Dec 10];21(11):1513–24. Available from: <http://doi.wiley.com/10.1002/sim.1184>

115. Zhang W. I-01 META-EPIDEMIOLOGY: BUILDING THE BRIDGE FROM RESEARCH EVIDENCE TO CLINICAL PRACTICE. *Osteoarthritis and Cartilage* [Internet]. 2010 Oct [cited 2020 Apr 10];18:S1. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1063458410002931>

116. Clarke M. The Cochrane Database of Systematic Reviews – Issue 4 2008. *Journal of Evidence-Based Medicine* [Internet]. 2008 [cited 2020 Apr 26];1(1):12–12. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-5391.2008.00009.x>

117. Moher D, Pham B, Jones A, Cook DJ, Jadad AR, Moher M, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998 Aug 22;352(9128):609–13.

118. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995 Feb 1;273(5):408–12.

119. Siersma V, Als-Nielsen B, Chen W, Hilden J, Gluud LL, Gluud C. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Statist Med* [Internet]. 2007 Jun 30 [cited 2019 Dec 10];26(14):2745–58. Available from: <http://doi.wiley.com/10.1002/sim.2752>

120. Courgeau D. Goldstein (H.) — Multilevel Statistical Models [Internet]. Vol. 52. Institut national d'études démographiques; 1997 [cited 2020 Apr 23]. Available from: [https://www.persee.fr/doc/pop\\_0032-4663\\_1997\\_num\\_52\\_4\\_6480](https://www.persee.fr/doc/pop_0032-4663_1997_num_52_4_6480)

121. Altman DG. Statistics Notes: Interaction revisited: the difference between two estimates. *BMJ* [Internet]. 2003 Jan 25 [cited 2019 Dec 10];326(7382):219–219. Available from: <http://www.bmj.com/cgi/doi/10.1136/bmj.326.7382.219>

122. Savović J, Jones H, Altman D, Harris R, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess*. 2012 Sep;16(35):1–82.

123. KJAERGARD L. quality of randomised clinical trials affects estimates of intervention efficacy. Proceedings of the 7th Cochrane colloquium, Università St Tommaso D'Aquino Rome [Internet]. 1999 [cited 2020 Aug 13];57. Available from: <https://ci.nii.ac.jp/naid/10011527240/>

124. Juni P, Tallon D, Egger M. Garbage in-garbage out? Assessment of the quality of controlled trials in meta-analyses published in leading journals. In: Proceedings of the 3rd symposium on systematic reviews: beyond the basics St Catherine's College Oxford: Centre for Statistics in Medicine. 2000.

125. Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ*. 2001 Jul 7;323(7303):42–6.

126. Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported

study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med.* 2012 Sep 18;157(6):429–38.

127. Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. *Ann Intern Med.* 2011 Jul 5;155(1):39–51.

128. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. *BMJ.* 2013 Apr 24;346:f2304.

129. Janiaud P, Cristea I-A, Ioannidis JPA. Industry-funded versus non-profit-funded critical care research: a meta-epidemiological overview. *Intensive Care Med* [Internet]. 2018 [cited 2020 Jul 9];44(10):1613–27. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6182357/>

130. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews* [Internet]. 2017 [cited 2020 Jul 9];(2). Available from: <https://www.readcube.com/articles/10.1002%2F14651858.MR000033.pub3>

131. Lexchin J. Sponsorship bias in clinical research. *Int J Risk Saf Med.* 2012;24(4):233–42.

132. Leyrat C, Caille A, Eldridge S, Kerry S, Dechartres A, Giraudeau B. Intervention effect estimates in cluster randomized versus individually randomized trials: a meta-epidemiological study. *Int J Epidemiol.* 2019 01;48(2):609–19.

133. Wallach JD, Ciani O, Pease AM, Gonsalves GS, Krumholz HM, Taylor RS, et al. Comparison of treatment effect sizes from pivotal and postapproval trials of novel therapeutics approved by the FDA based on surrogate markers of disease: a meta-epidemiological study. *BMC Med.* 2018 21;16(1):45.

134. Zhang Z, Ni H, Xu X. Do the observational studies using propensity score analysis agree with randomized controlled trials in the area of sepsis? *J Crit Care.* 2014 Oct;29(5):886.e9-15.

135. Dahabreh IJ, Sheldrick RC, Paulus JK, Chung M, Varvarigou V, Jafri H, et al. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. *Eur Heart J.* 2012 Aug;33(15):1893–901.

136. Zhang Z, Ni H, Xu X. Observational studies using propensity score analysis underestimated the effect sizes in critical care medicine. *J Clin Epidemiol.* 2014 Aug;67(8):932–9.

137. Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg.* 2014 Jan;259(1):18–25.

138. Li G, Holbrook A, Jin Y, Zhang Y, Levine MAH, Mbuagbaw L, et al. Comparison of treatment effect estimates of non-vitamin K antagonist oral anticoagulants versus warfarin between observational studies using propensity score methods and randomized controlled trials. *Eur J Epidemiol.* 2016;31(6):541–61.

139. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet*. 1999 Nov 27;354(9193):1896–900.
140. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Med* [Internet]. 2009 Jul 21 [cited 2020 Apr 7];6(7). Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2707010/>
141. Schiavo JH. PROSPERO: An International Register of Systematic Review Protocols. *Med Ref Serv Q*. 2019 Jun;38(2):171–80.
142. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine*. 2009;6(7):6.
143. Murad MH, Wang Z. Guidelines for reporting meta-epidemiological methodology research. *Evid Based Med* [Internet]. 2017 Aug [cited 2019 Dec 9];22(4):139–42. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5537553/>
144. Sindet-Pedersen C, Langtved Pallisgaard J, Staerk L, Gerds TA, Fosbøl EL, Torp-Pedersen C, et al. Comparative safety and effectiveness of rivaroxaban versus VKAs in patients with venous thromboembolism. A Danish nationwide registry-based study. *Thromb Haemost*. 2017 02;117(6):1182–91.
145. Larsen TB, Skjøth F, Kjældgaard JN, Lip GYH, Nielsen PB, Søgaard M. Effectiveness and safety of rivaroxaban and warfarin in patients with unprovoked venous thromboembolism: a propensity-matched nationwide cohort study. *Lancet Haematol*. 2017 May;4(5):e237–44.
146. Coleman CI, Peacock WF, Bunz TJ, Beyer-Westendorf J. Effectiveness and safety of rivaroxaban versus warfarin in patients with unprovoked venous thromboembolism: A propensity-score weighted administrative claims cohort study. *Thromb Res*. 2018;168:31–6.
147. Coleman CI, Turpie AGG, Bunz TJ, Beyer-Westendorf J. Effectiveness and safety of rivaroxaban versus warfarin in patients with provoked venous thromboembolism. *J Thromb Thrombolysis*. 2018 Oct;46(3):339–45.
148. Coleman CI, Turpie AGG, Bunz TJ, Baker WL, Beyer-Westendorf J. Effectiveness and safety of outpatient rivaroxaban versus warfarin for treatment of venous thromboembolism in patients with a known primary hypercoagulable state. *Thromb Res*. 2018;163:132–7.
149. Kohn CG, Bunz TJ, Beyer-Westendorf J, Coleman CI. Comparative risk of major bleeding with rivaroxaban and warfarin: Population-based cohort study of unprovoked venous thromboembolism. *Eur J Haematol*. 2019 Feb;102(2):143–9.
150. Coleman CI, Bunz TJ, Turpie AGG. Effectiveness and safety of rivaroxaban versus warfarin for treatment and prevention of recurrence of venous thromboembolism. *Thromb Haemost*. 2017 05;117(10):1841–7.
151. Wang L, Baser O, Wells P, Peacock WF, Coleman CI, Ferrmann GJ, et al. Overall Effectiveness of Rivaroxaban in Patients with Pulmonary Embolism. *Clinical Therapeutics*. 2017 Jul;39(7):1426-1436.e2.

152. Gaertner S, Cordeanu E-M, Nouri S, Faller A-M, Frantz A-S, Mirea C, et al. Rivaroxaban versus standard anticoagulation for symptomatic venous thromboembolism (REMOTEV observational study): Analysis of 6-month outcomes. *Int J Cardiol.* 2017 Jan 1;226:103–9.
153. Costa OS, Thompson S, Ashton V, Palladino M, Bunz TJ, Coleman CI. Rivaroxaban versus warfarin for treatment and prevention of recurrence of venous thromboembolism in African American patients: a retrospective cohort analysis. *Thromb J.* 2020;18:6.
154. Jun M, Lix LM, Durand M, Dahl M, Paterson JM, Dormuth CR, et al. Comparative safety of direct oral anticoagulants and warfarin in venous thromboembolism: multicentre, population based, observational study. *BMJ.* 2017 Oct 17;359:j4323.
155. López-Núñez JJ, Pérez-Andrés R, Di Micco P, Schellong S, Gómez-Cuervo C, Sahuquillo JC, et al. Direct Oral Anticoagulants or Standard Anticoagulant Therapy in Fragile Patients with Venous Thromboembolism. *TH Open.* 2019 Jan;3(1):e67–76.
156. Wakakura S, Hara F, Fujino T, Hamai A, Ohara H, Kabuki T, et al. Comparison of Direct Oral Anticoagulants and Warfarin in the Treatment of Deep Venous Thrombosis in the Chronic Phase. *International Heart Journal.* 2018 Jan 27;59(1):126–35.
157. Badreldin H. Hospital length of stay in patients initiated on direct oral anticoagulants versus warfarin for venous thromboembolism: a real-world single-center study. *Journal of Thrombosis and Thrombolysis.* 2018 Jul;46(1):16–21.
158. Basto AN, Fewel NP, Vo K, Stock EM, Ta M. Initiation of direct oral anticoagulants versus warfarin for venous thromboembolism: impact on time to hospital discharge. *Journal of Thrombosis and Thrombolysis.* 2018 Jan;45(1):51–5.
159. Fung K-P, Chan K-H, Ng V, Tsui P-T, You JHS. Health Economic Analysis of Rivaroxaban and Warfarin for Venous Thromboembolism Management in Chinese Patients. *Cardiovascular Drugs and Therapy.* 2019;33(3):331–7.
160. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ [Internet].* 2003 Sep 6 [cited 2020 May 18];327(7414):557–60. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC192859/>
161. Deeks JJ, Altman DG, Bradburn MJ. Statistical Methods for Examining Heterogeneity and Combining Results from Several Studies in Meta-Analysis. In: *Systematic Reviews in Health Care [Internet].* John Wiley & Sons, Ltd; 2008 [cited 2020 Aug 12]. p. 285–312. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470693926.ch15>
162. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials.* 1986 Sep;7(3):177–88.
163. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997 Sep 13;315(7109):629–34.
164. Altman DG, Bland JM. Interaction revisited: the difference between two estimates. *BMJ [Internet].* 2003 Jan 25 [cited 2020 Apr 16];326(7382):219. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1125071/>

165. Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in “meta-epidemiological” research. *Stat Med*. 2002 Jun 15;21(11):1513–24.
166. Schwarzer G, Carpenter JR, Rücker G. *Meta-Analysis with R* [Internet]. Springer International Publishing; 2015 [cited 2020 Aug 12]. (Use R!). Available from: <https://www.springer.com/gp/book/9783319214153>
167. Kreutz R, Mantovani LG, Haas S, Monje D, Schneider J, Bugge J-P, et al. XALIA-LEA: An observational study of venous thromboembolism treatment with rivaroxaban and standard anticoagulation in the Asia-Pacific, Eastern Europe, the Middle East, Africa and Latin America. *Thromb Res*. 2019 Apr;176:125–32.
168. Bui MH, Son NT, Viet PT, Hiep NH, Dinh TC. Oral Rivaroxaban Versus Standard Therapy in Acute Venous Thromboembolism Treatment for Vietnamese Patients. *Open Access Maced J Med Sci*. 2019 Dec 30;7(24):4255–9.
169. Kilcher G, Hummel N, Didden EM, Egger M, Reichenbach S. Rheumatoid arthritis patients treated in trial and real world settings: comparison of randomized trials with registries. *Rheumatology (Oxford)* [Internet]. 2018 Feb 1 [cited 2020 May 18];57(2):354–69. Available from: <https://academic.oup.com/rheumatology/article/57/2/354/4629381>
170. Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg*. 2014 Jan;259(1):18–25.
171. Li G, Holbrook A, Jin Y, Zhang Y, Levine MAH, Mbuagbaw L, et al. Comparison of treatment effect estimates of non-vitamin K antagonist oral anticoagulants versus warfarin between observational studies using propensity score methods and randomized controlled trials. *Eur J Epidemiol*. 2016;31(6):541–61.
172. Kuss O, Legler T, Börgermann J. Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. *J Clin Epidemiol*. 2011 Oct;64(10):1076–84.
173. Coleman CI, Antz M, Bowrin K, Evers T, Simard EP, Bonnemeier H, et al. Real-world evidence of stroke prevention in patients with nonvalvular atrial fibrillation in the United States: the REVISIT-US study. *Curr Med Res Opin*. 2016;32(12):2047–53.
174. Abraham NS, Singh S, Alexander GC, Heien H, Haas LR, Crown W, et al. Comparative risk of gastrointestinal bleeding with dabigatran, rivaroxaban, and warfarin: population based cohort study. *BMJ*. 2015 Apr 24;350:h1857.
175. Graham DJ, Reichman ME, Wernecke M, Zhang R, Southworth MR, Levenson M, et al. Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation. *Circulation*. 2015 Jan 13;131(2):157–64.
176. Hernandez I, Baik SH, Piñera A, Zhang Y. Risk of bleeding with dabigatran in atrial fibrillation. *JAMA Intern Med*. 2015 Jan;175(1):18–24.

177. Bengtson LGS, Lutsey PL, Chen LY, MacLehose RF, Alonso A. Comparative effectiveness of dabigatran and rivaroxaban versus warfarin for the treatment of non-valvular atrial fibrillation. *J Cardiol*. 2017 Jun;69(6):868–76.
178. Lauffenburger JC, Farley JF, Gehi AK, Rhoney DH, Brookhart MA, Fang G. Effectiveness and safety of dabigatran and warfarin in real-world US patients with non-valvular atrial fibrillation: a retrospective cohort study. *J Am Heart Assoc*. 2015 Apr 10;4(4).
179. Norby FL, Bengtson LGS, Lutsey PL, Chen LY, MacLehose RF, Chamberlain AM, et al. Comparative effectiveness of rivaroxaban versus warfarin or dabigatran for the treatment of patients with non-valvular atrial fibrillation. *BMC Cardiovasc Disord*. 2017 06;17(1):238.
180. Li XS, Deitelzweig S, Keshishian A, Hamilton M, Horblyuk R, Gupta K, et al. Effectiveness and safety of apixaban versus warfarin in non-valvular atrial fibrillation patients in “real-world” clinical practice. A propensity-matched analysis of 76,940 patients. *Thromb Haemost*. 2017 02;117(6):1072–82.
181. Lip GYH, Keshishian A, Kamble S, Pan X, Mardekian J, Horblyuk R, et al. Real-world comparison of major bleeding risk among non-valvular atrial fibrillation patients initiated on apixaban, dabigatran, rivaroxaban, or warfarin. A propensity score matched analysis. *Thromb Haemost*. 2016 Oct 28;116(5):975–86.
182. Amin A, Keshishian A, Trocio J, Dina O, Le H, Rosenblatt L, et al. Risk of stroke/systemic embolism, major bleeding and associated costs in non-valvular atrial fibrillation patients who initiated apixaban, dabigatran or rivaroxaban compared with warfarin in the United States Medicare population. *Curr Med Res Opin*. 2017;33(9):1595–604.
183. Seeger JD, Bykov K, Bartels DB, Huybrechts K, Zint K, Schneeweiss S. Safety and effectiveness of dabigatran and warfarin in routine care of patients with atrial fibrillation. *Thromb Haemost*. 2015 Nov 25;114(6):1277–89.
184. Lip GYH, Keshishian A, Li X, Hamilton M, Masseria C, Gupta K, et al. Effectiveness and Safety of Oral Anticoagulants Among Nonvalvular Atrial Fibrillation Patients. *Stroke*. 2018;49(12):2933–44.
185. Coleman CI, Peacock WF, Bunz TJ, Alberts MJ. Effectiveness and Safety of Apixaban, Dabigatran, and Rivaroxaban Versus Warfarin in Patients With Nonvalvular Atrial Fibrillation and Previous Stroke or Transient Ischemic Attack. *Stroke*. 2017;48(8):2142–9.
186. Vaughan Sarrazin MS, Jones M, Mazur A, Chrischilles E, Cram P. Bleeding rates in Veterans Affairs patients with atrial fibrillation who switch from warfarin to dabigatran. *Am J Med*. 2014 Dec;127(12):1179–85.
187. Villines TC, Schnee J, Fraeman K, Siu K, Reynolds MW, Collins J, et al. A comparison of the safety and effectiveness of dabigatran and warfarin in non-valvular atrial fibrillation patients in a large healthcare system. *Thromb Haemost*. 2015 Nov 25;114(6):1290–8.
188. Go AS, Singer DE, Toh S, Cheetham TC, Reichman ME, Graham DJ, et al. Outcomes of Dabigatran and Warfarin for Atrial Fibrillation in Contemporary Practice: A Retrospective Cohort Study. *Ann Intern Med*. 2017 Dec 19;167(12):845–54.
189. Adeboyeje G, Sylwestrzak G, Barron JJ, White J, Rosenberg A, Abarca J, et al. Major

Bleeding Risk During Anticoagulation with Warfarin, Dabigatran, Apixaban, or Rivaroxaban in Patients with Nonvalvular Atrial Fibrillation. *J Manag Care Spec Pharm*. 2017 Sep;23(9):968–78.

