



HAL
open science

Development and validation of risk scores in vascular diseases

François-Xavier Lapebie

► **To cite this version:**

François-Xavier Lapebie. Development and validation of risk scores in vascular diseases. Human health and pathology. Université Paul Sabatier - Toulouse III, 2021. English. NNT : 2021TOU30200 . tel-03663091

HAL Id: tel-03663091

<https://theses.hal.science/tel-03663091v1>

Submitted on 9 May 2022

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.



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier**

**Présentée et soutenue par
François-Xavier LAPEBIE**

Le 9 décembre 2021

**Développement et validation de scores de risque dans les
maladies vasculaires**

Ecole doctorale : **EDMITT - Ecole Doctorale Mathématiques, Informatique et
Télécommunications de Toulouse**

Spécialité : **Epidémiologie**

Unité de recherche :
**CERPOP - Centre d'Epidémiologie et de Recherche en santé des POPulations
de Toulouse**

Thèse dirigée par
Vanina BONGARD

Jury

M. Francis COUTURAUD, Rapporteur
Mme Nelly AGRINIER, Rapporteur
M. Jean FERRIERES, Examineur
Mme Alessandra BURA-RIVIERE, Examinatrice
Mme Vanina BONGARD, Directrice de thèse
M. Philippe LACROIX, Président

Développement et validation de scores de risque dans les maladies vasculaires

Remerciements

Je remercie chaleureusement le Pr Vanina BONGARD, pour son encadrement lors de cette thèse qu'elle me fait l'honneur de diriger, mais aussi pour son soutien depuis maintenant 5 ans, depuis mon inscription au M1R de santé publique qu'elle coordonne, suivi du M2R d'épidémiologie dont elle avait accepté d'encadrer le mémoire. Merci de m'avoir fait découvrir l'épidémiologie. Je la remercie également pour son implication dans la recherche clinique au sein du pôle cardiovasculaire et métabolique du CHU de Toulouse, dont notre service de médecine vasculaire bénéficie grandement. Je suis très reconnaissant de son accueil au sein du CERPOP, dont j'ai pu apprécier le dynamisme, et plutôt que sa conclusion, j'espère que ce travail sera le départ de ma collaboration.

J'exprime ma profonde gratitude au Pr Nelly AGRINIER et au Pr Francis COUTURAUD, pour me faire l'honneur d'être les rapporteurs de cette thèse. J'imagine le travail supplémentaire que cela doit représenter dans des emplois du temps surement chargés. Je suis très heureux que le Pr AGRINIER puisse juger en particulier le travail réalisé sur le registre COPART, étant donné son expertise, outre en épidémiologie, sur les cohortes et les essais cliniques dans les pathologies cardiovasculaires. Le Pr COUTURAUD est internationalement reconnu pour son expertise dans la maladie thrombo-embolique veineuse, et son équipe contribue largement à l'épidémiologie de la pathologie en France avec les études EPIGETBO. Je suis donc très honoré qu'il puisse juger le travail réalisé sur le registre RIETE.

Je remercie tout spécialement le Pr Alessandra BURA-RIVIERE, pour sa participation au jury de cette thèse bien sûr, mais au-delà, pour m'avoir recruté en 2015 dans son service, poussé rapidement sur la voie universitaire, et soutenu depuis, malgré mes (petits) défauts de caractère et un humour parfois inapproprié... J'espère que le Pr LACROIX vous avait prévenu. Même si je ne le dis pas souvent (jamais en fait), merci pour votre bienveillance, pour votre enseignement, pour m'avoir permis d'intégrer le CEMV, merci de m'avoir orienté vers l'épidémiologie et le Pr BONGARD... Et surtout (je sais que vous n'aimez pas), félicitations pour la une du Point !

Le Pr Philippe LACROIX a bien voulu participer à ce jury, ce à quoi je tenais beaucoup, et je l'en remercie profondément. J'ai découvert la médecine vasculaire avec lui au CHU de Limoges. Comme pour l'épidémiologie, il s'agissait de voies que je n'avais pas envisagées de prime abord, mais qui se sont révélées passionnantes et complémentaires. J'ai pu constater lors de ma formation ses grandes qualités, professionnelles et humaines, sa polyvalence, son implication dans l'enseignement et la recherche, et bien sûr sa passion contagieuse du marathon, à laquelle j'aurais su résister comme à la majorité des activités physiques. Ses travaux sur l'épidémiologie de la maladie athéromateuse, que ce soit en France ou en Afrique, son implication dans COPART dont il est l'un des créateurs, son expertise reconnue en médecine vasculaire, le rendent tout indiqué pour juger ce travail

Parmi de nombreuses activités, le Pr Jean FERRIERES dirige l'axe ARTERRE, au sein de l'équipe vieillissement du CERPOP, et je lui suis très reconnaissant de prendre le temps de participer à ce jury. Ses travaux sur la pathologie athéromateuse sont largement connus, tout comme son expertise en épidémiologie. Je suis très honoré d'être le second médecin vasculaire, après le Pr Victor ABOYANS dont il avait dirigé le travail, à réaliser une thèse dans son unité. Je le remercie pour son accueil et sa gentillesse. J'espère pouvoir continuer à collaborer avec lui et le Pr BONGARD dans le cadre du CERPOP.

Je remercie toutes les personnes avec lesquelles j'ai pu collaborer dans le cadre des registres COPART et RIETE. Merci au Pr Joël CONSTANS, de Bordeaux, avec qui j'ai eu la chance de réaliser mon stage inter-CHU, au Pr Victor ABOYANS, notamment pour ses relectures exigeantes mais toujours pertinentes des articles issus de COPART, merci au Pr Manuel MONREAL de m'avoir donné accès au registre RIETE qu'il a créé et au Pr Laurent BERTOLETTI, qui le coordonne pour la France. Merci bien sûr aux assistantes et infirmières de recherche, notamment Delphine BAUDOIN et Brigitte PELVET pour Toulouse, pour la qualité de leur travail et du recueil des données.

Au moment de terminer cette thèse, outre mes collègues actuels du service de médecine vasculaire du CHU Rangueil surtout ceux qui partent (notamment le Dr Thomas JAMMES, le Dr Arthur THERY, et le Dr Lucas VERSET), j'ai une pensée pour mes anciens collègues du

service de médecine interne A et polyclinique médicale du CHU Dupuytren de Limoges : le Dr Holy BEZANAHARY, qui a notamment dirigé ma thèse de médecine et qui me fait l'amitié de se déplacer pour la soutenance de celle-ci, le Pr Elisabeth VIDAL, qui m'a orienté vers la médecine vasculaire (« - Bon, t'as réfléchi au DESC ? », « - Je pensais à maladies infectieuses ou immunologie clinique... », « - Non ! Il y a trop d'internistes qui font maladies infectieuses. Et immunologie clinique, ça ne sert pas pour nous. Tu vas faire médecine vasculaire ! C'est bien ça, et en plus ça t'apprend l'écho. J'en ai parlé à Philippe, il est d'accord. », « - Euh...OK. »), le Dr Paul CARRIER, le Dr Anne CYPierre, le Pr Anne-Laure FAUCHAIS, le Dr Guillaume GONDRAN, le Dr Eric LIOZON, le Pr Véronique LOUSTAUD-RATTI, le Pr Kim LY, le Dr Clothilde MARTEL, le Dr Sylvie NADALON, et le Dr Sylvain PALAT.

Je ne voulais pas dépasser une page pour les remerciements, car ça devient vite ridicule et surtout gênant pour ceux qui les lisent (je souhaite qu'ils soient peu nombreux), mais je suis obligé d'exprimer ma reconnaissance à mon épouse, Laetitia, en plus de mes excuses pour le temps pris par ma thèse, à mes filles, Margaux et Alix même si elles ne m'auront pas vraiment aidé pour ce travail, à mes parents, Pierre et Marie-Christine. Je ne remercie pas mes frères (Nicolas, Thomas, Arnaud et Jean-Baptiste), le reste de ma famille et belle-famille car ça n'aurait pas tellement de sens pour une thèse, mais au moins les voilà cités.

Table des matières

1. Objectifs et structure de la thèse	12
2. Modélisation en médecine	13
2.1. But d'un modèle	14
2.1.1. Modèle descriptif.....	14
2.1.2. Modèle explicatif.....	14
2.1.3. Modèle prédictif	14
2.2. Sélection de la base de données pour la dérivation du modèle	15
2.3. Gestion des données manquantes	16
2.4. Choix du type de modèle	21
2.4.1. Le modèle de Cox	21
2.4.2. Le modèle de Fine et Gray	24
2.5. Sélection des variables à inclure.....	25
2.5.1. Raisons de la sélection des variables, et nombre maximal de variables explicatives	25
2.5.2. Sélection des variables par un algorithme à partir des données	27
2.5.3. Sélection des variables sur une connaissance préalable, formalisée par un graphe acyclique dirigé	29
2.5.4. Cas particulier de la sélection sur des données issues d'une imputation multiple	33
2.6. Evaluation de la performance du modèle	34
2.6.1. Mesures de la performance du modèle.....	34
2.6.1.1. Calibration.....	35
2.6.1.2. Discrimination.....	36
2.6.1.3. Autres paramètres estimant la performance du modèle	37
2.6.2. Validation interne	38
2.6.2.1. Méthodes habituelles	38