190. Laliberté F, Cloutier M, Nelson WW, Coleman CI, Pilon D, Olson WH, et al. Real-world comparative effectiveness and safety of rivaroxaban and warfarin in nonvalvular atrial fibrillation patients. *Curr Med Res Opin*. 2014 Jul;30(7):1317–25.

191. Nielsen PB, Skjøth F, Søgaard M, Kjældgaard JN, Lip GYH, Larsen TB. Effectiveness and safety of reduced dose non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study. *BMJ* [Internet]. 2017 Feb 10 [cited 2020 Jul 30];356. Available from: <https://www.bmj.com/content/356/bmj.j510>

192. Larsen TB, Skjøth F, Nielsen PB, Kjældgaard JN, Lip GYH. Comparative effectiveness and safety of non-vitamin K antagonist oral anticoagulants and warfarin in patients with atrial fibrillation: propensity weighted nationwide cohort study. *BMJ*. 2016 Jun 16;353:i3189.

193. Gorst-Rasmussen A, Lip GYH, Bjerregaard Larsen T. Rivaroxaban versus warfarin and dabigatran in atrial fibrillation: comparative effectiveness and safety in Danish routine care. *Pharmacoepidemiol Drug Saf*. 2016;25(11):1236–44.

194. Larsen TB, Rasmussen LH, Skjøth F, Due KM, Callréus T, Rosenzweig M, et al. Efficacy and safety of dabigatran etexilate and warfarin in “real-world” patients with atrial fibrillation: a prospective nationwide cohort study. *J Am Coll Cardiol*. 2013 Jun 4;61(22):2264–73.

195. Sørensen R, Gislason G, Torp-Pedersen C, Olesen JB, Fosbøl EL, Hvidtfeldt MW, et al. Dabigatran use in Danish atrial fibrillation patients in 2011: a nationwide study. *BMJ Open*. 2013 May 3;3(5).

196. Staerk L, Gislason GH, Lip GYH, Fosbøl EL, Hansen ML, Lamberts M, et al. Risk of gastrointestinal adverse effects of dabigatran compared with warfarin among patients with atrial fibrillation: a nationwide cohort study. *Europace*. 2015 Aug;17(8):1215–22.

197. Larsen TB, Gorst-Rasmussen A, Rasmussen LH, Skjøth F, Rosenzweig M, Lip GYH. Bleeding events among new starters and switchers to dabigatran compared with warfarin in atrial fibrillation. *Am J Med*. 2014 Jul;127(7):650-656.e5.

198. Larsen TB, Rasmussen LH, Gorst-Rasmussen A, Skjøth F, Lane DA, Lip GYH. Dabigatran and warfarin for secondary prevention of stroke in atrial fibrillation patients: a nationwide cohort study. *Am J Med*. 2014 Dec;127(12):1172-1178.e5.

199. NACORA\_CNAMTSjuillet2014.pdf [Internet]. [cited 2020 Jul 30]. Available from: [https://www.snds.gouv.fr/download/Epidemio/NACORA\\_CNAMTSjuillet2014.pdf](https://www.snds.gouv.fr/download/Epidemio/NACORA_CNAMTSjuillet2014.pdf)

200. Maura G, Blotière P-O, Bouillon K, Billionnet C, Ricordeau P, Alla F, et al. Comparison of the short-term risk of bleeding and arterial thromboembolic events in nonvalvular atrial fibrillation patients newly treated with dabigatran or rivaroxaban versus vitamin K antagonists: a French nationwide propensity-matched cohort study. *Circulation*. 2015 Sep 29;132(13):1252–60.

201. Bouillon K, Bertrand M, Maura G, Blotière P-O, Ricordeau P, Zureik M. Risk of bleeding and arterial thromboembolism in patients with non-valvular atrial fibrillation either maintained on a vitamin K antagonist or switched to a non-vitamin K-antagonist oral anticoagulant: a retrospective, matched-cohort study. *Lancet Haematol*. 2015 Apr;2(4):e150-159.
202. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med*. 2011 Sep 8;365(10):883–91.
203. Granger CB, Alexander JH, McMurray JJV, Lopes RD, Hylek EM, Hanna M, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011 Sep 15;365(11):981–92.
204. Yavuz B, Ayturk M, Ozkan S, Ozturk M, Topaloglu C, Aksoy H, et al. A real world data of dabigatran etexilate: multicenter registry of oral anticoagulants in nonvalvular atrial fibrillation. *J Thromb Thrombolysis*. 2016 Oct;42(3):399–404.
205. Kodani E, Atarashi H, Inoue H, Okumura K, Yamashita T, Origasa H, et al. Beneficial Effect of Non-Vitamin K Antagonist Oral Anticoagulants in Patients With Nonvalvular Atrial Fibrillation - Results of the J-RHYTHM Registry 2. *Circ J*. 2016;80(4):843–51.
206. Ho JCS, Chang AM, Yan BP, Yu CM, Lam YY, Lee VWY. Dabigatran compared with warfarin for stroke prevention with atrial fibrillation: experience in Hong Kong. *Clin Cardiol*. 2012 Dec;35(12):E40-45.
207. Aslan O, Yaylali YT, Yildirim S, Yurtdas M, Senol H, Ugur-Yildiz M, et al. Dabigatran Versus Warfarin in Atrial Fibrillation: Multicenter Experience in Turkey. *Clin Appl Thromb Hemost*. 2016 Mar;22(2):147–52.
208. Yap LB, Eng DTS, Sivalingam L, Rusani BI, Umadevan D, Muhammad Z, et al. A Comparison of Dabigatran With Warfarin for Stroke Prevention in Atrial Fibrillation in an Asian Population. *Clin Appl Thromb Hemost*. 2016 Nov;22(8):792–7.
209. Korenstra J, Wijtvliet EPJ, Veeger NJGM, Geluk CA, Bartels GL, Posma JL, et al. Effectiveness and safety of dabigatran versus acenocoumarol in “real-world” patients with atrial fibrillation. *Europace*. 2016 Sep;18(9):1319–27.
210. Russo-Alvarez G, Martinez KA, Valente M, Bena J, Hu B, Luxenburg J, et al. Thromboembolic and Major Bleeding Events With Rivaroxaban Versus Warfarin Use in a Real-World Setting. *Ann Pharmacother*. 2018 Jan;52(1):19–25.
211. Jacobs V, May HT, Bair TL, Crandall BG, Cutler MJ, Day JD, et al. Long-Term Population-Based Cerebral Ischemic Event and Cognitive Outcomes of Direct Oral Anticoagulants Compared With Warfarin Among Long-term Anticoagulated Patients for Atrial Fibrillation. *Am J Cardiol*. 2016 Jul 15;118(2):210–4.
212. Chan Y-H, Kuo C-T, Yeh Y-H, Chang S-H, Wu L-S, Lee H-F, et al. Thromboembolic, Bleeding, and Mortality Risks of Rivaroxaban and Dabigatran in Asians With Nonvalvular Atrial Fibrillation. *J Am Coll Cardiol*. 2016 27;68(13):1389–401.
213. Norby FL, Bengtson LGS, Lutsey PL, Chen LY, MacLehose RF, Chamberlain AM, et al. Comparative effectiveness of rivaroxaban versus warfarin or dabigatran for the treatment of

patients with non-valvular atrial fibrillation. *BMC Cardiovasc Disord.* 2017 06;17(1):238.

214. Nishtala PS, Gnjjidic D, Jamieson HA, Hanger HC, Kaluarachchi C, Hilmer SN. “Real-world” haemorrhagic rates for warfarin and dabigatran using population-level data in New Zealand. *Int J Cardiol.* 2016 Jan 15;203:746–52.

215. Avgil-Tsadok M, Jackevicius CA, Essebag V, Eisenberg MJ, Rahme E, Behlouli H, et al. Dabigatran use in elderly patients with atrial fibrillation. *Thromb Haemost.* 2016 Jan;115(1):152–60.

216. Graham DJ, Reichman ME, Wernecke M, Zhang R, Southworth MR, Levenson M, et al. Cardiovascular, bleeding, and mortality risks in elderly Medicare patients treated with dabigatran or warfarin for nonvalvular atrial fibrillation. *Circulation.* 2015 Jan 13;131(2):157–64.

217. Larsen TB, Rasmussen LH, Gorst-Rasmussen A, Skjøth F, Lane DA, Lip GYH. Dabigatran and warfarin for secondary prevention of stroke in atrial fibrillation patients: a nationwide cohort study. *Am J Med.* 2014 Dec;127(12):1172-1178.e5.

218. Larsen TB, Gorst-Rasmussen A, Rasmussen LH, Skjøth F, Rosenzweig M, Lip GYH. Bleeding events among new starters and switchers to dabigatran compared with warfarin in atrial fibrillation. *Am J Med.* 2014 Jul;127(7):650-656.e5.

219. Yu AXY, Malo S, Svenson LW, Wilton SB, Hill MD. Temporal Trends in the Use and Comparative Effectiveness of Direct Oral Anticoagulant Agents Versus Warfarin for Nonvalvular Atrial Fibrillation: A Canadian Population-Based Study. *J Am Heart Assoc.* 2017 Oct 28;6(11).

220. Yao X, Abraham NS, Sangaralingham LR, Bellolio MF, McBane RD, Shah ND, et al. Effectiveness and Safety of Dabigatran, Rivaroxaban, and Apixaban Versus Warfarin in Nonvalvular Atrial Fibrillation. *J Am Heart Assoc.* 2016 13;5(6).

221. Sjögren V, Byström B, Renlund H, Svensson PJ, Oldgren J, Norrving B, et al. Non-vitamin K oral anticoagulants are non-inferior for stroke prevention but cause fewer major bleedings than well-managed warfarin: A retrospective register study. *PLoS ONE.* 2017;12(7):e0181000.

222. Forslund T, Wettermark B, Andersen M, Hjemdahl P. Stroke and bleeding with non-vitamin K antagonist oral anticoagulant or warfarin treatment in patients with non-valvular atrial fibrillation: a population-based cohort study. *Europace.* 2018 01;20(3):420–8.

223. Kohsaka S, Murata T, Izumi N, Katada J, Wang F, Terayama Y. Bleeding risk of apixaban, dabigatran, and low-dose rivaroxaban compared with warfarin in Japanese patients with non-valvular atrial fibrillation: a propensity matched analysis of administrative claims data. *Curr Med Res Opin.* 2017;33(11):1955–63.

224. Vinogradova Y, Coupland C, Hill T, Hippisley-Cox J. Risks and benefits of direct oral anticoagulants versus warfarin in a real world setting: cohort study in primary care. *BMJ.* 2018 04;362:k2505.

225. Ellis MH, Neuman T, Bitterman H, Dotan SG, Hammerman A, Battat E, et al. Bleeding in patients with atrial fibrillation treated with dabigatran, rivaroxaban or warfarin: A

retrospective population-based cohort study. *Eur J Intern Med.* 2016 Sep;33:55–9.

226. Bowman L, Mafham M, Stevens W, Haynes R, Aung T, Chen F, et al. ASCEND: A Study of Cardiovascular Events in Diabetes: Characteristics of a randomized trial of aspirin and of omega-3 fatty acid supplementation in 15,480 people with diabetes. *Am Heart J.* 2018;198:135–44.

227. Mues KE, Liede A, Liu J, Wetmore JB, Zaha R, Bradbury BD, et al. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. *Clin Epidemiol [Internet].* 2017 May 9 [cited 2020 Feb 3];9:267–77. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5433516/>

228. Chung H, Deshpande G, Zolotarjova J, Quimbo RA, Kern DM, Cochetti PT, et al. Health Plan Enrollment and Disenrollment of Individuals With and Without Established Chronic Disease in a U.S. Commercially Insured and Medicare Advantage Population. *JMCP [Internet].* 2019 Apr 30 [cited 2020 Feb 3];25(5):612–20. Available from: <https://www.jmcp.org/doi/10.18553/jmcp.2019.25.5.612>

229. Kulaylat AS, Schaefer EW, Messaris E, Hollenbeak CS. Truven Health Analytics MarketScan Databases for Clinical Research in Colon and Rectal Surgery. *Clin Colon Rectal Surg [Internet].* 2019 Jan [cited 2020 Feb 3];32(1):54–60. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6327721/>

230. Optum Labs: Building A Novel Node In The Learning Health Care System | Health Affairs [Internet]. [cited 2020 Feb 3]. Available from: [https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2014.0038?url\\_ver=Z39.88-2003&rfr\\_id=ori%3Arid%3Acrossref.org&rfr\\_dat=cr\\_pub%3Dpubmed&](https://www.healthaffairs.org/doi/full/10.1377/hlthaff.2014.0038?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub%3Dpubmed&)

231. Data Partners | Sentinel Initiative [Internet]. [cited 2020 Feb 3]. Available from: <https://www.sentinelinitiative.org/sentinel/data/data-partners>

## 7 Annexes

### **Annexe A1. Assurance de santé et bases de données de santé informatisées américaines**

Tableau A1: Bases de données américaine .....	147
Figure A1: Assurance de santé aux États-Unis en 2019 .....	150

### **Annexe 2. Tableaux et figures annexes de la partie 3 et 4**

Tableau A2: Lignes directrices pour les rapports de Prisma.....	151
Tableau A3 : Caractéristiques des patients dans les études incluses ( partie 3).....	153
Tableau A4: Équation de recherche utilisée dans PubMed (partie 3) .....	157
Tableau A5 : Caractéristiques des études incluses dans la partie 3.....	158
Tableau A6: Caractéristiques des études incluses dans la partie 4.....	163
Tableau A7: Équation de recherche utilisée dans PubMed (partie 4) .....	168
Figure A2: Comparaison des estimations des effets du traitement en fonction du type d'études sur les hémorragies intracrâniennes .....	160
Figure A3: Comparaison des estimations des effets du traitement en fonction du type d'études sur les hémorragies gastro-intestinales.....	161
Figure A4: RHR et IC à 95% pour quantifier la différence d'estimation des effets traitement entre les études observationnelles et les ECR pour les critères secondaires .....	162

<b>Annexe 3. Article soumis</b> .....	165
---------------------------------------	-----

## **Annexe 1. Assurance de santé et bases de données de santé informatisées américaines**

### A/ Les composantes publiques du système de santé américain

1. Le programme Medicare<sup>(227)</sup> est un programme d'assurance santé fédéral instauré en 1965. Il couvre les personnes âgées de  $\geq 65$  ans ainsi que les personnes  $< 65$  ans avec certaines incapacités. Ce programme se décline en plusieurs plans ou parties :

- La Partie A couvre les dépenses liées à une hospitalisation.
- La Partie B couvre les visites chez le médecin ainsi que d'autres services médicaux parmi lesquels les examens de dépistage du diabète, des maladies cardiaques ou encore de certains cancers.
- La Partie C («Medicare Advantage plan») est une couverture fournie par les compagnies d'assurances privées<sup>(228)</sup>. Elle constitue une couverture additionnelle à celle apportée par les parties A&B.
- La Partie D est également un plan privé destiné aux personnes ayant déjà souscrit les plans A&B. Elle couvre les coûts liés aux médicaments sur ordonnance. En 2014, environ 70 % des bénéficiaires de Medicare étaient inscrits au plan D.

2. Le programme Medicaid a été instauré en 1965, tout comme Medicare. Il s'agit d'un programme cofinancé par le gouvernement fédéral mais géré par chacun des États. Il est destiné aux individus et familles possédant de faibles revenus et aux personnes handicapées. Les différents États et le District de Columbia sont responsables de l'administration du programme Medicaid. La plupart des États administrent un programme Medicaid de façon personnalisée. Il existe 50 programmes Medicaid différents selon l'État. Selon la loi fédérale, les États ont l'obligation d'assurer une couverture de santé couvrant les consultations chez les médecins généralistes, les hospitalisations, les soins préventifs ou encore les soins liés à la santé mentale. Medicaid offre ainsi un ensemble assez complet de prestations, notamment des médicaments sur ordonnance. Ce programme prend ainsi en charge les actes de santé des femmes enceintes, des enfants, des personnes âgées, des handicapés et des parents très pauvres. Les conditions d'éligibilité au programme ne sont pas identiques selon les États, plus au moins restrictives. Elles sont basées sur les revenus des personnes. Les adultes sans enfants ne sont pas couverts.