2.6.2.2.	Cas particulier de la validation interne sur des données issues d'une imputation multiple	41
2.6.3.	Validation externe	42
2.6.4.	Evaluation de l'impact de l'implémentation du modèle dans la pratique clinique	44
3.	Applications	45
3.1.	Cohortes utilisées.....	45
3.1.1.	L'artériopathie athéromateuse des membres inférieurs et la cohorte COPART	45
3.1.2.	La maladie thromboembolique veineuse et la cohorte RIETE	47
3.2.	Logiciel utilisé.....	49
3.3.	Différence entre modèle explicatif et modèle descriptif, et exemples de graphes acycliques dirigés	50
3.3.1.	Article "Mortality, Cardiovascular and Limb events in Patients with Symptomatic Lower Extremity Artery Disease and Diabetes"	50
3.3.1.1.	Manuscrit	50
3.3.1.2.	Résumé et remarques	83
3.3.2.	Article "Impact of angiotensin receptor blockers on mortality after hospitalization for symptomatic lower extremity artery disease"	83
3.3.2.1.	Manuscrit	83
3.3.2.2.	Résumé et remarques	111
3.4.	Validation externe des résultats d'essais contrôlés randomisés dans une cohorte : article "External applicability of the COMPASS and VOYAGER-PAD trials on patients with symptomatic lower extremity artery disease in France: the COPART registry"	112
3.4.1.	Manuscrit.....	112
3.4.2.	Résumé et remarques	155
3.5.	Imputation multiple des données manquantes par équations chaînées : article "Identification of factors associated with recurrence of cancer-associated venous thromboembolism after discontinuation of anticoagulant therapy: the RIETE registry"	156
3.5.1.	Manuscrit.....	156

3.5.2.	Résumé et remarques	184
3.6.	Validation interne d'un modèle prédictif par <i>bootstrap</i>	189
3.6.1.	Présentation du modèle multivarié	189
3.6.2.	Estimation de la performance apparente du modèle multivarié (performance sur la base de données initiales).....	192
3.6.3.	Tirage au sort avec remplacement des échantillons <i>bootstrap</i>	193
3.6.4.	Evaluation de la performance du modèle multivarié sur l'échantillon <i>bootstrap</i> (<i>bootstrap performance</i>)	195
3.6.5.	Evaluation de la performance du modèle multivarié avec coefficients β estimés sur l'échantillon <i>bootstrap</i> , sur la base de données initiale (<i>test performance</i>).....	196
3.6.6.	Calcul de l'optimisme du modèle multivarié estimé sur l'échantillon <i>bootstrap</i>	198
3.6.7.	Génération de 100 échantillons <i>bootstrap</i> et estimations des paramètres de la performance du modèle.....	198
3.6.8.	Calcul de l'optimisme et correction de la performance du modèle multivarié évalué dans la base de données initiale.....	201
4.	Conclusion	205
5.	Bibliographie.....	207

Liste des abréviations

95%CI : *95% confidence interval*

AAMI : Artériopathie athéromateuse des membres inférieurs

ABI : *Ankle-brachial index*

ACEI : *Angiotensin-converting-enzyme inhibitors*

AIC : *Akaike information criterion*

ALE : *Adverse limb event*

ALI : *Acute limb ischemia*

ANRS : Agence nationale pour la recherche sur le syndrome d'immunodéficience acquise et les hépatites virales

ARA2 : Antagoniste des récepteurs de l'angiotensine de type 2

ARB : *Angiotensin-receptor blockers*

AT1R : *AT1 subtype of the angiotensin II receptor*

AUC : *Area under the receiver-operating characteristic curve*

AVC : Accident vasculaire cérébral

BMI : *Body mass index*

BMS : Bristol-Myers Squibb

CAD : *Coronary artery disease*

CAT : *Cancer-associated thrombosis*

CépiDc : Centre d'épidémiologie sur les causes médicales de décès

CHMP : *Committee for medicinal products for human use*

CHU : Centre hospitalier universitaire

CKD-EPI : *Chronic kidney disease-epidemiology collaboration*

CLTI : *Chronic limb-threatening ischemia*

COMPASS : *Cardiovascular outcomes for people using anticoagulation strategies*

COPART : Cohorte de patients artériopathies

COPD : *Chronic obstructive pulmonary disease*

CRP : *C-reactive protein*

CV : *Cardiovascular*

CYP3A4 : *Cytochrome P450 isoenzyme 3A4*

DAG : *Directed acyclic graph*

DAPT : *Dual antiplatelet therapy*

DASH : *D-dimer, age, sex, hormonal therapy*

DFG : *Débit de filtration glomérulaire*

DGOS : *Direction générale de l'offre de soins*

DOA : *Direct oral anticoagulant*

DVT : *Deep vein thrombosis*

EGB : *Echantillon généraliste des bénéficiaires*

eGFR : *Estimated glomerular filtration rate*

EMA : *European medicines agency*

EP : *Embolie pulmonaire*

EPV : *Events per variable*

ESVS : *European society for vascular surgery*

ETEV : *Evènement thromboembolique veineux*

EUCLID : *Examining use of ticagrelor in peripheral artery disease*

FDA : *Food and drug administration*

GLP-1 : *Glucagon peptide-1*

HDL : *High-density lipoprotein*

HERDOO2 : *Hyperpigmentation, edema, or redness in either leg, D-dimer level ≥ 250 $\mu\text{g/L}$, obesity with body mass index ≥ 30 kg/m^2 , or older age, ≥ 65 years*

HIV : *Human immunodeficiency virus*

HOPE : *Heart outcomes prevention evaluation*

HR : *Hazard ratio*

IC : *Intermittent claudication*

IC95% : *Intervalle de confiance à 95%*

ICTHIC : *International conference on thrombosis and hemostasis issues in cancer*

IEC : *Inhibiteur de l'enzyme de conversion de l'angiotensine*

INSERM : *Institut national de la santé et de la recherche médicale*

IPS : *Indice de pression systolique*

IQR : *Interquartile range*

IS : *Ischemic stroke*

ISTH : *International society of thrombosis and haemostasis*

LASSO : *Least absolute shrinkage and selection operator*

LDL : *Low-density lipoprotein*

LEAD : *Lower extremity artery disease*

LE-DVT : *Lower extremity deep vein thrombosis*

LMWH : *Low molecular weight heparin*

LOCF : *Last observation carried forward*

MACE : *Major adverse cardiovascular events*

MALE : *Major adverse limb events*

MAR : *Missing at random*

MCAR : *Missing completely at random*

MI : *Myocardial infarction*

MICE : *Multiple imputation by chained equations*

MNAR : *Missing not at random*

MSPE : *Mean squared prediction error*

MTEV : *Maladie thromboembolique veineuse*

NA : *Not applicable*

NS : *Non-significant*

NYHA : *New York heart association*

OAD : *Oral antidiabetic drug*

ONTARGET : *Telmisartan alone and in combination with ramipril global endpoint trial*

OR : *Odds ratio*

PAD : *Peripheral artery disease*

PE : *Pulmonary embolism*

PESI : *Pulmonary embolism severity index*

P-gp : *P-glycoprotein*

PMM : *Predictive mean matching*

PS : *Propensity score*

PTS : *Post-thrombotic syndrome*

RAS : *Renin-angiotensin system*

RASI : *Renin-angiotensin system inhibitors*

RCT : *Randomized controlled trial*

REACH : *Reduction of atherothrombosis for continued health*

RIETE : *Registro Informatizado de la Enfermedad ThromboEmbolica venosa*

ROC : *Receiver operating characteristic*

RR : Risque relatif

RWD : *Real-world data*

SCA : Syndrome coronarien aigu

SD : *Standard deviation*

SFMV : Société française de médecine vasculaire

SGLT2 : *Sodium glucose transporter-2*

sHR : Subhazard ratio

SNDS : Système national des données de santé

SRAA : Système rénine angiotensine aldostérone

SVT : *Superficial vein thrombosis*

TBI : *Toe-brachial index*

TcPO₂ : Pression d'oxygène mesurée à travers la peau, ou *transcutaneous oxygen pressure*

TRIPOD : *Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis*

TVP : Thrombose veineuse profonde

TVS : Thrombose veineuse superficielle

UE-DVT : *Upper extremity deep vein thrombosis*

UFH : *Unfractionated heparin*

UMR : Unité mixte de recherche

VKA : *Vitamin K antagonist*

VOYAGER-PAD : *Vascular outcomes study of acetylsalicylic acid along with rivaroxaban in endovascular or surgical limb revascularization for peripheral artery disease*

VTE : *Venous thromboembolism*

WIFI : *Wound, Ischemia, and foot Infection*

XATOA : *Xarelto plus acetylsalicylic acid: treatment patterns and outcomes in patients with atherosclerosis*

1. Objectifs et structure de la thèse

En médecine en général, et notamment pour les pathologies cardiovasculaires, le choix d'un traitement et de sa durée ou d'un examen complémentaire se base sur la probabilité estimée par le clinicien d'une maladie donnée ou de survenue d'un évènement particulier (décès ou autre). Ces prédictions peuvent s'appuyer sur des modèles multivariés. Les méthodes statistiques utilisées pour leur développement et leur validation font l'objet de cette thèse.

La première partie de ce travail résume la méthodologie de la modélisation en médecine, en évoquant le but du modèle, le choix du type d'équation, la sélection des variables, et l'étude de sa performance. La prise en compte des données manquantes est aussi abordée.

La seconde partie regroupe plusieurs applications de ces méthodes aux cohortes COPART (COhorte de Patients ARTériopathes) et RIETE (*Registro Informatizado de la Enfermedad ThromboEmbolica venosa*).