Medicaid peut également offrir, selon l'État, une couverture qui n'est pas incluse dans Medicare, comme les soins d'optométrie et les soins dentaires.

3. Autres systèmes publics:

- Le programme d'assurance maladie pour enfants de l'État (S-CHIP) a été conçu en 1997 pour couvrir des enfants non assurés des familles dont les revenus sont relativement faibles mais trop élevés pour bénéficier de Medicaid. S-CHIP et Medicaid partagent souvent des structures administratives et financières communes.
- Le Veteran's administration (VA) est un programme administré au niveau fédéral pour les vétérans de l'armée. Les soins de santé sont assurés par l'État qui possède des hôpitaux et des cliniques de VA.

B/ Les assurances de santé privées aux États-Unis

1. L'assurance santé de l'employeur représente le principal moyen par lequel les citoyens américains bénéficient d'une assurance maladie. L'employeur prend à sa charge la plus grande partie des coûts et une cotisation est également prélevée sur le salaire de l'employé de façon similaire aux mutuelles d'entreprise en France. Les fournisseurs d'assurances santé privées se déclinent essentiellement en 3 catégories : les « Health Maintenance organizations », les « Preferred Provider Organizations » et les « Indemnity Insurances ». Les prestations « services » varient considérablement en fonction du régime d'assurance maladie concerné. Certains régimes couvrent les médicaments sur ordonnance, tandis que d'autres ne le font pas. Un cas particulier est représenté par les entreprises qui sont "auto-assurées" : le cas de la grande entreprise « General Motors ».

2. L'assurance à titre individuel, qui couvre certaines personnes qui ne peuvent pas bénéficier de l'assurance de leur employeur.

Les bases de données de santé américaines intégrant l'ensemble des données disponibles pour les recherches en matière de soins de santé sont détaillées dans le table A1 ci-dessous.

Tableau A1: Bases de données américaine

1- Market scan <sup>(229)</sup>	
A/ Bases de données principales	
<b>Commercial Claims and Encounters (CCAЕ)</b>	Contient l'expérience en soins de santé des assurés dans le secteur privé. Elle comprend les demandes de remboursement (en hospitalisation et en consultation externe) de frais médicaux/ pharmaceutiques et des frais de procédure.
<b>Medicare Supplemental</b>	Contient l'expérience en soins de santé des personnes dont l'assurance complémentaire Medicare est payée par les employeurs. La part couverte par Medicare (représentée sous forme de coordination des prestations) et la part payée par l'employeur sont incluses dans cette base de données.
<b>Multi-State Medicaid</b>	Reflète l'utilisation des services de soins par les personnes couvertes par les programmes Medicaid dans de nombreux état géographiquement répartis (11 États). La base de données contient l'expérience commune en soins de santé des adhérents Medicaid, couverts au titre des régimes de paiement à l'acte et d'assurance-maladie. Elle contient les dossiers des services d'hospitalisation, les admissions en hospitalisation, les services ambulatoires et les demandes de remboursement des médicaments sur ordonnance, ainsi que des informations sur les soins de longue durée. De plus des données démographiques supplémentaires sur l'ethnie et le statut du handicap sont aussi incluses dans cette base.
B/ Bases de données supplémentaires	
<b>Health and productivity management</b>	Intègre des données sur l'absence au travail (arrêt ou vacance), l'indemnisation des travailleurs et l'invalidité à court et à long terme pour déterminer les coûts directs et indirects des états pathologiques.
<b>Benefit plan design database</b>	Comprend entre autres des données sur les soins à domicile, la physiothérapie, etc.
<b>Laboratory database</b>	Ajoute des résultats de laboratoire aux éléments de données contenus dans les bases de données complémentaires commerciales et Medicare. Elle contient les résultats des analyses de laboratoire des patients hospitalisés et ambulatoires depuis 2005.
<b>Health risk assessment database</b>	Fournit des données sur des facteurs de risques comme le tabagisme, la consommation d'alcool, l'indice de masse corporelle et l'exercice physique
<b>Dental database</b>	Fournit des données sur des soins dentaires et médicaux
<b>Hospital drug database</b>	Fournit des données plus détaillées sur la consommation de médicaments en milieu hospitalier, notamment la durée du traitement et la voie d'administration.

B/ Bases de données supplémentaires	
<b>Inpatient drug link file</b>	Fournit un lien entre les principales données du MarketScan et les données des patients hospitalisés.
<b>EMR-claims linked data (EMR: electronic medical record)</b>	Élargit les données cliniques disponibles par MarketScan en établissant un lien déterministe et probabiliste avec les données disponibles du EMR.
<b>Mortality file</b>	Comprend des données de mortalité d'un sous-groupe de patients pour lesquels la mortalité a été identifiée via un lien avec les enregistrements de la SSA (Social Security Administration)
<b>Oncology EMR database</b>	Comprend des données sur l'histologie, le type/le stade de la tumeur et les traitements du cancer
<b>Oncology EMR-claims linked dataset</b>	Élargit les données disponibles via MarketScan en établissant un lien avec les données d'oncologie clinique disponibles
<p><b>2- Base de données du SHS (Symphony Health Solutions) :</b> base de données qui contient les demandes d'ordonnances jugées dans les États-Unis pour tous les types de paiement. L'échantillon global représente près de 30 000 pharmacies, 1 000 hôpitaux, 800 établissements de soins ambulatoires et 80 000 cabinets médicaux. Cette base de données contient des plans commerciaux, Medicare Part D, les paiements en espèces, les programmes d'assistance et Medicaid.</p>	
<p><b>3- Optum Labs Data Warehouse (OLDW)<sup>(230)</sup>:</b> base de données sur les soins de santé qui contient les demandes de remboursement de 100 millions de personnes inscrites à d'autres régimes d'assurances commerciales ou à Medicare Advantage sur une période de 20 ans. Elle intègre des données de santé issues des assurances commerciales et Medicare Advantage.</p>	
<p>4- Healthcore integrated research environment (HIRE) : elle intègre les données sur les demandes de remboursement des frais médicaux et pharmaceutiques de près de 40 millions de membres ayant une assurance commerciale dans 14 régimes régionaux de santé aux États-Unis.</p>	
<p><b>5- Pharmetrics:</b> Elle contient les données issues des assurances de santé suivantes :</p> <ul style="list-style-type: none"> <li>C - Commercial</li> <li>K - State Children's Health Insurance Program (CHIP)</li> <li>M - Medicaid</li> <li>R - Medicare Risk (presently known as Medicare Advantage)</li> <li>S - Self-Insured</li> <li>T - Medicare Cost (Medicare Supplemental)</li> </ul>	

## **6-Sentinel : système électronique national de la FDA<sup>(231)</sup>**

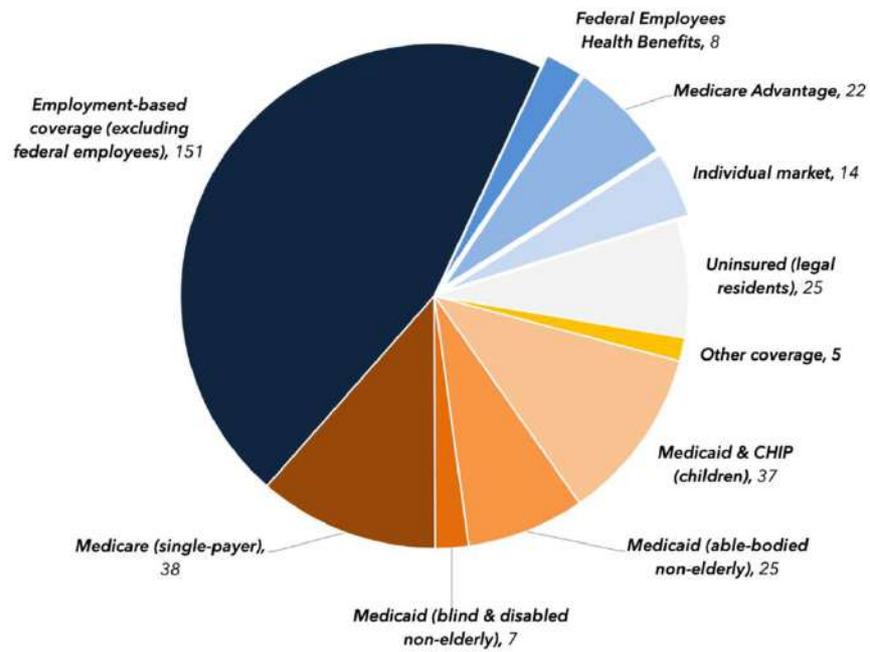
Surveille la sécurité des produits médicaux réglementés par la FDA, y compris les médicaments, les vaccins, les produits biologiques et les produits médicaux. Ce système est composé de données dont la qualité a été vérifiée et qui sont détenues par 18 institutions partenaires. Les données de ces institutions partenaires sont générées à partir des interfaces entre les patients et le système de santé américain par l'intermédiaire de leurs assureurs et de leurs prestataires. Par exemple, le patient fictif Bob (illustré ci-dessous) est un homme qui est inscrit dans un plan de santé et qui a fait plusieurs visites médicales entre 2011 et 2014.

### **Institutions partenaires des données**

- 1. Aetna, a CVS Health company**
- 2. Blue cross blue shield of massachusetts**
- 3. Duke university school of medicine: department of population health sciences (100% medicare fee-for-service data)**
- 4. Harvard Pilgrim Health Care Institute**
- 5. HCA Healthcare**
- 6. HealthCore, Inc (Anthem, Inc data)**
- 7. HealthPartners Institute**
- 8. Humana, Inc.**
- 9. Kaiser permanente colorado institute for health research**
- 10. Kaiser permanente center for health research hawaii**
- 11. Kaiser foundation health plan of the mid-atlantic states, inc.**
- 12. Kaiser permanente northern california, division of research**
- 13. Kaiser permanente northwest center for health research**
- 14. Kaiser permanente washington health research institute**
- 15. Marshfield clinic research institute**
- 16. Optum (optuminsight life sciences inc. And optum labs®)**
- 17. Vanderbilt university medical center, department of health policy (tennessee medicaid data)**

Figure A1: Assurance de santé aux États-Unis en 2019

Health Insurance Enrollment in America, 2019 (Millions)



Source: FREOPP— The Foundation of research for research for equal opportunity

## Annexe 2. Tableaux et figures annexes de la partie 3 et 4

Tableau A2: Lignes directrices pour les rapports de Prisma

Section/topic	#	Checklist item
Title	1	Identify the report as a systematic review, meta-analysis, or both.
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ ) for each meta-analysis.
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.

Section/topic	#	Checklist item
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.

Tableau A3: Caractéristiques des patients dans les études incluses ( partie 3)

Etude, année	AOD, dose, durée	Nombre total des patients	Age		Sexe masculin (%)		Insuffisance rénale (%)		Antécédents ETEV (%)		Cancer (%)		% PE ± TVP		% TVP	
			AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK
<b>Essais contrôlés randomisés</b>																
<b>RE-COVER, 2009</b>	Dabigatran, 150 mg bid, 6 mo	2539	55*	54,4*	58	58,9	NC	NC	25,7	25,4	5	4,5	31	31	69	69
<b>EINSTEIN-DVT, 2010</b>	Rivaroxaban, 15 mg bid pour 21 days puis 20mg od, (3,6,12) mo	3449	55,8	56,4	57,4	56,4	6,6	7	19	19,2	6,8	5,2	0	0	100	100
<b>EINSTEIN-PE, 2012</b>	Rivaroxaban, 15 mg bid pour 21 jours puis 20mg od, (3,6,12) mo	4832	57,9	57,5	54	51,7	8,6	7,9	18,8	20,3	8,6	7,9	100	100	0	0
<b>AMPLIFY,2013</b>	Apixaban, 10 mg bid pour 7jours puis 5 mg bid,6 mo	5395	57,2	56,7	58,3	59,1	6	5,5	17,2	15,1	2,5	2,8	35	35	65	65
<b>HOKUSAI-VTE, 2013</b>	Edoxaban, 60 mg od ou 30 mg if 30<Crcl< 50ml/min ou poids< 60 kg, 3 à 12 mo	8240	55,7	55,9	57,3	57,2	6,5	6,6	19	17,9	9,2	9,5	40	40	60	60

Etude, année	AOD, dose, durée	Nombre total des patients	Age		Sexe masculin (%)		Insuffisance rénale (%)		Antécédents ETEV (%)		Cancer (%)		% PE ± TVP		% TVP	
			AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK
<b>Essais contrôlés randomisés</b>																
<b>RE-COVERII, 2014</b>	Dabigatran, 150 mg bid, 6mo	2568	54,7	55,1	61	61,2	NC	NC	19	15,8	3,9	3,9	31	31	69	69
<b>Etudes de cohortes prospectives avant SP</b>																
<b>XALIA, 2015</b>	Rivaroxaban, 15 mg pour 21 jours puis 20 mg, ≥ 3 mo	4768	2149	59*	66*	55	52	4	10	24	22	6	19	8	12	92
<b>REMOTEV, 2016</b>	Rivaroxaban, 15 mg pour 21 days puis 20 mg, 3 à 6 mo	419	111	62,2	70,7	49	46,8	15,2	44,1	29,2	37,8	2,6	8,1	90	91,9	10
<b>SWIVTER 2016</b>	Rivaroxaban, 15 mg pour 21 days puis 20 mg, 3 mo	2062	417	56	65	56,1	51,7	7	12,8	22,1	23,6	9,6	27,5	38	66	62
<b>XALIA-LEA, 2019</b>	Rivaroxaban, 15 mg bid pour 21± 3 jours puis 20 mg od, ≥ 3 mo	1687	59,6	58	48,5	44,8	9,7	14,9	11,7	13,7	16,8	17,2	31,4	40,8	68,6	59,2
<b>Bui et al, 2019</b>	Rivaroxaban, 15 mg bid pour 21 puis 20 mg od, 6 mo	187	57,08	60,2	36,1	25,2	8,4	12,5	3,6	2,9	6	8,4	14,4	11,5	85,6	88,5

Etude, année	AOD, dose, durée	Nombre total des patients	Age		Sexe masculin (%)		Insuffisance rénale (%)		Antécédents ETEV (%)		Cancer (%)		% PE ± TVP		% TVP	
			AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK
Etudes de cohortes prospectives après SP																
<b>SWIVTER, 2016</b>	Rivaroxaban, 15 mg pour 21 jours puis 20 mg, 3 mo	834	56	57	56,1	54,9	7	8,6	22,1	25,9	9,6	11	NC	NC	NC	NC
Etudes sur bases de données de santé informatisées avant SP																
<b>Sindet-Pedersen et al, 2017</b>	Rivaroxaban, 15 mg pour 21 jours puis 20 mg, 6 mo	12318	66*	66*	54	54,2	2,3	5,5	12,5	16,9	11,2	14,3	41	38	59	62
<b>Jun et al, 2017</b>	AOD, 3mo	148972	62,9	62,1	47,5	48,8	9,1	15,5	NC	NC	14,3	17,4	NC	NC	NC	NC
<b>Wang et al, 2017</b>	Rivaroxaban	4849	65	67	94	95	21	27	NC	NC	37	35	100	100	0	0
<b>Guo et al, 2020</b>	Apixaban	25139	53	52	56	54	8	11	NC	NC	NC	NC	45	45	55	55
<b>Bounameaux et al, 2020</b>	AOD	7987	59	61	51	52	25	20	17	16	4	4	38	40	70	60

Etude, année	AOD, dose, durée	Nombre total des patients	Age		Sexe masculin (%)		Insuffisance rénale (%)		Antécédents ETEV (%)		Cancer (%)		% PE ± TVP		% TVP	
			AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK	AOD	AVK
<b>coleman et al, 2018 (2)</b>	Rivaroxaban, 1mo to 12 mo	17618	57*	59*	48,6	48,9	8,6	7,3	NC	NC	26,3	28,5	44,9	43,3	55,1	56,7
Etudes sur bases de données de santé informatisées après SP																
<b>Coleman et al, 2018 (1)</b>	Rivaroxaban, 1mo to 12 mo	47342	56*	56*	53,9	51,6	3,5	3,9	NC	NC			46,7	48	53,3	52
<b>Dawass et al, 2020</b>	Apixaban, 12 mo	36907	58,9	59,3	48,3	47,9	13,4	14,4	NC	NC	16,2	17				
<b>Weycker et al, 2018</b>	Apixaban, 5mg, mean 116 days	35756	60	60	51,9	51,9	13,4	13,4	NC	NC	1,2	1,1	41	41	59	59
<b>Costa et al 2020</b>	Rivaroxaban, 3,6, 12 mo	4939	50*	51*	43,6	43,6	14,7	16,6	NC	NC	3,5	3,8	18,6	19,1	81,4	80,9
<b>Baviera et al, 2020</b>	AOD(Rivaroxaban, Apixaban), 3,6, 12 mo	9362	72,8	72,8	43,9	44,7	NC	NC	NC	NC	13,2	13,2	77,8	76,3	22,2	23,7
<b>Jun et al, 2017</b>	AOD, 3mo	59525	62,8	64,7	46,6	46,4	6,9	7,6	NC	NC	14,5	14,3	NA	NA	NA	NA
Etudes de cohortes rétrospectives																
<b>Badreldin et al, 2018</b>	AOD	441	62	62	60	52	NC	NC	18	19	NC	NC	39	43	61	57
<b>Basto et al, 2018</b>	AOD	107	64	61	95	91	20	31	NC	NC	NC	NC	64	63	36	37
<b>Wakakura et al, 2018</b>	AOD	807	70	69	32	17	29	33	NC	NC	33	29	30	30	70	70
<b>Fung et al, 2019</b>	Rivaroxaban	181	63	62	41	47	12	10	2	0	0	0	28	23	72	77

bid: deux fois par jour, od: une fois par jour, mo: mois, IQR: : Ecart interquartile; NC: non connu, \*: Median