2. Modélisation en médecine

Un modèle statistique est une représentation simplifiée de la réalité, dont l'objectif est de comprendre les causes à l'origine d'un phénomène. Un modèle peut ainsi décrire un phénomène, l'expliquer ou le prédire. Les modèles statistiques ont des applications variées dans de nombreux champs de recherche impliquant des données empiriques. En médecine, ils aident notamment à la prise de décision. Un modèle statistique multivarié relie en général une variable dépendante (variable d'intérêt, ou variable à expliquer), à plusieurs variables indépendantes (variables explicatives), et quantifie la force de ce lien [1]. Les registres représentent une des principales sources de dérivation de ces modèles.

Avant de développer un modèle statistique multivarié, plusieurs décisions doivent être prises, qui affecteront le modèle et donc les conclusions de la recherche [2]. Ces décisions concernent notamment :

- le but du modèle,
- la sélection de la base de données pour la dérivation du modèle,
- l'évaluation de la qualité des données et la gestion des valeurs manquantes dans les variables d'intérêt,
- le type de modèle : régression logistique, modèle des risques proportionnels de Cox...
- la sélection des variables indépendantes qui devront être évaluées pour l'inclusion dans le modèle final (variables candidates),
- la transformation éventuelle des variables quantitatives : transformation en variables catégorielles, transformation polynomiale,
- la stratégie de sélection des variables indépendantes pour le modèle final, à partir des variables candidates,
- l'évaluation de la qualité du modèle : robustesse en présence d'observations influentes, interaction entre les prédicteurs, correction d'un surajustement du modèle aux données ayant servies à le dériver (utilisation d'un facteur de contraction), étude de la stabilité (reproductibilité) du modèle, validation interne et validation externe du modèle.

Certains de ces points sont développés dans les sections suivantes, et leur application aux cohortes COPART et RIETE dans la partie 3 constitue l'objectif de cette thèse.

2.1. But d'un modèle

2.1.1. Modèle descriptif

Un modèle descriptif capture l'association entre une variable dépendante et des variables indépendantes, mais sans hypothèse sur un lien de causalité ; on estime des associations conditionnelles (« toutes choses égales par ailleurs ») [3].

L'identification des facteurs associés au risque d'amputation majeure du membre inférieur au cours de l'artériopathie athéromateuse des membres inférieurs (AAMI) dans le travail présenté en partie 3.3.1 fournit un exemple de modèle descriptif.

2.1.2. Modèle explicatif

Un modèle explicatif teste une hypothèse causale entre un groupe de variables indépendantes, et la variable dépendante [3]. On évalue donc l'effet causal d'une exposition (facteur de risque supposé) à l'aide de l'estimation d'effets ajustés, en général sur la base d'études observationnelles.

Au cours de l'AAMI, l'étude du lien de causalité entre la prescription d'un antagoniste des récepteurs de l'angiotensine de type 2 (ARA2) et le décès, dans le travail présenté en partie 3.3.2, ainsi que l'étude du lien de causalité entre le diabète et les événements cardiovasculaires (MACE : *major adverse cardiovascular events*), dans le travail présenté en partie 3.3.1, sont des exemples de modèles explicatifs.

2.1.3. Modèle prédictif

Un modèle prédictif multivarié prédit, pour un individu donné, la survenue de la variable dépendante (événement d'intérêt) ou sa valeur, en fonction des variables indépendantes, ou prédicteurs [2,3]. Les modèles prédictifs multivariés se répartissent en modèles diagnostiques et modèles pronostiques [4]. Les modèles diagnostiques estiment la probabilité qu'une maladie soit présente au moment de la prédiction chez un sujet donné. Les modèles pronostiques estiment la probabilité qu'un événement survienne au cours d'une période donnée dans le futur. Les modèles diagnostiques sont développés à partir d'études épidémiologiques transversales, alors que les modèles pronostiques s'appuient sur les études épidémiologiques longitudinales [4].

Des recommandations sur la dérivation et la validation d'un modèle prédictif en médecine, dénommées TRIPOD (*Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis*) ont été édictées en 2015 [4].

Un modèle prédictif a été dérivé à partir de la cohorte RIETE pour prédire le risque de récurrence des événements thromboemboliques veineux (ETE) à l'arrêt du traitement anticoagulant chez les patients présentant un cancer dans la partie 3.6.

2.2. Sélection de la base de données pour la dérivation du modèle

Les sources de données de santé sont multiples :

- bases de données médico-administratives : centre d'épidémiologie sur les causes médicales de décès (CépiDc), échantillon généraliste des bénéficiaires (EGB), système national des données de santé (SNDS),
- études observationnelles,
- études expérimentales, dont les essais cliniques.

Les essais contrôlés randomisés apportent le plus haut niveau de preuve sur l'efficacité et les effets secondaires éventuels d'une intervention médicale (médicament ou autre). La population incluse ainsi que le déroulement du protocole sont strictement contrôlés afin de maximiser la validité interne des résultats. Mais cela peut limiter la validité externe de l'étude (transposabilité), la population étudiée pouvant ne pas être représentative de la population ciblée. En particulier, les patients inclus dans les études contrôlées randomisées ont souvent un profil

de risque plus faible que la population dont ils sont issus [5]. Enfin, l'objectif de l'étude, souvent unique, est défini a priori.