Tableau A4: Équation de recherche utilisée dans PubMed (partie 3)

#16	(((((((((((#1) AND#2) AND#3) AND#4) AND#5) NOT#3)NOT#4)NOT#5)NOT#6)NOT#7)NOT#8)NOT#9)NOT#10)NOT#11)NOT#12)NOT#13) NOT#14)NOT#15)
#15	NOT ("post-thrombotic syndrome"[Title] OR "postthrombotic syndrome"[Title])
#14	NOT ("secondary prevention"[Title] OR "extended"[Title])
#13	NOT ("elderly"[Title] OR "older patients"[Title])
#12	NOT ("obese"[Title] OR "obesity"[Title])
#11	NOT ("cancer*" [Title] OR "oncology patients"[Title] OR "malignancy"[Title] OR "malignancies"[Title] OR "multiple myeloma"[Title])
#10	NOT ("surgery"[Title] OR "surgeries"[Title] OR "knee"[Title] OR "hip"[Title] OR "arthroplasty"[Title] OR "arthroplasties"[Title])
#9	NOT "switching"[Title]
#8	NOT "adherence"[Title]
#7	NOT ("post hoc analysis"[Title] OR "sub-analysis"[Title] OR "subanalysis"[Title] OR "subgroup"[Title] OR "subgroups"[Title])
#6	NOT ("recommendations"[Title] OR "design"[Title] OR "protocol"[Title] OR "review"[Title] OR "meta-analysis"[Title] OR "case series"[Title/Abstract] OR "case report"[Title/Abstract])
#5	("2009/01/01"[Date - Publication] : "3000"[Date - Publication])
#4	((clinical trials, randomized[MeSH Terms]) OR (controlled clinical trials, randomized[MeSH Terms]) OR (cohort studies[MeSH Terms]) OR (cohort studies, historical[MeSH Terms]) OR (double blind method[MeSH Terms]) OR (double blind methods[MeSH Terms]) OR (double blind studies[MeSH Terms]) OR (double blind study[MeSH Terms]) OR (method, double blind[MeSH Terms]) OR (registries[MeSH Terms]) OR (registry[MeSH Terms]) OR (retrospective studies[MeSH Terms]) OR (retrospective study[MeSH Terms]) OR (Comparative study[Publication type]) OR (retrospective cohort[Title/Abstract]) OR (prospective study[Title/Abstract]))
#3	((vitamin K antagonists[Title]) OR (vitamin K antagonist[Title]) OR (Vitamin K/antagonists and inhibitors[MeSH Terms]) OR (VKA[Title]) OR (VKAs[Title]) OR (warfarin/adverse effects[MeSH Terms]) OR (warfarin/therapeutic use[MeSH Terms]) OR (warfarin[Title]) OR (vitamin K antagonist[Title/Abstract]) OR (standard therapy[Title/Abstract]) OR (standard anticoagulant therapy[Title/Abstract]) OR (conventional anticoagulation[Title/Abstract]))
#2	((anticoagulant agents[MeSH Terms]) OR (anticoagulant drugs[MeSH Terms]) OR (anticoagulants[MeSH Terms]) OR (agents, anticoagulant[MeSH Terms]) OR (anticoagulants/adverse effects[MeSH Terms]) OR (anticoagulants/therapeutic use[MeSH Terms]) OR (Factor Xa Inhibitors/therapeutic use[MeSH Terms]) OR (Factor Xa Inhibitors/adverse effects[MeSH Terms]) OR (dabigatran[Title]) OR (rivaroxaban[Title]) OR (apixaban[Title]) OR (edoxaban[Title]) OR (direct oral anticoagulants[Title]))
#1	((deep venous thrombosis[MeSH Terms]) OR (pulmonary embolism[MeSH Terms]) OR (venous thromboembolism[MeSH Terms]) OR (venous thrombosis[MeSH Terms]) OR (venous thromboembolism/drug therapy[MeSH Terms]) OR (Pulmonary Embolism / drug therapy[MeSH Terms]) OR (deep venous thrombosis / drug therapy[MeSH Terms]) OR (venous thrombosis[Title]) OR (venous thromboembolism[Title]))

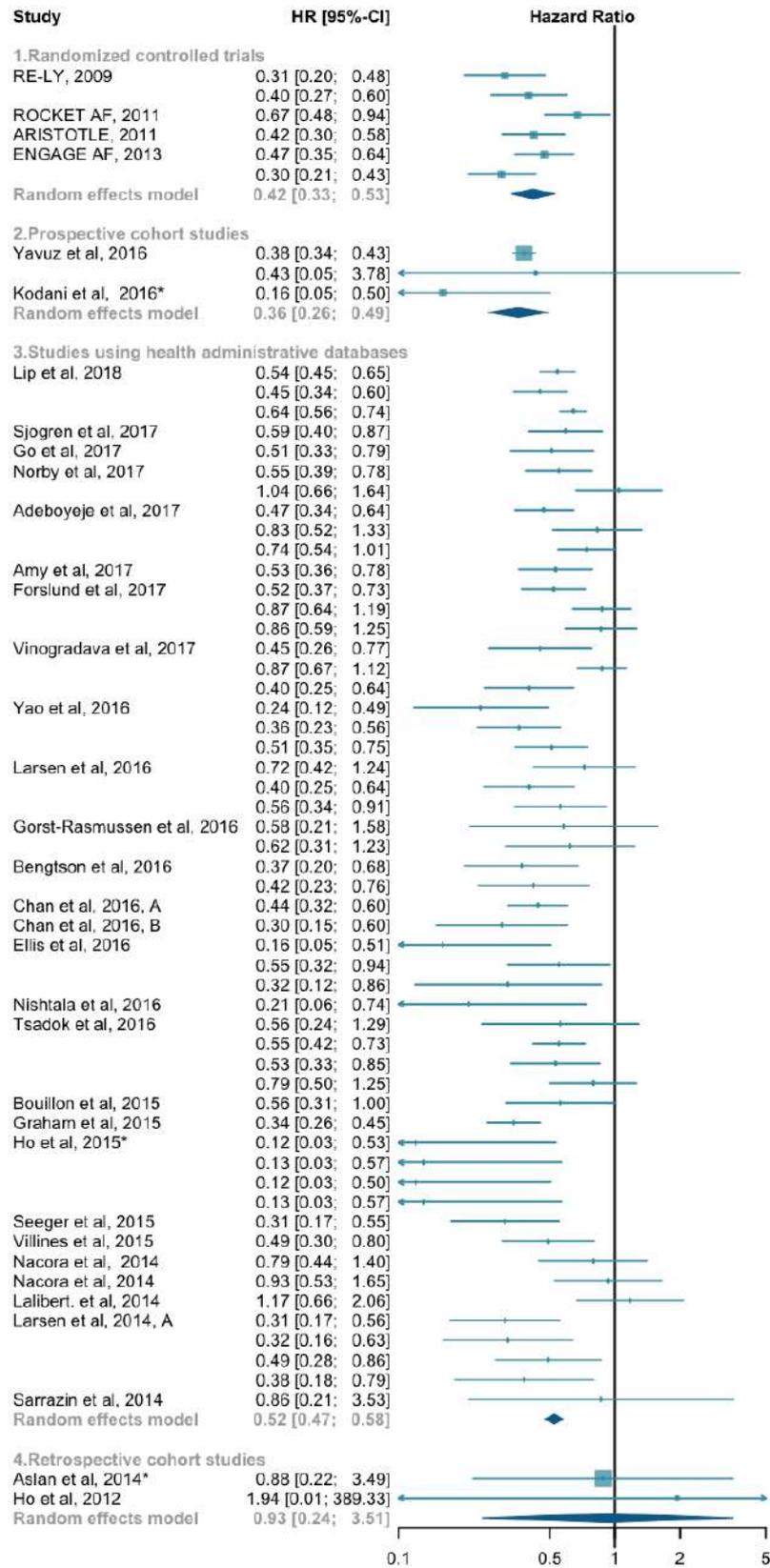
Tableau A5: Caractéristiques des études incluses dans la partie 3

Study, year	No. of patients	Experimental design	Database	Recruitment period	DOAC
<b>Randomized controlled trials</b>					
RE-COVER, 2009 (25)	2,564	Double blind	/	Apr. 2006 - Nov. 2008	Dabigatran
EINSTEIN-DVT, 2010 (26)	3,449	PROBE	/	Mar. 2007 - Sept. 2009	Rivaroxaban
EINSTEIN-PE, 2012 (27)	4,832	PROBE	/	Mar. 2007 - Mar. 2011	Rivaroxaban
AMPLIFY, 2013 (28)	5,395	Double blind	/	Aug. 2008 - Aug. 2012	Apixaban
HOKUSAI-VTE, 2013 (29)	8,292	Double blind	/	Jan. 2010 - Oct. 2012	Edoxaban
RE-COVER II, 2014 (30)	2,589	Double blind	/	Jun. 2008 - Oct. 2010	Dabigatran
<b>Prospective cohort studies</b>					
XALIA, 2015 (31)	4,515	PS adjustment	/	Jun. 2012 - Mar. 2014	Rivaroxaban
SWIVTER, 2016 (32)	834	PS matching	/	Nov. 2012 - Feb. 2015	Rivaroxaban
XALIA-LEA, 2019 (33)	1,687	No adjustment	/	Jun. 2014 - Oct. 2015	Rivaroxaban
Bui et al, 2019 (34)	187	No adjustment	/	Jan. 2017 - Aug. 2018	Rivaroxaban
<b>Living clinical databases</b>					
Gaertner et al., 2017 (35)	419	No adjustment	REMOTEV Registry	Nov. 2013 - Jul. 2015	Rivaroxaban
Bounameaux et al., 2020 (36)	7,987	PS weighting	GARFIELD-VTE Registry	Mar. 2014 - Jan. 2017	Apixaban, Dabigatran, Edoxaban or Rivaroxaban
<b>Living health administrative databases</b>					
Sindet-Pedersen et al, 2017 (18)	12,318	Multivariate model	Danish administrative registries (National Prescription Registry, National Patient Register, Civil Registration System)	Feb. 2012 - Sep. 2016	Rivaroxaban
Jun et al, 2017 (37)	59,525	PS matching	US Market scan + Canada	Jan. 2009 - Mar. 2016	Apixaban, Dabigatran or Rivaroxaban
Wang et al., 2017 (38)	813	PS matching	Veterans Health Administration	Oct. 2010 - Sept. 2015	Rivaroxaban
Coleman et al, 2018 (23)	36,853	PS weighting	US Truven Market scan	Jan. 2012 - Dec. 2016	Rivaroxaban

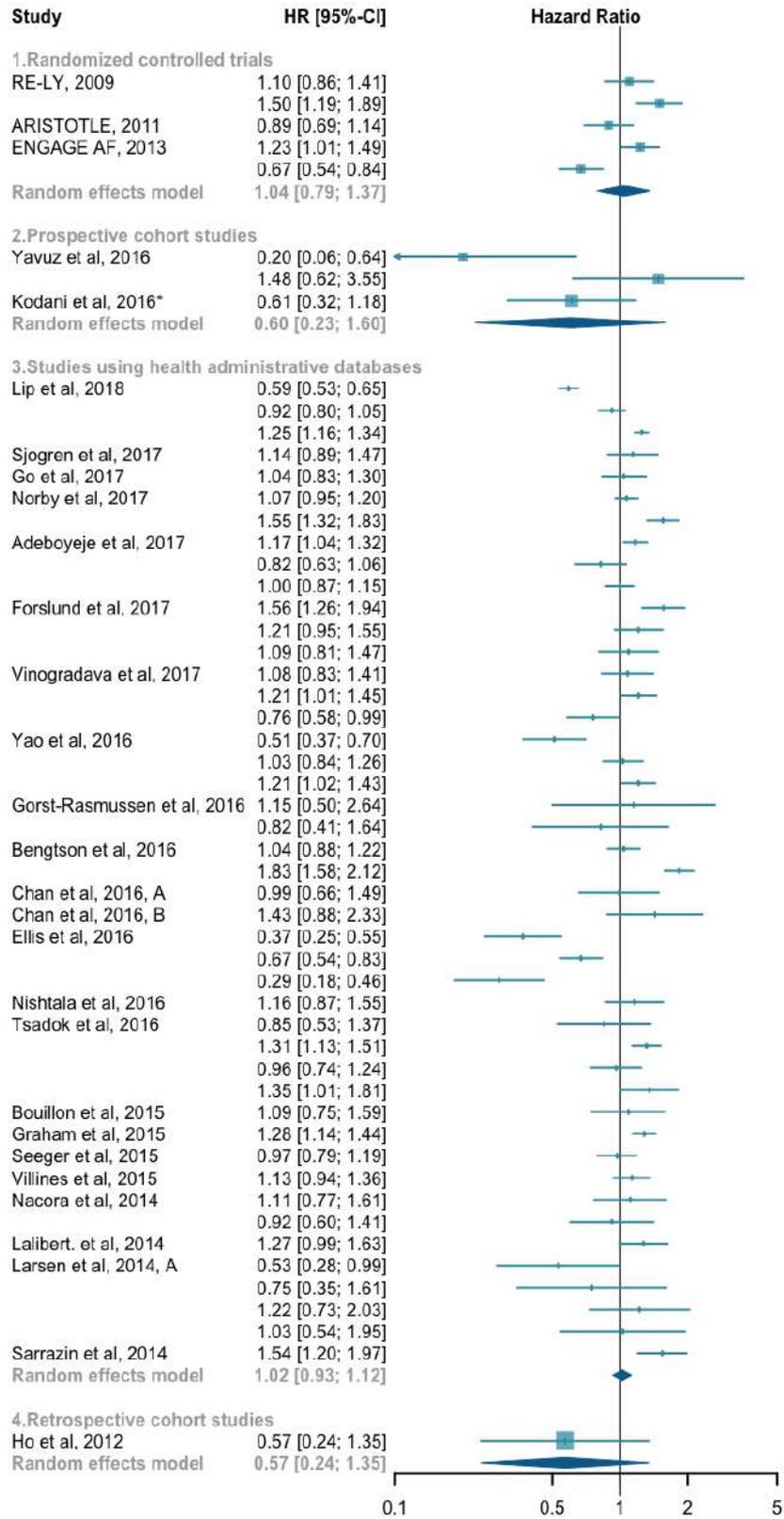
<b>Living health administrative databases</b>					
<b>Coleman et al, 2018 (24)</b>	17,618	PS weighting	US Truven Market scan	Jan. 2012 - Dec. 2016	Rivaroxaban
<b>Weycker et al, 2018 (39)</b>	35,756	PS matching	U.S. private health care claims databases (MarketScan Database, PharMetrics Database, Optum Database, Humana Database)	Sept. 2014 - Jun. 2017	Apixaban
<b>Costa et al, 2020 (40)</b>	4,939	PS weighting	US Optum EHR	Nov. 2012 - Sep. 2018	Rivaroxaban
<b>Dawwas et al, 2020 (41)</b>	36,907	PS matching	US Truven Market scan commercial +Medicare supplemental	Jan. 2014 - Dec. 2017	Apixaban
<b>Baviera et al, 2020 (42)</b>	9,362	PS matching	administrative health database of Lombardy	Jan. 2013 - Dec. 2017	Apixaban, Dabigatran, Edoxaban, Rivaroxaban
<b>Guo et al., 2020 (43)</b>	17,716	PS matching	IQVIA PharMetrics Real-World Data Adjudicated Claims	Sept. 2014 - Sept. 2018	Apixaban
<b>Retrospective cohort studies</b>					
<b>Badreldin et al., 2018 (44)</b>	441	No adjustment	Patients admitted to Brigham and Women's Hospital (USA)	Aug. 2015 - May 2017	Apixaban, Dabigatran, Edoxaban, Rivaroxaban
<b>Basto et al., 2018 (45)</b>	107	No adjustment	Veterans Affairs hospital (USA)	Jan. 2010 - Dec. 2016	Apixaban, Dabigatran, Rivaroxaban
<b>Wakakura et al., 2018 (46)</b>	807	No adjustment	Department of Cardiovascular Medicine (Tokyo)	Jan. 2009 - Dec. 2014	Apixaban, Dabigatran, Edoxaban, Rivaroxaban
<b>Fung et al., 2019 (47)</b>	181	No adjustment	Patients admitted to the Princess Margaret Hospital (China)	Mar. 2012 - Feb. 2017	Rivaroxaban

PS: propensity score; DOAC: direct oral anticoagulant; PROBE: Prospective Randomized Open Blinded End-point.

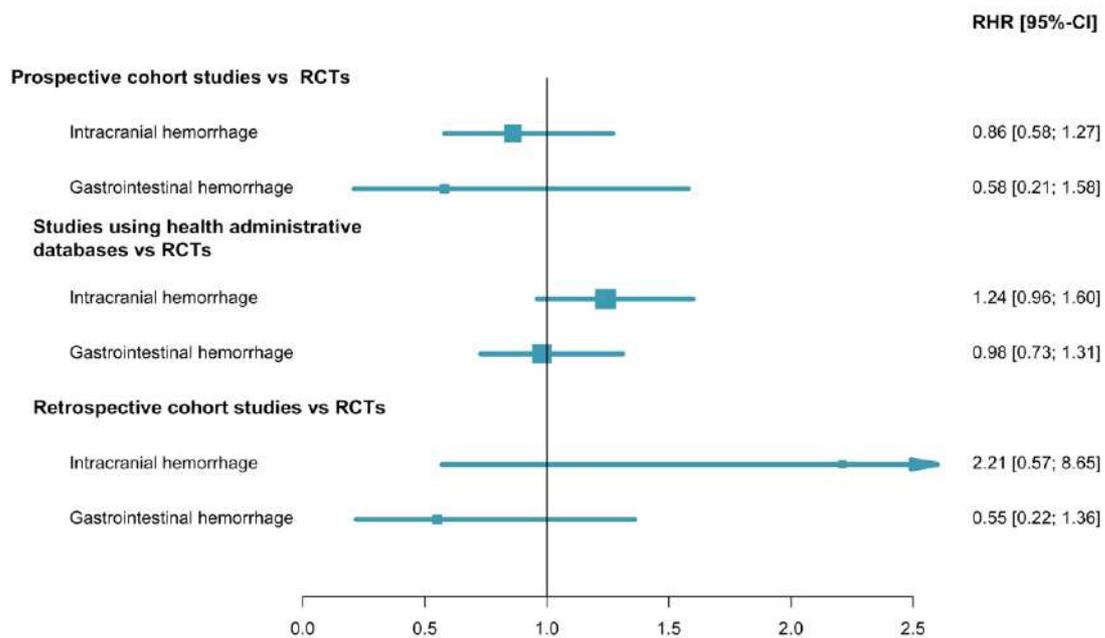
**Figure A2:** Comparaison des estimations des effets du traitement en fonction du type d'études sur les hémorragies intracrâniennes



**Figure A3:** Comparaison des estimations des effets du traitement en fonction du type d'études sur les hémorragies gastro-intestinales



**Figure A4:** RHR et intervalles de confiance à 95% des RHR pour quantifier la différence d'estimation des effets traitement entre les études observationnelles et les ECR pour les critères secondaires



**Tableau A6: Caractéristiques des études incluses dans la partie 4**

<b>Etude, année, pays</b>	<b>Nombre des patients</b>	<b>Plan expérimental</b>	<b>AOD, dose, durée de traitement</b>	<b>Période de recrutement</b>		<b>Résultats étudiés</b>	<b>HR/ RR, collecté/ calculé</b>
<b>Essais randomisés contrôlés</b>							
<b>RE-LY, 2009, multinationale (44pays)</b>	18,113 patients, AVK-naïf et expérimenté	En ouvert	Dabigatran, 110, 150 mg bid, 6 mois, durée médiane de suivi : 2 année	Décembre 2005	Décembre 2007	AVC ischémique/ES, HM, HGI, HIC	RR, collecté
ROCKET AF, 2011, multinationale (45 pays)	14,264 patients	Double aveugle	Rivaroxaban, 15,20 mg od, durée médiane de suivi : 590 jours for the per-protocol analysis and 707 days for the intention-to-treat population.	Décembre 2006	Juin 2009	AVC ischémique/ES, HM, HIC	HR, collecté
<b>ARISTOTLE, 2011, multinationale (40 pays)</b>	18,201 patients	Double aveugle	Apixaban, 2. 5, 5 mg bid, >12 mo,	Décembre 2006	Avril 2010	AVC ischémique/ES, HM, HGI, HIC	HR, collecté
<b>ENGAGE AF, 2013, multinationale (46 pays)</b>	21,105 patients	Double aveugle	Edoxaban, 30, 60 mg, durée médiane de traitement : 907 jours	Novembre 2008,	Novembre 2010.	AVC ischémique/ES, HM, HGI, HIC	HR, collecté
<b>Études utilisant des bases de données de santé informatisées</b>							
<b>Lip et al, 2018, USA</b>	321,182 patients, naïfs	Appariement sur SP	Dabigatran, 75, 150 mg bid, Apixaban, 2. 5, 5 mg bid, Rivaroxaban, 10, 15/20 mg od; 12 mo	Janvier 2013	Septembre 2015	AVC ischémique/ES, HM, HGI, HIC	HR, collecté
<b>Sjogren et al, 2017, Suède</b>	49,011 patients, naïfs+ expérimentes	Appariement sur SP	Dabigatran 110 150 mg bid, apixaban, ( 2. 5, 5 )mg bid, rivaroxaban, 15,20 mg od; <1 année	Juillet 2011	Décembre 2014	AVC ischémique/ES, HM, HGI, HIC	HR, collecté
<b>Go et al, 2017, USA</b>	50,578 patients, naïfs	Appariement sur SP	Dabigatran	Novembre 2010	Mai 2014	AVC ischémique, HIC, HGI	HR, collecté

Etude, année, pays	Nombre des patients	Plan expérimental	AOD, dose, durée de traitement	Période de recrutement		Résultats étudiés	HR/ RR, collecté/ calculé
Nielsen et al, 2017, Danemark	55,644 patients, naïfs	Pondération inverse sur SP	Dabigatran 110 mg, Rivaroxaban 15 mg, Apixaban 2.5 mg	Apixaban : Décembre 2012, dabigatran: Aout 2011, rivaroxaban: Février 2012	Février 2016	AVC ischémique /ES, HM	HR, collecté
Norby et al, 2017, USA	133,740 patients, naïfs+ switch	Appariement sur SP	Rivaroxaban 10,15,20 mg, 12 mo	Janvier 2010	Décembre 2014	AVC ischémique, HGI, HIC	HR, collecté
Adeboyeje et al, 2017, USA	44,075 patients, naïfs	Pondération inverse sur SP	dabigatran 150 mg bid, rivaroxaban 15 mg, apixaban 2.5 mg	Novembre 2010	Février 2016	HIC, HM	HR, collecté
Amy et al, 2017, Canada	34, 965 patients, naïfs + switch	Modèle de régression multiple	dabigatran, rivaroxaban, apixaban, >= 6mo	Janvier 2009	Juin 2015	AVC ischémique, HIC	HR, collecté
Forslund et al, 2017, Suède	22,198 patients, naïfs	Pondération inverse sur SP	Dabigatran, Rivaroxaban, Apixaban	Janvier 2012	Décembre 2015	AVC ischémique, HM, HGI, HIC	HR, collecté
Koshaka et al, 2017, Japan	38,662 patients, naïfs +switcher	Appariement sur SP	Dabigatran, 110,150 mg bid, apixaban 2.5,5 mg bid, rivaroxaban 10,15 od,	Mars 2011	Mars 2016	HM	HR, collecté
Vinogradava et al, 2017, UK	103,270 patients, naïfs	Modèle de régression multiple	Dabigatran, 150 mg bid, rivaroxaban 20 mg, apixaban 10 mg ; 5mo, 9mo	Janvier 2011	Mars 2016	AVC ischémique, HM, HGI, HIC	HR, collecté
Yao et al, 2016, USA	76,354 patients, naïfs + expérimenté	Appariement sur SP	Dabigatran, 75,150 mg bid, apixaban 5 mg bid, rivaroxaban ,10,15,20 mg od, >=12mo	Octobre 2010	Juin 2015	AVC/ ES, HM, HGI, HIC	HR, collecté

Etude, année, pays	Nombre des patients	Plan expérimental	AOD, dose, durée de traitement	Période de recrutement		Résultats étudiés	HR/ RR, collecté/ calculé
<b>Larsen et al, 2016, Danemark</b>	61,678 patients, naïfs	Pondération inverse sur SP	Dabigatran, 150 mg bid, apixaban 5 mg bid, rivaroxaban 20 mg od,	Apixaban : décembre 2012, warfarin et dabigatran: aout 2011, rivaroxaban: Février2012	Novembre 2015.	AVC ischémique ou ES, HIC, HM	HR, collecté
<b>Gorst-Rasmussen et al, 2016, Danemark</b>	22,358 patients, naïfs	Ajustement sur PS	Rivaroxaban 15, 20 mg	Février 2012	Aout 2014	AVC /ES/ AIT, HM, HGI, HIC	HR, collecté
<b>Bengtson et al, 2016, USA</b>	56,688 patients, naïfs +switch	Appariement sur SP	Dabigatan75, 150mg bid, Rivaroxaban10, 15, 20 mg	Dabigatan : octobre 2010, rivaroxaban : novembre 2011	Décembre 2012	AVC ischémique, HGI, HIC	HR, collecté
<b>Chan et al, 2016, A, Taiwan</b>	19,853 patients, naïfs + expérimenté	Pondération inverse sur SP	Dabigatran 110, 150 bid	Juin 2012,	Décembre2013	AVC ischémique, HM, HGI, HIC	HR, collecté
<b>Chan et al, 2016, B, Taiwan</b>	9,167 patients, naïfs + expérimenté	Pondération inverse sur SP	Rivaroxaban 10, 15 mg od	Février 2013	Décembre 2013	AVC ischémique ou ES, HM, HGI, HIC	HR, collecté
<b>Ellis et al, 2016, Israel</b>	18,249 patients, naïfs	Modèles de régression multiple	Dabigatran, 150/110 mg bid, rivaroxaban, >=3mo	Janvier 2011	December2013	HM, HGI, HIC	RR, calculé
<b>Nishtala et al, 2016, New zeeland</b>	8,770 patients, naïfs	Appariement sur SP	Dabigatran 110, 150 mg bid	Juillet 2011	Décembre 2012	HGI, HIC	HR, collecté
<b>Tsadok et al 2016, Canada</b>	63,110 patients, naïfs + expérimenté	Ajustement sur PS	Dabigatran, 150,110 mg bid, 30 days	Avril 2011	March 2013	AVC/AIT, GIB, ICH	HR, collecté
<b>Bouillon et al, 2015, France</b>	17,410 patients, naïfs +switch	Modèle de régression multiple	Dabigatran 75,110, 150 mg bid, rivaroxaban 10,15 20 mg	Janvier 2011	Novembre 2012,	AVC ischémique ou ES, HGI, HIC	HR, collecté pour AVC ischémique ou ES ; RR calculé pour HGI, HIC

<b>Etude, année, pays</b>	<b>Nombre des patients</b>	<b>Plan expérimental</b>	<b>AOD, dose, durée de traitement</b>	<b>Période de recrutement</b>		<b>Résultats étudiés</b>	<b>HR/ RR, collecté/ calculé</b>
<b>Graham et al, 2015, USA</b>	134,414 patients, naïfs	Appariement sur SP	Dabigatran 75, 150 mg bid, 6 mo	Octobre 2010	Décembre 2012	AVC ischémique, HM, HGI, HIC	HR, collecté
<b>Ho et al, 2015, Hong kong</b>	1,821 patients, naïfs	Modèle de régression multiple	Dabigatran	Juillet 1997	Décembre 2011	AVC ischémique, HIC	RR, calculé
<b>Seeger et al, 2015, USA</b>	73,831 patients, naïfs	Appariement sur SP	Dabigatran 150mg	October2010	Décembre 2012	AVC ou ES, HM, HIC, HGI	HR, collecté
<b>Villines et al, 2015, USA</b>	19,853 patients, naïfs	Appariement sur SP	Dabigatran 150, 75mg	Octobre 2009	Juillet 2013	AVC, HM, HGI, HIC	HR, collecté
<b>Nacora et al, 2014, France</b>	71,589 patients, naïfs	Méthode de Cochran-Mantel-Haenszel	Dabigatran 75,110, 150 mg bid, rivaroxaban 10,15, 20 mg	AOD : juin 2012 AVK : Juin 2011	AOD : novembre 2012, AVK : novembre 2011	AVC ischémique ou ES, HM, HGI, HIC	HR, collecté
<b>Laliberté et al, 2014, USA</b>	18,270 patients, naïfs	Appariement sur SP	Rivaroxaban, >6mo	Mai 2011	Juillet 2012	AVC ou ES, HM, HGI, HIC	HR, collecté
<b>Larsen et al, 2014, A</b>	33945 patients, naïfs+ switch	Appariement sur SP	Dabigatran 110, 150 mg	Dabigatran: August 2011, AVK: August 2009	Dabigatran: Mai 2013, AVK: Mai 2013	HM, HGI, HIC	HR, collecté
<b>Larsen et al, 2014, B</b>	6,141 patients, naïfs + expérimenté	Appariement sur SP	Dabigatran 110,150 mg bid	Dabigatran: Aout 2011, AVK: Aout 2009	Dabigatran: Mai 2013, AVK: Mai 2013	AVC	HR, collecté
<b>Sarrazin et al, 2014, USA</b>	85,344 patients, switch	Modèle de régression multiple	Dabigatran 150 mg, 15 mo	2010	2012	HGI, HIC	HR, collecté

Etude, année, pays	Nombre des patients	Plan expérimental	AOD, dose, durée de traitement	Période de recrutement	Résultats étudiés	HR/ RR, collecté/ calculé	
<b>Etudes de cohortes rétrospectives</b>							
<b>Russo-Alvarez et al, 2018, USA</b>	944 patients, naïfs	Appariement sur SP	Rivaroxaban, durée médiane du traitement : 9 mo	Janvier 2012	Juillet 2016	AVC ischémique/AIT/ETE V, HM	HR, collecté
<b>Jacobs et al, 2016, USA</b>	5,254patients, naïfs	Appariement sur SP	DOAC (Rivaroxaban, Apixaban, Dabigatran)	Juin 2010	Décembres 2014	AVC/AIT, HM	HR, collecté
<b>Korenstra et al, 2016, Netherlands</b>	766 patients, naïfs + switch	Appariement sur SP	Dabigatran, 110,150 mg bid	Janvier 2010	Décembre 2012	AVC ou ES, HM	HR, collecté
<b>Yap et al, 2015, Malaysia</b>	1000, naïfs+ switch	Pas de correction	Dabigatran, 110 ,150 mg	Janvier 2009	Décembres 2013	AVC ischémique, HM	HR, collecté
<b>Aslan et al, 2014, Turquie</b>	439 patients, naïfs	Pas de correction	Dabigatran, 110,150 mg	Mars 2012	Décembre 2013	AVC ischémique, HIC, HM	RR, calculé
<b>Ho et al, 2012, Hong kong</b>	244 patients, naïfs	Pas de correction	Dabigatran, 110 mg bid	Janvier 2010.	Novembre 2011	AVC, HIC, HM, HGI	HR, collecté
<b>Etudes de cohortes prospectives</b>							
<b>Kodani et al, 2016, Japan</b>	6,616 patients, naïfs+ switch	Régression multiple	AOD: Dabigatran, rivaroxaban, apixaban et edoxaban	Dabigatran : mars 2011 ; rivaroxaban: avril 2012; apixaban: février 2013, edoxaban: septembre 2014	Durée totale de suivi :5 ans	AVC ischémique/ AIT/ES, HIC, HM, HGI	RR, calculé
<b>Yavuz et al, 2016, USA</b>	555 patients, naïfs	Pas de correction	Dabigatran 110, 150 mg bid ;18 mo	Mai 2013	Novembre 2014	AVC, HIC, HM, HGI	RR, collecté

bid: deux fois par jour, od: une fois par jour, mo: mois, ES: embolie systémique, HIC: hémorragie intracrânienne, HGI: hémorragies gastro-intestinales, HM: hémorragies majeurs , AIT: accident ischémique transitoire

Tableau A7: Équation de recherche utilisée dans PubMed (partie 4)