Les études observationnelles ont notamment pour but l'identification de facteurs de risque, d'indicateurs pronostiques, et la détermination de l'étiologie des maladies. Ces données sont majoritairement issues des cohortes et registres. Selon la définition de la direction générale de l'offre de soins (DGOS), un registre, ou cohorte, est un recueil continu et exhaustif de données nominatives intéressant un ou plusieurs événements de santé dans une population géographiquement définie, à des fins de recherche et de santé publique, par une équipe ayant les compétences appropriées. Selon la définition de l'agence nationale pour la recherche sur le syndrome d'immunodéficience acquise et les hépatites virales (ANRS), une cohorte doit être conçue pour répondre à plusieurs questions de recherche épidémiologique, clinique, biologique ou de santé publique même si certaines ne sont pas encore formulées de manière précise au démarrage de la cohorte. Les études observationnelles apportent donc d'autres types de données que les essais cliniques [6].

Les études observationnelles sont cependant soumises à un risque élevé de biais, affectant la validité interne et limitant leur niveau de preuve [7]. Les biais de sélection discriminent la population ; l'échantillon n'est pas représentatif de la population ciblée. Les biais d'information résultent d'une mesure incorrecte de l'exposition à un facteur de risque et/ou de l'évènement étudié, surtout si ces erreurs se répartissent inégalement entre les groupes étudiés. Un facteur de confusion est associé avec l'exposition et affecte l'évènement étudié, sans être un facteur de médiation entre l'exposition et l'évènement sur la chaîne de causalité. Le choix du type de modèle, des variables à inclure et la prise en compte des données manquantes devront corriger en partie ces biais.

Les modèles descriptifs et prédictifs sont généralement dérivés à partir d'études observationnelles. Pour ce travail, les cohortes COPART et RIETE ont été utilisées, et sont présentées dans la partie 3.

A noter qu'une méthode permettant de calculer la taille de la base nécessaire au développement d'un modèle prédictif a été proposée [8].

2.3. Gestion des données manquantes

Les données manquantes sont des valeurs jamais enregistrées, ou perdues accidentellement, ou fausses et donc effacées. Elles peuvent concerner la variable dépendante, ou les variables indépendantes.

Les données manquantes sont très fréquentes dans les études observationnelles, mais peuvent aussi être présentes dans les essais cliniques. Elles peuvent conduire à une estimation biaisée des paramètres, tels que les moyennes, les proportions, ou les coefficients de régression d'un modèle, ainsi que de leurs erreurs-type, donc à des intervalles de confiance et des tests de significativité incorrects [9]. Cela menace les validités interne et externe de l'étude. La perte d'information introduite par les données manquantes conduit également à une perte de précision dans les estimations se traduisant par une majoration de la variance, et à une perte de puissance statistique. Enfin, cela introduit un déséquilibre dans les données, avec des nombres de sujets différant selon les variables considérées : lors d'une analyse multivariée, cela conduira à la suppression de nombreuses observations en cas d'analyse sur les seuls cas complets.

On distingue trois mécanismes, ou lois de probabilité, à l'origine des données manquantes [9] :

- Manquant complètement au hasard (MCAR : *missing completely at random*) : la probabilité qu'une valeur soit manquante est la même pour tous les individus, ne dépend pas des valeurs observées des variables, et ne dépend pas de sa valeur non observée. Cela introduit une perte de précision et de puissance, mais pas de biais.
- Manquant au hasard (MAR : *missing at random*) : la probabilité qu'une donnée soit manquante dépend des valeurs observées des variables, mais pas de sa valeur non observée. Cela introduit une perte de précision et de puissance, ainsi qu'un biais qui pourra être minimisé avec des méthodes appropriées.
- Manquant non au hasard (MNAR : *missing not at random*) : la probabilité qu'une donnée soit manquante dépend de sa valeur non observée. C'est le mécanisme le plus problématique, car la donnée manquante est informative. Outre la perte de précision et de puissance, cela introduit un biais difficile à corriger.

Prenons le cas de la cohorte COPART. Des mesures de pression artérielle systolique de cheville sont réalisées à l'entrée du patient dans le service. Imaginons que l'appareil de mesure soit cassé le jour de l'admission : les pressions de cheville seront manquantes dans le registre, selon un mécanisme MCAR. Toujours dans cet exemple, prenons le cas de la médiocalcose, qui correspond à une calcification des artères de jambe, aboutissant à leur incompressibilité ; dans ce cas-là, les pressions de cheville ne peuvent être relevées, et la donnée sera manquante. On

sait que la médiacalcosse est plus fréquente chez les patients diabétiques. La probabilité que la donnée « pression de cheville » soit manquante dépendra donc d'une autre variable observée, en l'occurrence le diabète, mais pas de la vraie valeur de pression (non observée) : le mécanisme est MAR. La mesure de la pression artérielle systolique à l'orteil est réputée plus fiable que la pression de cheville. Elle peut être réalisée par photopléthysmographie, ou par laser. La photopléthysmographie n'est pas fiable pour les valeurs basses de pression. Imaginons que le centre investigateur ne dispose que d'un appareil de mesure de pression d'orteil par photopléthysmographie. La probabilité que la donnée « pression d'orteil » soit manquante dépendra donc de sa vraie valeur (non observée), donc la donnée n'est pas manquante au hasard : le mécanisme est MNAR.