#10	((((((((#1) AND#2)AND#3)NOT#4)NOT#5)AND#6)NOT#7)NOT#8)NOT#9)
#9	("adherence"[Title])
#8	("cost"[Title])
#7	("design"[Title])
#6	("review"[Publication Type])
#5	("hemorrhage/epidemiology"[MeSH Terms] OR "hemorrhage/chemically induced"[MeSH Terms] OR "myocardial ischemia/prevention and control"[MeSH Terms] OR "gastrointestinal hemorrhage/chemically induced"[MeSH Terms] OR "hemorrhage/complications"[MeSH Terms] OR "gastrointestinal diseases/chemically induced"[MeSH Terms] OR "gastrointestinal diseases/mortality"[MeSH Terms] OR "ischemic attack, transient/prevention and control"[MeSH Terms] OR "intracranial hypertension/chemically induced"[MeSH Terms] OR "thromboembolic events"[Title/Abstract] OR "stroke/chemically induced"[MeSH Terms] OR "warfarin/adverse effects"[MeSH Terms] OR "stroke/mortality"[MeSH Terms] )
#4	("recommendations"[Title])
#3	("anticoagulants/therapeutic use"[MeSH Terms] OR "anticoagulants/adverse effects"[MeSH Terms] OR "rivaroxaban/therapeutic use"[MeSH Terms] OR "rivaroxaban/adverse effects"[MeSH Terms] OR "dabigatran/therapeutic use"[MeSH Terms] OR "dabigatran/adverse effects"[MeSH Terms] OR "warfarin/therapeutic use"[MeSH Terms] OR "dabigatran"[MeSH Terms] OR "dabigatran"[Title/Abstract] OR "rivaroxaban"[Title/Abstract] OR "apixaban"[Title/Abstract] OR "edoxaban"[Title/Abstract] )
#2	("atrial fibrillation"[Title/Abstract])
#1	("cohort studies"[MeSH Terms] OR "observational studies as topic"[MeSH Terms] OR "follow up studies/methods"[MeSH Terms] OR "follow up studies"[MeSH Terms] OR "multicenter study"[Publication Type] OR "comparative study"[Publication Type] OR "real world"[Title/Abstract] OR "real life"[Title/Abstract] OR "nationwide study"[Title/Abstract])

## Annexe 3. Article soumis

### Manuscript Details

<b>Manuscript number</b>	JCE_2020_554
<b>Title</b>	Observational studies versus randomized clinical trials with direct oral anticoagulants: a meta-epidemiological approach
<b>Article type</b>	Original article

### Abstract

Background: Randomized controlled trials (RCTs) are often criticized for including populations that are too highly selected. Health Authorities consequently encourage "real-life" post-marketing cohort studies comprising unselected populations. We aimed to determine the differences between RCTs and observational studies as regards their patient populations and efficacy/safety results. Materials and methods. A systematic literature review was conducted to identify randomized and observational studies including patients with venous thromboembolism (VTE) receiving direct oral anticoagulants (DOACs) or conventional treatment. Patient characteristics were compared between RCTs and observational studies. For efficacy and safety data, ratios of Hazard ratio (HR) obtained from RCTs and epidemiological studies respectively were computed. Results: Six RCTs (27,096 patients) and eighteen observational studies (368,563 patients) were identified and analyzed. Compared to RCTs, benefit of DOACs on efficacy endpoint (recurrent VTE) and safety endpoint (major bleeding) was estimated to be 41% higher (Ratio of HRs 0.59, [95% CI, 0.36 to 0.99]) and 23% higher (Ratio of HRs 0.77, [95% CI, 0.47 to 1.27]) respectively with prospective cohort studies. On the contrary, studies using health administrative databases found a higher benefit on efficacy (Ratio of HRs 0.82, [95% CI, 0.67 to 1.02]) and a lower benefit on safety (Ratio of HRs 1.30, [95% CI, 0.94 to 1.79]). Conclusions: Prospective studies do not seem to recruit patients less selected than those of RCTs but overestimate the effectiveness of DOACs. On the contrary, population in studies using health administrative databases may reflect the "real life" treated population. These studies provide a lower estimate of the treatment effect concerning safety, even if the result remains favorable to DOACs for efficacy and safety.

<b>Keywords</b>	direct oral anticoagulants; venous thromboembolism; randomized controlled trial; prospective cohort; studies using health administrative databases; meta-analysis
<b>Manuscript region of origin</b>	Europe
<b>Corresponding Author</b>	Silvy Laporte
<b>Corresponding Author's Institution</b>	Unité de Recherche Clinique Innovation Pharmacologie
<b>Order of Authors</b>	Maissa Safieddine, Céline Chapelle, Cyril Ferdynus, Laurent Bertoletti, Patrick Mismetti, Michel Cucherat, Silvy Laporte
<b>Suggested reviewers</b>	Philippe Girard, Francesco Salvo, Walter Ageno

### Submission Files Included in this PDF

#### File Name [File Type]

cover letter JCE.docx [Cover Letter]

highlights.docx [Highlights]

VTE manuscript vf.docx [Manuscript File]

Conflict of Interest statement.docx [Conflict of Interest]

Supplementary materials VTE sy ms.docx [Supplementary Material]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

Dear Editor,

On behalf of the authors, we wish to submit our manuscript, entitled "Observational studies versus randomized clinical trials with direct oral anticoagulants: a meta-epidemiological approach" for consideration for publication in the *Journal of Clinical Epidemiology*.

Randomized controlled trials (RCTs) are often criticized for including populations that are too highly selected. Health Authorities consequently encourage "real-life" post-marketing cohort studies comprising unselected populations. Our aim was to determine the differences between RCTs, and observational studies in terms of population's characteristics and efficacy/safety results. The novelty in our approach is to have distinguished prospective cohort studies and cohort studies based on medico-administrative database. For this purpose, a meta-epidemiological method was conducted in studies having assessed direct oral anticoagulants in the treatment of venous thromboembolism.

Six RCTs (27,096 patients), 5 prospective cohort studies (9,123 patients) and 10 cohort studies using health administrative databases were identified and analyzed.

Characteristics of patients included in prospective cohort studies were similar to those included in the RCTs, indicating that the implementation of prospective cohort studies did not lead to the inclusion of unselected patients when performed in a short time after RCTs. Not surprisingly, despite the adjustment to take into account confounding factors, prospective cohort studies were associated with a significant increase of treatment efficacy and safety compared to RCTs.

In contrast, baseline characteristics of patients included in historical cohort studies based on claims database records differed substantially and are closer to "real-life" patients. Compared to RCTs, cohort studies based on medico-administrative database were associated with an exaggeration of efficacy results, but a reduction of safety effect of about 30%.

The results of the present meta-epidemiologic study suggest that prospective cohort studies may not represent always a better real-life than RCTs. Cohort studies based on claims database records bring less enthusiastic information than RCTs, but persistence of confounding factors in such studies remain questionable. We believe that these findings will interest the broad readership of the *Journal of Clinical Epidemiology* as they may have important implications.

This work is original, has not been published elsewhere, and is not currently under consideration for publication elsewhere. This final version of the manuscript was read and approved by all authors for submission to the *Journal of Clinical Epidemiology*.

We thank you for your kind consideration of this manuscript and look forward to hearing from you soon.

Yours faithfully,

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Key findings

- When specific post-marketing prospective cohort studies are performed in a short time after RCTs, selection of included patients is the same

What this adds to what was known?

- Prospective cohort studies had significantly larger estimates of treatment benefit on efficacy, compared to RCTs
- Cohort studies based on health administrative database had smaller estimates of treatment benefit on safety compared to RCTs

What is the implication and what should change now?

- In all cases, despite of adjustment, the part of residual confusion is not excluded in observational studies

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59

**Observational studies versus randomized clinical trials with direct oral anticoagulants: a meta-epidemiological approach**

Maissa Safieddine<sup>a,b</sup>, Celine Chapelle<sup>a,c</sup>, Edouard Ollier<sup>a,c</sup>, Cyril Ferdynus<sup>b</sup>, Laurent Bertoletti<sup>c,d</sup>, Patrick Mismetti<sup>a,c,d</sup>, Michel Cucherat<sup>e</sup>, Silvy Laporte<sup>a,c</sup>

<sup>a</sup> Unité de Recherche Clinique, Innovation, Pharmacologie, CHU Saint-Etienne, Hôpital Nord, F-42055 Saint-Etienne; France.

<sup>b</sup> Unité de Soutien Méthodologique, INSERM, CIC1410, CHU de la Réunion, Saint-Denis, France

<sup>c</sup> SAINBIOSE U1059, Université Jean Monnet, Univ. Lyon, INSERM, F-42023 Saint-Etienne ; F-CRIN INNOVTE network ; France.

<sup>d</sup> Service de Médecine Vasculaire et Thérapeutique, CHU Saint-Etienne, Hôpital Nord, F-42055, Saint-Etienne, France.

<sup>e</sup> Service de Pharmacologie, HCL; UMR CNRS 5558 Evaluation et Modélisation des Effets Thérapeutiques, Université Claude Bernard Lyon 1, Lyon, France.

Number of words in the abstract: 249

Number of words in the main body of the text: 3590

Founding source: Scholarship from Tyr municipality-Lebanon

**Corresponding author:**

Silvy Laporte, MSc, PhD,  
CHU Saint-Etienne, Hôpital Nord  
Bâtiment Recherche  
Avenue Albert Raimond  
42055 Saint-Etienne, France  
Phone : +33 7 77 12 05 72  
Fax : + 33 4 77 12 78 20  
E-mail : Silvy.laporte@chu-st-etienne.fr

**ABSTRACT** (238 words, max. 250)

**Background:** Randomized controlled trials (RCTs) are criticized for including patients that are over-selected. Health Authorities consequently encourage “real-world” post-marketing cohort studies. Our objective was to determine the differences between RCTs and observational studies as regards their populations and efficacy/safety results.

**Materials and methods:** A systematic review was conducted to identify RCTs and observational studies including patients with venous thromboembolism (VTE) receiving direct oral anticoagulants (DOACs) or conventional treatment. Ratios of Hazard Ratio (RHR) comparing epidemiological studies (prospective and retrospective cohort studies and studies using living databases) to RCTs were computed.

**Results:** Six RCTs (27,121 patients) and twenty observational studies (248,971 patients) were identified and analyzed. Prospective cohort studies seemed to recruit patients who were no less selected than those of RCTs whereas other type of observational studies may reflect the population treated in real life. Among observational studies, prospective cohort studies yielded the most favorable estimates of treatment effect compared to RCTs. These studies were associated with a non-significant 33% increase in efficacy estimate (RHR 0.67, [95% CI, 0.39-1.18]) but no effect on safety estimate. Studies using living databases were associated with non-significant trends toward a greater effect on efficacy (RHR 0.82, [0.66-1.01]) and a smaller effect on safety (RHR 1.33, [0.96-1.84]).

**Conclusions:** Overall, in this clinical setting, an exaggeration of the treatment efficacy estimate was seen with observational studies compared to RCTs. As the presence of residual confounding cannot be excluded, these results should be interpreted cautiously.

**Keywords:** direct oral anticoagulants; randomized controlled trial; observational studies; meta-epidemiological study; residual confounding

## INTRODUCTION

Although randomized controlled trials (RCTs) allow control of selection and measurement biases, they are subject to criticism with regard to the population included. RCTs are often challenged for not being sufficiently inclusive of variables associated with the patient pool representative of the target population treated in current practice [1]. RCTs in many areas of medicine and surgery commonly exclude patients on the grounds of gender, age, severity of disease, comorbidities and/or concomitant drugs, race, access to emergency units, educational status and socioeconomic class [2–5]. The patients included in RCTs therefore have a high probability of differing from those who are eligible but not enrolled, resulting in a selective recruitment into these trials [2]. Another frequent criticism is that RCTs are generally too small to safely and accurately assess the benefit-risk ratio of the treatment evaluated. Faced with this lack of generalizability and the small number of patients included in these trials, Health Authorities encourage the implementation of large post-marketing cohort studies in order to gather data on “real-life” treated populations, more broadly representative of the population receiving the evaluated medication in clinical practice. Yet this incentive is flawed. First, there is no recommendation to favor one study design over another (prospective rather than retrospective, specific study or use of living databases). Second, as such studies are not randomized, they are subject to confounding and there is no recommendation to try to address the issue of selection bias and confounders. The statistical analysis of these observational studies may incorporate a correction mechanism based on adjustment or matching, generally using the propensity score (PS), to allow a comparable distribution of measured baseline variables between treated and untreated subjects. However, such methods can only adjust for confounding factors that have actually been measured [6]. The problem is to determine which observational study design can best capture all the known confounding factors.

We assumed that patients enrolled in RCTs and observational studies may differ in terms of characteristics that could modify treatment effects and might lead to differences in treatment effect estimates concerning the risks of VTE or major bleeding. We therefore sought to determine the actual differences between the populations included in RCTs and observational studies with regard to studies involving the assessment of direct oral anticoagulants (DOACs) for the treatment of venous thromboembolism (VTE), using meta-epidemiological methods and analysis of the consequences of this approach in terms of estimating the efficacy/safety ratio of these products.

## METHODS

The internal protocol of this meta-epidemiological study was written in accordance with the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) guidance and the PRISMA checklist is provided in Appendix 1.

### *Literature search*

We performed a systematic review of published RCTs and observational studies comparing DOACs to vitamin-K antagonists (VKAs) for the treatment of VTE with regard to recurrent VTE and major bleeding. An exhaustive literature search, both manual and computer-assisted, was performed from January 1, 2009 (first DOAC studied) to August 30, 2020. The computer-assisted search was carried out in various electronic databases (PubMed, Cochrane Database and Google Scholar). We constructed a search equation, including a panel of relevant keywords to identify articles in PubMed (Appendix 2). The search was not restricted to articles published in English.

### *Study Selection*

Two of the authors (MS and SL) independently evaluated studies for potential inclusion, disagreements being resolved by discussion. To be included in the meta-analysis, studies had to meet all the following criteria: (i) pivotal phase III RCTs as well as observational studies, (ii) comparing DOACs versus VKA administered for at least 3 months for the initial treatment of VTE. The comparisons were limited to DOACs approved or under regulatory review for the treatment of VTE, namely direct thrombin inhibitors (dabigatran) and factor Xa inhibitors (rivaroxaban, apixaban, and edoxaban) versus oral warfarin or VKA as the comparator. Three subcategories of observational studies were defined a priori in the internal protocol: 1) prospective cohort studies comprising observational studies with prospective recruitment and construction of a specific database for the study, 2) studies using living databases, comprising observational studies with prospective recruitment included in a living clinical database (i.e. a registry) or a living health administrative database and 3) retrospective cohort studies with retrospective recruitment.

Studies conducted on the same database were considered duplicates if more than half the recruitment period overlapped with that of another study with a larger enrollment and evaluating the same treatment groups in the same patients. We reviewed every study, and when duplicates were identified, we chose the study with the largest sample size. Studies with objectives not focusing on the effectiveness or safety profiles of anticoagulants, as well as

protocols and reviews that did not provide estimates of treatment effects in terms of efficacy or safety profiles were excluded.

#### *Data extraction*

Data were independently extracted by two of the authors (MS and SL). In the event of discrepancies between the reviewers, a consensus was reached by discussion. For each study, we extracted details of the study characteristics (year of publication, names of authors, country in which the study was conducted, number of patients in each group), study design, treatment information (type and dose), patient characteristics (mean age, gender, renal insufficiency, history of VTE, cancer) and outcome measures. We extracted treatment effect estimates for DOACs versus VKA, such as the hazard ratio (HR) or relative risk (RR) for VTE recurrence and major bleeding. Among observational studies, we differentiated prospective cohort studies (conducted specifically to investigate the defined treatments in the defined indication), retrospective cohort studies involving a retrospective collection of available data in patient medical records, and studies using health administrative databases or healthcare registries with prospective recording of patient information. For observational studies, we also extracted the matching or adjustment methods used (none, propensity score [PS] or multivariate model). For these studies, patient characteristics both before and after adjustment were extracted; HR and RR were extracted after matching or adjustment according to the PS or after adjustment using a multivariate model.

#### *Endpoint for efficacy and safety synthesis*

The primary efficacy outcome was that used as the primary endpoint in all randomized studies evaluating DOACs in this indication, i.e., recurrence of VTE. This efficacy outcome is a composite endpoint including fatal pulmonary embolism (PE), non-fatal symptomatic PE, and symptomatic recurrence of deep-vein thrombosis (DVT). The main safety outcome was major bleeding according to the definition issued by the International Society on Thrombosis and Haemostasis and including: fatal bleeding, bleeding causing a fall in hemoglobin level of 20 g/L or more or leading to transfusion of two or more units of whole blood or red blood cells, symptomatic bleeding in a critical area or organ (such as intracranial, intraocular, intraspinal, retroperitoneal, intra-articular pericardial, or intramuscular with compartment syndrome) [7].