L'examen des données permet de comprendre le mécanisme en jeu, en recherchant les variables associées au caractère manquant d'une variable donnée par une régression logistique, en étudiant les schémas d'associations entre les variables avec données manquantes, en comparant les proportions de données manquantes entre les différents groupes étudiés.

On distingue enfin deux structures des données manquantes :

- Monotone, lorsque les variables présentant des valeurs manquantes peuvent être ordonnées en fonction de la proportion des valeurs manquantes qu'elles contiennent. On peut dire que les variables Y_1, \dots, Y_k sont ordonnées selon une structure monotone si pour $j = 1, \dots, k-1$, tous les cas contenant des données manquantes pour Y_j présentent également des données manquantes pour $Y_{>j}$.
- Arbitraire, lorsque les données manquantes sont réparties uniformément dans l'ensemble de la base de données.

Plusieurs méthodes d'analyses sont proposées en présence de données manquantes [10] :

- L'analyse sur les seuls cas complets (*case complete analysis*) : seuls les sujets n'ayant aucune valeur manquante pour chacune des variables étudiées sont analysés. En excluant potentiellement un grand nombre de sujets, elle induit une perte importante de précision et de puissance, introduit un biais de sélection si le mécanisme de données manquantes n'est pas MCAR, voire MAR ne dépendant pas de la variable à expliquer, et fausse le processus de sélection des variables lors de l'analyse multivariée. Une analyse sur les cas complets peut cependant se justifier en présence d'une faible proportion d'observations (<5%) contenant des données manquantes [2].

- L'analyse sur les données disponibles (*available case analysis*) : les effectifs seront différents selon la variable étudiée. Cette méthode pose des problèmes similaires à l'analyse sur les cas complets lors d'une analyse multivariée.
- La création d'une catégorie additionnelle pour les variables catégorielles, regroupant les valeurs manquantes de la variable en leur attribuant une valeur fixe, et la création d'un indicateur de donnée manquante pour les variables continues. Cela permet de conserver l'intégralité des sujets pour l'analyse, mais quel que soit le mécanisme de données manquantes (même MCAR), les estimations sont toujours biaisées, et sont difficiles à interpréter car la catégorie additionnelle regroupe en fait des modalités très différentes. Cette méthode n'est pas recommandée.
- L'imputation des données manquantes [11].

Dans l'imputation simple des données manquantes, chaque donnée manquante est remplacée par une seule valeur plausible, permettant de générer une base de données complète. Pour les données MCAR, cette valeur peut être la moyenne des valeurs observées chez les sujets ayant des données complètes ; pour les données MAR, il peut s'agir de la moyenne des valeurs observées sur les sujets ayant les mêmes caractéristiques, de la valeur prédite par un modèle de régression estimé sur l'échantillon complet, d'un tirage au sort avec remise parmi les réponses complètes, ou d'un remplacement par la valeur observée chez le sujet le plus proche défini par un jeu de variables. Ces méthodes peuvent produire des estimateurs non biaisés, mais sous-estiment la variance car l'incertitude liée à la présence de valeurs imputées n'est pas prise en compte, et modifient les associations entre variables. A noter aussi dans les études longitudinales, la méthode *last observation carried forward* (LOCF), qui remplace une valeur manquante par la dernière valeur observée, en faisant l'hypothèse que la valeur de la variable concernée est constante avec le temps.

L'imputation multiple crée plusieurs valeurs possibles d'une valeur manquante. Cela permet de prendre en compte l'incertitude liée à l'imputation, de préserver les aspects importants des distributions et les relations importantes entre les variables, et de minimiser la perte de puissance statistique. Elle est valide pour les données MAR si les hypothèses du modèle d'imputation sont correctes ; l'implémentation est possible avec des données MNAR mais complexe et de validité incertaine. Son concept clé est d'utiliser la distribution des valeurs observées pour estimer un jeu de valeurs plausibles pour chaque valeur manquante. Les étapes de l'imputation multiple sont [9] :

- Remplacer chaque valeur manquante par M valeurs tirées de la distribution possible de la donnée manquante conditionnellement aux données observées (*posterior predictive distribution*), en général prédites par un modèle de régression, afin de créer M bases complètes, où les valeurs imputées sont différentes dans chaque base.
- Analyser séparément chaque base de données complètes, ce qui permet d'obtenir M estimations du paramètre d'intérêt (moyenne, proportion, coefficient de régression...).
- Combiner ces M estimations selon les règles de Rubin pour obtenir un estimateur unique non biaisé ainsi que son erreur-type, qui prennent ainsi en compte l'incertitude liée à la donnée manquante et l'incertitude liée à son imputation.