#### *Statistical Analyses*

First, using descriptive statistics, we compared the populations enrolled in RCTs and observational studies including assessment of DOACs in the treatment of VTE. The characteristics of the patients enrolled in these studies were compared before and after application of the PS. Categorical variables extracted as percentages as well as continuous

variables extracted as means were summarized as an unweighted median for each study design.

Second, to compare the treatment effect of DOACs versus VKA between RCTs and observational studies (prospective cohort studies, studies using health administrative databases, or retrospective cohort studies), we calculated for each type of study, the pooled HR and its 95% confidence interval (CI) for VTE recurrence and major bleeding. The study data were aggregated using the logarithm of the HR method with inverse variance weighting [8,9]. When the value of the HR was not reported in the paper, it was estimated by the RR. For each meta-analysis, statistical heterogeneity between the studies was checked using Cochran's Q statistic and study consistency was quantified by means of the  $I^2$  statistic [10]. In the event of significant heterogeneity ( $p$  value less than 0.10), a random-effect model was used [11,12].

A  $P$ -value for the association  $<0.05$  was considered statistically significant. The results of the meta-analysis were presented graphically by a forest plot, including the effect size expressed as the HR with the corresponding 95% CI. A HR equal to 1 indicates no difference between the treatments, a HR  $<1$  indicates that DOACs are more effective than VKAs and a HR  $>1$  indicates that VKAs are more effective than DOACs. Begg funnel plots were used to explore potential publication bias and display the small- versus large-study effect according to the type of study [13]. Moreover, to quantify the difference in estimated treatment effect between observational studies and RCTs, we calculated the ratio of the HRs (RHR) by dividing the HR values obtained in observational studies by those obtained in RCTs. The RHR summarizes the association between the type of study and the magnitude of the treatment effect. A RHR  $< 1$  would imply that the estimates of treatment effect were higher with observational studies compared to RCTs, whereas a RHR  $> 1$  would imply that the estimates of treatment effect were lower with observational studies. We derived 95% CIs using robust standard errors allowing for heterogeneity between the studies included [14-16].

The meta-analysis was performed using R statistical software, version 3.5.1 (meta package, downloaded from [www.r-project.org](http://www.r-project.org)) [17].

#### *Sensitivity analyses*

Sensitivity analyses for both endpoints (recurrent VTE and major bleeding) were performed by excluding the observational studies that did not involve the use of any methods of adjusting for confounders in the analysis phase.

## RESULTS

### *Study Identification*

The search procedure identified 413 potentially eligible references, of which 366 were excluded after review of their titles and abstracts (Appendix 3). After assessment of the full texts, a further 17 references were excluded. Finally, 30 references were considered eligible for inclusion in the meta-analysis. These studies included two duplicate entries in a Danish database with a large overlap of the recruitment period [18,19]. As the second publication concerned a subgroup of 5004 patients with unprovoked venous thromboembolism within the study reported in the first publication (including 12,318 patients), only this larger study was retained [18]. Five studies conducted on the same American database (US Truven Market scan) were found [20-24]. Three studies [20-22] were excluded because they evaluated the same patients as two other larger studies [23,24], and more than half of their enrollment periods overlapped with those of the other studies. Finally, 26 studies (totalling 276,092 patients) were included: 6 RCTs [25-30], 4 prospective cohort studies [31-34], 12 studies using living databases (2 using living clinical databases [35-36] and 10 using living health administrative databases [18,23,24,37-43]), and 4 retrospective cohort studies [44-47].

### *Characteristics of patients treated with DOACs*

The characteristics of DOAC-treated patients by type of study are described in Table 1. The DOAC-treated patients in prospective cohort studies were similar to those included in RCTs regarding disease severity, the median age being 58 and 56 years, respectively (Table 1). In contrast, compared to the patients enrolled in RCTs, the DOAC-treated patients included in retrospective cohort studies or studies using living health administrative databases were 7 years older in terms of median age at baseline and presented more comorbidities.

### *Comparison of treatment effect estimates: prospective cohort studies versus RCTs*

The treatment effect of DOACs compared to VKA estimated in RCTs corresponded to a non-significant decrease of 11% (HR 0.89 [95% CI 0.76 to 1.05]) in the risk of recurrent VTE and a significant reduction of 39% (HR 0.61, [95% CI, 0.45 to 0.83]) in the risk of major bleeding (Figures 1 and 2). The treatment effect of DOACs estimated in prospective cohort studies corresponded to a non-significant decrease of 40% in the risk of recurrent VTE and a significant decrease of 43% in the risk of major bleeding (HR 0.60, [95% CI, 0.35 to 1.02] and HR 0.57, [95% CI, 0.37 to 0.89], respectively) (Figures 1 and 2).

Compared to RCTs, prospective cohort studies were associated with a non-significant 33% increase in efficacy estimate (RHR 0.67, [95% CI, 0.39 to 1.18]) (Figure 3). This result was

unchanged after exclusion of studies that did not use any methods of adjustment for confounders [33,34] (RHR 0.93, [95% CI, 0.57 to 1.53]). As regards the safety endpoint, no association was found between the type of study and the treatment effect (RHR 0.93, [95% CI, 0.55 to 1.60]) (Figure 3), a similar absence of any association being observed after the exclusion of studies not including any methods of adjustment for cofounders [35,36] (RHR 1.30, [95% CI, 0.64 to 2.61]).

*Comparison of treatment effect estimates: studies using living databases versus RCTs*

In studies using living databases, treatment with DOACs was associated with a 27% decrease in the efficacy endpoint (HR 0.73, [95% CI, 0.63 to 0.83]) and a 19% decrease in the safety endpoint (HR 0.81, [95% CI, 0.72 to 0.90]) (Figures 2 and 3). Compared to RCTs, studies using living databases showed a non-significant 18% increment in estimated treatment effect with respect to the efficacy endpoint (RHR 0.82, [95% CI, 0.66 to 1.01]) (Figure 3); the magnitude of this increment was similar after the exclusion of studies that did not use any method of adjustment for confounders [35] (RHR 0.82, [95% CI, 0.67 to 1.01]). This trend was reversed as regards the safety endpoint (RHR 1.33, [95% CI, 0.96 to 1.84]) (Figure 3), this result remaining unchanged even after the exclusion of studies that did not use any method of adjustment for confounders [35].

*Comparison of treatment effect estimates: retrospective cohort studies versus RCTs*

In retrospective cohort studies, DOACs were associated with a 70% decrease in the risk of VTE (HR 0.30, [95% CI, 0.03 to 0.97]) with no difference in the risk of major bleeding (HR 1.03, [95% CI, 0.40 to 2.66]) (Figures 1 and 2). Compared to RCTs, retrospective cohort studies showed a non-significant exaggeration of the estimated treatment effect as regards the efficacy endpoint (RHR 0.34, [95% CI, 0.10 to 1.12]) (Figure 3). In contrast, with respect to the safety endpoint, the estimated risk with DOACs compared to VKA was 69% higher than in RCTs although this difference was not significant (RHR 1.69, [95% CI, 0.62 to 4.57]) (Figure 3). None of the retrospective cohort studies used any method of adjustment for confounders.

*Comparison of treatment effect estimates: observational studies versus RCTs*

Overall, compared to RCTs, observational studies tended to overestimate the treatment effect in terms of the efficacy endpoint (RHR 0.80, [95% CI, 0.65 to 0.98]), but to reduce the estimated treatment effect as regards the safety endpoint (RHR 1.31, [95% CI, 0.95 to 1.81]) (Figure 3).

*Detection of publication bias and small-study effect*

The funnel plots presenting the analysis of recurrent VTE and major bleeding are shown in Figure 4A and 4B, respectively. The asymmetrical plots plead for an apparent overestimation of the treatment effect with respect to both efficacy and safety endpoints, but the Egger test was not significant.

## **DISCUSSION**

In this meta-epidemiological study, even though the patients included in prospective cohort studies were close to those included in RCTs in the defined indication, these studies overestimated the magnitude of the treatment effect for both efficacy and safety endpoints compared to RCTs. In contrast, the patients included in retrospective studies and studies using living databases were closer to the real-life setting, with patients about 10 years older than those included in RCTs. However, we also observed a non-significant exaggeration of the efficacy estimate but an underestimation of the safety estimate in these studies.

To the best of our knowledge, this is the first meta-epidemiological study to differentiate subcategories of observational studies based on their design in terms of recruitment and data collection, the aim being to test the consistency of the results. In prospective cohort studies, the choice of collecting potential confounding factors is at the discretion of the investigator, a protocol is generally registered and is freely accessible, minimizing the possibility of p-hacking; in contrast, exhaustivity is not guaranteed and this may explain the possible selection of patients as in RCTs, and also why patient characteristics are superimposable on the characteristics of patients included in RCTs. In the indication evaluated, it seems that the implementation of prospective cohort studies did not lead to the inclusion of less selected patients compared to RCTs, as the same clinical research sites agreed to participate in both RCTs and prospective cohort studies. Among the different types of observational study analyzed, prospective cohort studies yielded the most favorable estimates of treatment effect compared to RCTs.

In contrast, studies using living databases have the advantage of being exhaustive but do not allow flexibility with regard to the variables analyzed, as this is necessarily constrained by the range of variables already collected in the database. In such studies, a lesser selection of the patient population is observed, as already shown in other clinical settings. Indeed, previous meta-epidemiological analyses comparing RCTs and observational studies also found systematic differences in the baseline characteristics of the patients included, similarly showing that the patients included in RCTs were younger and presented a shorter disease duration compared to those enrolled in observational studies [48].

Finally, retrospective cohort studies permit neither exhaustivity nor the choice of confounding factors to be measured and should not be considered as having an acceptable design for pharmaco-epidemiological purposes.

Whatever the population enrolled, this meta-epidemiological study shows that observational studies are associated with a statistically significant 20% increase in treatment efficacy compared to RCTs. Basically, pharmaco-epidemiological studies are encouraged as under real-life conditions, less optimistic estimates of the treatment effects than those observed in RCTs are expected; when the opposite is observed, residual confounding cannot be excluded, and the results obtained are questionable. In practice, such observational studies may be used as promotional arguments on the basis of falsely promising results, and no longer play their initial role of vigilance. As a first step, it is important to estimate the magnitude of the overestimation of the treatment effect, which is an indirect estimate of residual confounding. One implication of such findings could be the necessity "to correct" treatment effect estimates derived from observational studies, rather than trusting falsely reassuring results.

Several meta-analyses have compared the treatment effect estimates obtained in RCTs and observational studies employing a PS in different clinical settings, with the conclusion that the estimates of the latter studies were in agreement with those of RCTs [49-51]. However, the results were the same as those presented in this study, i.e. an enhancement of the treatment effect by observational studies compared to RCTs.

Regarding safety, in contrast, whatever the study design, observational studies were associated with a non-significant decrease of about 30% in the magnitude of the effect of DOACs effect. Several hypotheses can be put forward to try to explain this finding. First, controlling for confounding (using adjustment or matching) might for the most part be based on variables that are primarily prognostic factors for efficacy. Lack of data on specific variables related to bleeding risk is an issue in these studies, so calculating a reliable and complete PS is difficult. Second, in unselected patients, we cannot exclude a less promising result than that observed in observational studies, and that is exactly the role we expect these studies to play. The difficulty is that we cannot retain either the first or the second hypothesis and should therefore believe in the most pessimistic information.

This meta-epidemiological study has some limitations. The first is a lack of statistical power to identify differences in RHR between subcategories of observational studies. The same type of research should be performed in other clinical settings so as to obtain precise and conclusive results for each type of observational study. The second limitation is that it is impossible to impute the difference between RCTs and observational studies to either a genuine difference

between RCTs and real-world studies or a residual confounding which is inevitable in a real-life setting. Even when using adjustment or matching techniques, observational studies would hardly be capable of replicating the results of RCTs. They should only be used if their estimated results are not exaggerated compared to those of RCTs. These results should probably be adjusted using a correction factor that remains to be precisely estimated.

In all cases, even though not perfect, observational studies could be useful for investigating comparative effectiveness, in which case a supposed “over” or “under” estimation could be comparable between similar drugs, as well as for evaluating the risk of rare or unexpected adverse events, not captured in RCTs.

### **CONCLUSION**

The population included in prospective cohort studies was similar to that enrolled in RCTs. However, the population included in other types of observational study seemed to be more representative of patients treated in “real-life”. Overall, in the clinical setting concerned, an exaggeration of the treatment efficacy estimate was observed in observational studies compared to RCTs. Safety results, as expected in real-world patients, were less promising when estimated in observational studies. In all cases, the presence of residual confounding cannot be excluded in observational studies and their results should therefore be interpreted with caution.

**Declaration of competing interest:** None

### **REFERENCES**

- [1] Moustafa F, Pesavento R, di Micco P, González-Martínez J, Quintavalla R, Peris M-L, et al. Real-life Use of Anticoagulants in Venous Thromboembolism with a Focus on Patients with Exclusion Criteria for Direct Oral Anticoagulants. *Clin Pharmacol Ther.* 2018 Apr;103(4): 684-91.
- [2] Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?” *Lancet.* 2005 Jan 1-7;365(9453):82-93.
- [3] Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. *JAMA.* 1992 Sep 16;268(11):1417-22.

-

assessing the influence of study characteristics on treatment effects in "meta epidemiological"

Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported



- [18] Sindet-Pedersen C, Langtved Pallisgaard J, Staerk L, Gerds TA, Fosbøl EL, Torp-Pedersen C, et al. Comparative safety and effectiveness of rivaroxaban versus VKAs in patients with venous thromboembolism. A Danish nationwide registry based study. *Thromb Haemost.* 2017 Jun 2;117(6):1182-91.
- [19] Larsen TB, Skjøth F, Kjældgaard JN, Lip GYH, Nielsen PB, Søgaard M. Effectiveness and safety of rivaroxaban and warfarin in patients with unprovoked venous thromboembolism: a propensity-matched nationwide cohort study. *Lancet Haematol.* 2017 May;4(5): e237-e244.
- [20] Coleman CI, Turpie AGG, Bunz TJ, Baker WL, Beyer-Westendorf J. Effectiveness and safety of outpatient rivaroxaban versus warfarin for treatment of venous thromboembolism in patients with a known primary hypercoagulable state. *Thromb Res.* 2018 Mar;163:132-7.
- [21] Kohn CG, Bunz TJ, Beyer-Westendorf J, Coleman CI. Comparative risk of major bleeding with rivaroxaban and warfarin: Population-based cohort study of unprovoked venous thromboembolism. *Eur J Haematol.* 2019 Feb;102(2):143–9.
- [22] Coleman CI, Bunz TJ, Turpie AGG. Effectiveness and safety of rivaroxaban versus warfarin for treatment and prevention of recurrence of venous thromboembolism. *Thromb Haemost.* 2017 Oct 5;117(10):1841-7.
- [23] Coleman CI, Peacock WF, Bunz TJ, Beyer-Westendorf J. Effectiveness and safety of rivaroxaban versus warfarin in patients with unprovoked venous thromboembolism: A propensity-score weighted administrative claims cohort study. *Thromb Res.* 2018 Aug;168: 31-6.
- [24] Coleman CI, Turpie AGG, Bunz TJ, Beyer-Westendorf J. Effectiveness and safety of rivaroxaban versus warfarin in patients with provoked venous thromboembolism. *J Thromb Thrombolysis.* 2018 Oct;46(3):339-45.
- [25] Schulman S, Kearon C, Kakkar AK, Mismetti P, Schellong S, Eriksson H, et al. Dabigatran versus Warfarin in the Treatment of acute Venous Thromboembolism. *N Engl J Med.* 2009 Dec 10;361(24):2342-52.
- [26] EINSTEIN Investigators, Bauersachs R, Berkowitz SD, Brenner B, Buller HR, Decousus H, et al. Oral rivaroxaban for symptomatic venous thromboembolism. *N Engl J Med.* 2010 Dec 23;363(26):2499-510.
- [27] EINSTEIN–PE Investigators, Büller HR, Prins MH, Lensin AW, Decousus H, Jacobson BF, et al. Oral Rivaroxaban for the Treatment of Symptomatic Pulmonary Embolism. *N Engl J Med.* 2012 Apr 5;366(14):1287-97.
- [28] Agnelli G, Buller HR, Cohen A, Curto M, Gallus AS, Johnson M, et al. Oral apixaban for the treatment of acute venous thromboembolism. *N Engl J Med.* 2013 Aug 29;369(9):799-808.
- [29] Hokusai-VTE Investigators, Büller HR, Décousus H, Grosso MA, Mercuri M, Middeldorp S, et al. Edoxaban versus warfarin for the treatment of symptomatic venous thromboembolism. *N Engl J Med.* 2013 Oct 10;369(15):1406-15.

- [30] Schulman S, Kakkar AK, Goldhaber SZ, Schellong S, Eriksson H, Mismetti P, et al. Treatment of Acute Venous Thromboembolism with Dabigatran or Warfarin and Pooled Analysis. *Circulation*. 2014 Feb 18;129(7):764-72.
- [31] Ageno W, Mantovani LG, Haas S, Kreutz R, Monje D, Schneider J, et al. Safety and effectiveness of oral rivaroxaban versus standard anticoagulation for the treatment of symptomatic deep-vein thrombosis (XALIA): an international, prospective, non-interventional study. *Lancet Haematol*. 2015 Jan;3(1):e12-21.
- [32] Kucher N, Aujesky D, Beer JH, Mazzolai L, Baldi T, Banyai M, et al. Rivaroxaban for the treatment of venous thromboembolism the swiss venous thromboembolism registry (SWIVTER). *Thromb Haemost*. 2016 Aug 30;116(3):472-9.
- [33] Kreutz R, Mantovani LG, Haas S, Monje D, Schneider J, Bugge J-P, et al. XALIA-LEA: An observational study of venous thromboembolism treatment with rivaroxaban and standard anticoagulation in the Asia-Pacific, Eastern Europe, the Middle East, Africa and Latin America. *Thromb Res*. 2019 Apr;176:125-32.
- [34] Bui MH, Son NT, Viet PT, Hiep NH, Dinh TC. Oral Rivaroxaban Versus Standard Therapy in Acute Venous Thromboembolism Treatment for Vietnamese Patients. *Open Access Maced J Med Sci*. 2019 Dec 20;7(24):4255-9.
- [35] Gaertner S, Cordeanu E-M, Nouri S, Faller A-M, Frantz A-S, Mirea C, et al. Rivaroxaban versus standard anticoagulation for symptomatic venous thromboembolism (REMOTEV observational study): analysis of 6-month outcomes. *Inter J Cardiol*. 2017 Jan 1;226:103-9.
- [36] Bounameaux H, Haas S, Farjat AE, Ageno W, Weitz JI, Goldhaber SZ, et al. Comparative effectiveness of oral anticoagulants in venous thromboembolism: GARFIELD VTE. *Thromb Res*. 2020 Jul;191:103-112.
- [37] Jun M, Lix LM, Durand M, Dahl M, Paterson JM, Dormuth CR, et al. Comparative safety of direct oral anticoagulants and warfarin in venous thromboembolism: multicentre, population based, observational study. *BMJ*. 2017 Oct 17;359:j4323.
- [38] Wang L, Baser O, Wells P, Peacock WF, Coleman CI, Fermann GJ, et al. Overall Effectiveness of Rivaroxaban in Patients with Pulmonary Embolism. *Clin Ther*. 2017 Jul ;39(7):1426-1436.e2.
- [39] Weycker D, Li X, Wygant GD, Lee T, Hamilton M, Luo X, et al. Effectiveness and Safety of Apixaban versus Warfarin as Outpatient Treatment of Venous Thromboembolism in U.S. Clinical Practice. *Thromb Haemost*. 2018 Nov;118(11):1951-61.
- [40] Costa OS, Thompson S, Ashton V, Palladino M, Bunz TJ, Coleman CI. Rivaroxaban versus warfarin for treatment and prevention of recurrence of venous thromboembolism in African American patients: a retrospective cohort analysis. *Thromb J*. 2020 Apr 8;18:6.

- [41] Dawwas GK, Smith SM, Dietrich E, Lo-Ciganic W-H, Park H. Comparative effectiveness and safety of apixaban versus warfarin in patients with venous thromboembolism. *Am J Health Syst Pharm*. 2020 Jan 24;77(3):188-95.
- [42] Baviera M, Barcella L, Vannini T, Colacioppo P, Marzona I, Tettamanti M, et al. Treatment of venous thromboembolism in Northern Italy: A population-based study from 2013 to 2018. *Thromb Res*. 2020 Apr;188:97-9.
- [43] Guo JD, Rajpura J, Hlavacek P, Keshishian A, Sah J, Delinger R, et al. Comparative Clinical and Economic Outcomes Associated with Warfarin Versus Apixaban in the Treatment of Patients with Venous Thromboembolism in a Large U.S. Commercial Claims Database. *J Manag Care Spec Pharm*. 2020 Aug;26(8):1017-1026.
- [44] Badreldin H. Hospital length of stay in patients initiated on direct oral anticoagulants versus warfarin for venous thromboembolism: a real-world single-center study. *J Thromb Thrombolysis*. 2018 Jul;46(1):16-21.
- [45] Basto AN, Fewel NP, Vo K, Stock EM, Ta M. Initiation of direct oral anticoagulants versus warfarin for venous thromboembolism: impact on time to hospital discharge. *J Thromb Thrombolysis*. 2018 Jan;45(1):51-5.
- [46] Wakakura S, Hara F, Fujino T, Hamai A, Ohara H, Kabuki T, et al. Comparison of Direct Oral Anticoagulants and Warfarin in the Treatment of Deep Venous Thrombosis in the Chronic Phase. *Int Heart J*. 2018 Jan 27;59(1):126-35.
- [47] Fung KP, Chan KH, Ng V, Tsui PT, You JHS. Health Economic Analysis of Rivaroxaban and Warfarin for Venous Thromboembolism Management in Chinese Patients. *Cardiovasc Drugs Ther*. 2019 Jun;33(3):331-7.
- [48] Kilcher G, Hummel N, Didden EM, Egger M, Reichenbach S. Rheumatoid arthritis patients treated in trial and real world settings: comparison of randomized trials with registries. *Rheumatology (Oxford)*. 2018 Feb 1;57(2):354-69.
- [49] Lonjon G, Boutron I, Trinquart L, Ahmad N, Aim F, Nizard R, et al. Comparison of treatment effect estimates from prospective nonrandomized studies with propensity score analysis and randomized controlled trials of surgical procedures. *Ann Surg*. 2014 Jan;259(1): 18-25.
- [50] Li G, Holbrook A, Jin Y, Zhang Y, Levine MAH, Mbuagbaw L, et al. Comparison of treatment effect estimates of non-vitamin K antagonist oral anticoagulants versus warfarin between observational studies using propensity score methods and randomized controlled trials. *Eur J Epidemiol*. 2016 Jun;31(6):541-61.
- [51] Kuss O, Legler T, Borgermann J. Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. *J Clin Epidemiol*. 2011 Oct;64(10):1076–84.

**Table 1.** Characteristics of DOAC-treated patients by type of study before any adjustment or matching

	Randomized controlled trials (6 studies)	Prospective cohort studies (4 studies)	Studies using living databases (12 studies)	Retrospective cohort studies (4 studies)
Number of studies with available information	6	4	6	4
Age, years	56 (55-57)	58 (57-59)	63 (60-64)	63 (63-65)
Male gender (%)	58 (57-58)	52 (45-55)	53 (50-55)	50 (39-69)
Renal insufficiency (%) median (IQR)	7 (7-7)	8 (6-9)	12 (9-20)	12 (12-20)
History of VTE (%)	19 (19-19)	17 (10-23)	17 (15-23)	10 (6-14)
Cancer (%)	5 (4-6)	8 (6-11)	11 (4-14)	17 (8-25)
PE at inclusion (%)	33 (31-39)	23 (13-33)	45 (41-90)	35 (30-45)

\*only one study. DOACs: direct oral anticoagulants; VKA: vitamin K antagonist; PE: pulmonary embolism; IQR: inter-quartile range; PS: propensity score; NA: Not available.

Figure 1. Comparison of treatment effect estimates (Hazard Ratio) for recurrent VTE between randomized controlled trials and observational studies.

Figure 2. Comparison of treatment effect estimates (Hazard Ratio) for major bleeding between randomized controlled trials and observational studies.

Figure 3. Quantification of the differences in estimated treatment effect between observational studies and RCTs based on the ratio of hazard ratios (RHR).

Legend: HAD: health administrative databases; 95% CI: 95% confidence interval; RCTs: randomized controlled trials; RHR: ratio of Hazard Ratios; VTE: venous thromboembolism; DOACs: direct oral anticoagulants; VKA: vitamin-K antagonists

\* A RHR < 1 would imply that the estimates of treatment effect in observational studies were higher compared to RCTs, whereas a RHR > 1 would imply that the estimates of treatment effect were lower.

Figure 4. Funnel plot of included studies for recurrent VTE (A) and major bleeding (B).

Legend: Circles: RCTs; squares: prospective cohort studies; triangles: studies using living databases, stars: retrospective cohort studies. Egger test for recurrent VTE and major bleeding:  $p = 0.60$  and  $p = 0.11$  respectively.

Figure 1.

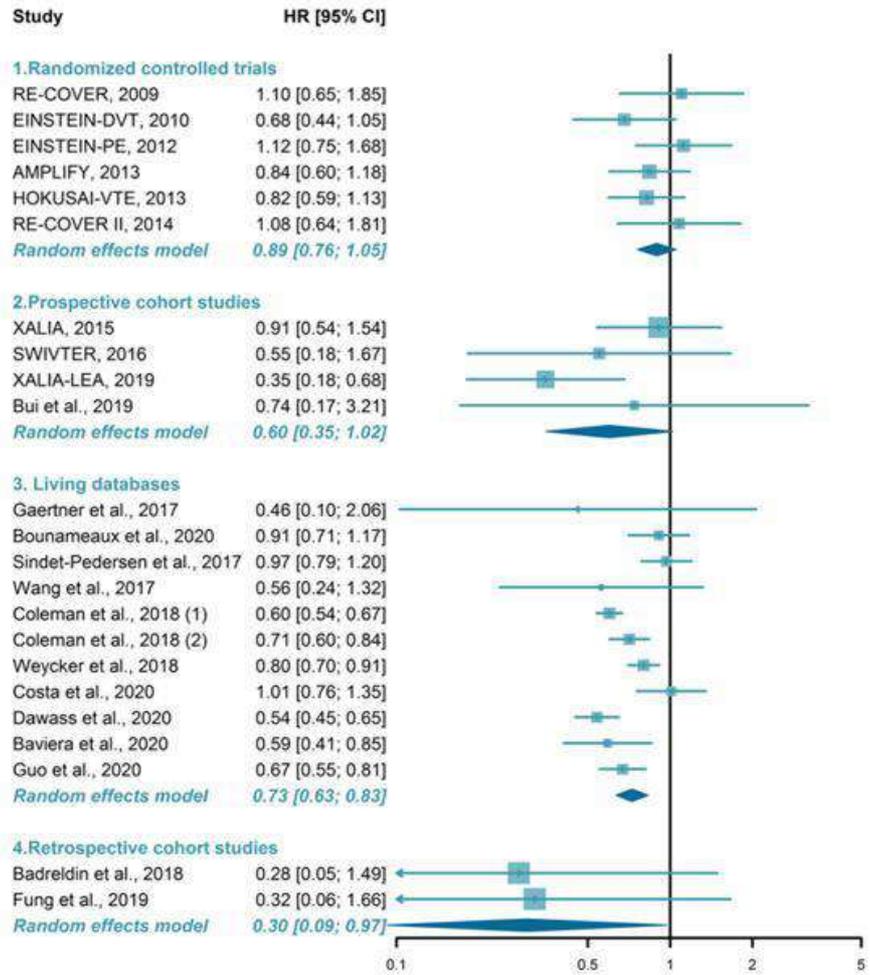


Figure 2.

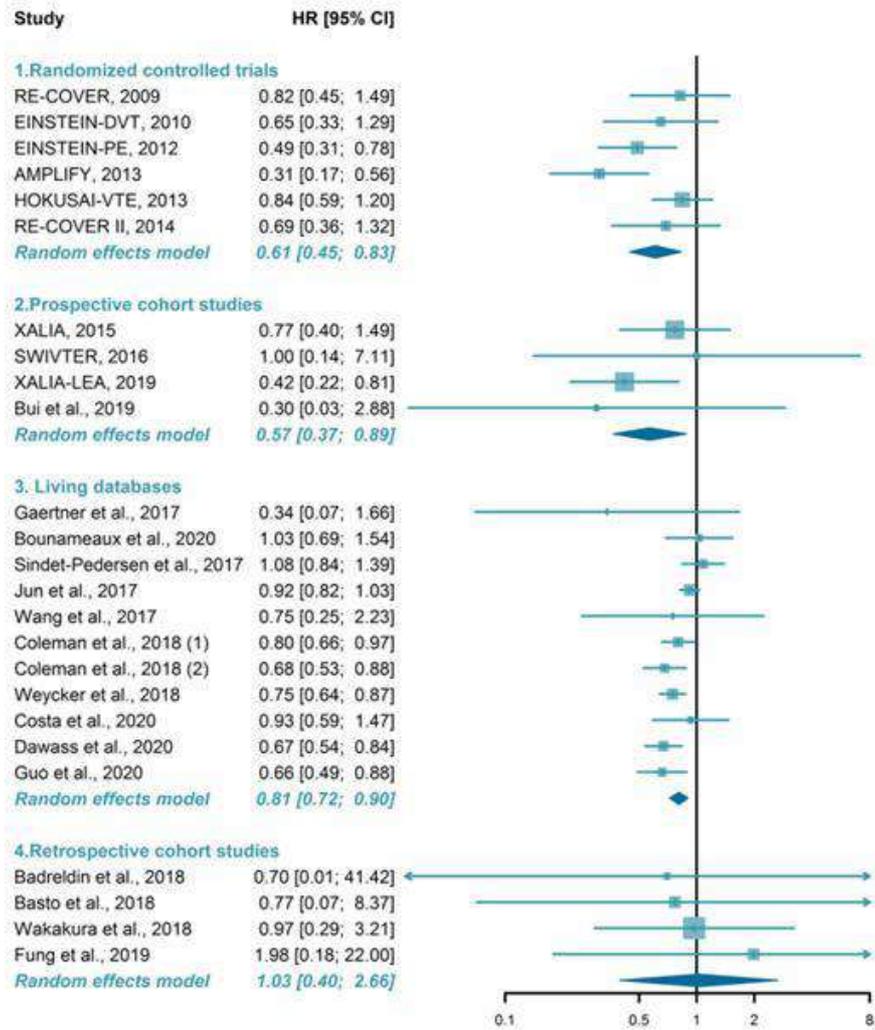


Figure 3.

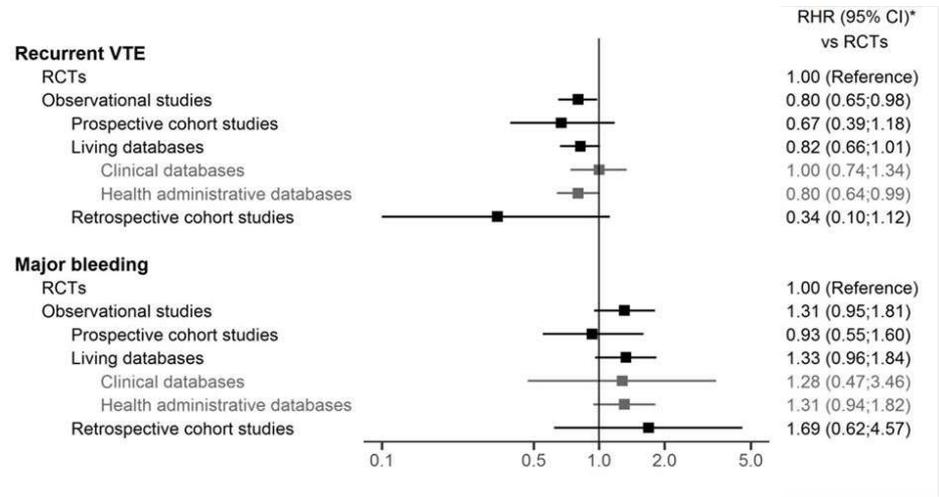
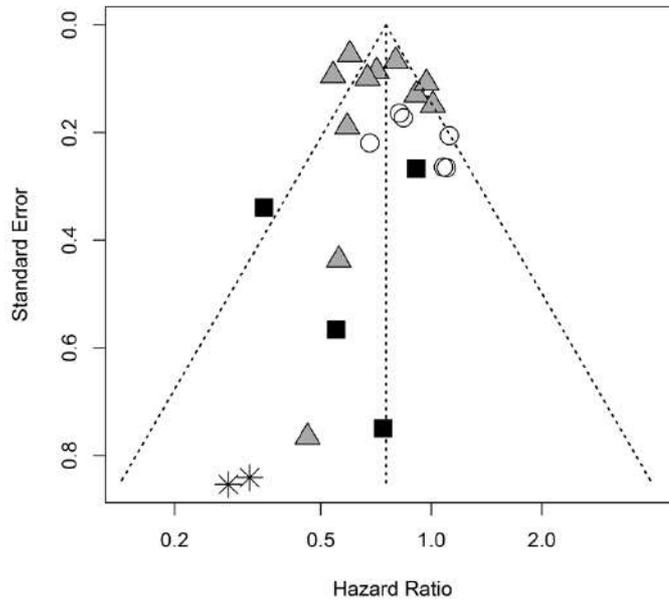


Figure 4.

A.



B.