Plusieurs méthodes d'imputation multiple ont été proposées. L'imputation multiple par équations chaînées (MICE : *multiple imputation by chained equations*) est l'une des plus utilisées [9,12]. Les valeurs manquantes sont d'abord remplacées par un tirage au sort avec remise à partir des valeurs observées, puis la première variable avec données manquantes x_1 est régressée sur toutes les autres variables, uniquement sur la population sans données manquantes pour x_1 . Les valeurs manquantes de x_1 sont ensuite remplacées par tirage au sort dans la distribution possible de la valeur manquante conditionnellement aux données observées. La seconde variable avec données manquantes x_2 est alors régressée sur toutes les autres variables, dans la population sans données manquantes pour x_2 , mais en utilisant les valeurs imputées de x_1 . Les valeurs manquantes de x_2 sont alors tirées au sort dans la distribution possible de la valeur manquante conditionnellement aux données observées. La procédure est répétée pour l'ensemble des variables avec données manquantes, ce qui constitue un cycle. Plusieurs cycles sont répétés pour produire un premier jeu de données imputées, puis l'ensemble est répété M fois pour donner M jeux de données imputées. Ces jeux de données respectent les caractéristiques de la distribution des données observées (variabilité et corrélations entre les variables).

La méthode MICE prend en compte tous les types de variables (quantitatives, binaires, catégorielles ordonnées ou non) car chaque variable est imputée en utilisant son propre modèle d'imputation. Un modèle de régression linéaire est en général utilisé pour les variables quantitatives normalement distribuées, et un appariement d'après la moyenne prédictive (PMM : *predictive mean matching*) pour les variables quantitatives de distribution non normale. Un modèle de régression logistique est utilisé pour les variables qualitatives binaires. Un modèle de régression logistique multinomiale est utilisé pour les variables qualitatives (non

ordonnées ou ordonnées). Pour éviter les biais, le modèle d'imputation doit inclure toutes les variables du modèle final analysé, en particulier l'évènement d'intérêt [9]. Il faut aussi inclure les variables auxiliaires, qui sont les variables corrélées avec les variables contenant des données manquantes ou qu'on pense associées au fait d'être manquant. Les sujets avec des valeurs manquantes dans la variable dépendante sont habituellement exclus, l'imputation de cette variable n'étant pas consensuelle.

Le nombre de bases imputées est sujet à discussion. Certains auteurs suggèrent que ce nombre soit au moins égal au pourcentage de cas incomplets [9].

La qualité du modèle d'imputation doit ensuite être évaluée ; plusieurs méthodes ont été proposées, reposant notamment sur la comparaison entre les valeurs imputées et les valeurs observées [13].

Une imputation des données manquantes selon la méthode MICE a été réalisée dans le cadre de la cohorte RIETE et est présentée dans la partie 3.5.

2.4. Choix du type de modèle

En fonction du type de la variable dépendante, les modèles souvent utilisés en médecine sont : le modèle linéaire, la régression logistique, et le modèle semi-paramétrique de Cox. La régression logistique est généralement utilisée pour les études transversales et pour les études longitudinales avec un suivi similaire entre les patients et sans perdus de vue, alors que le modèle de Cox est utilisé pour les études longitudinales avec un suivi long. Des alternatives à ces modèles de régression existent, tels que les arbres de régression, les réseaux neuronaux et les machines à vecteurs de support, mais nécessitent des bases de données en grande dimension (plus de 100 000 observations) et ne sont donc pas applicables en médecine.

2.4.1. Le modèle de Cox

L'analyse de survie s'applique à un évènement d'intérêt qui survient au cours du temps. Elle est donc particulièrement adaptée aux modèles prédictifs à but pronostique. Elle s'appuie sur la fonction de survie $S(t)$ et la fonction de risque instantanée $h(t)$. La fonction de survie est la probabilité qu'un individu survive jusqu'à un temps t , donc également la probabilité que l'évènement d'intérêt survienne après un temps t . La fonction de risque instantanée décrit le taux de survenue de l'évènement d'intérêt chez les sujets à risque de présenter l'évènement (sujets qui n'ont pas encore présenté l'évènement d'intérêt).

L'inclusion des sujets dans une étude longitudinale de type cohorte ayant lieu tout au long de la période étudiée, pour chaque individu il existe une date d'entrée dans l'étude différente, ou date d'origine, et une date de fin d'observation. Cette date peut correspondre à la date de survenue de l'évènement d'intérêt, à la date de survenue d'un autre évènement empêchant ensuite la survenue de l'évènement d'intérêt (évènement compétitif), à la date de fin de l'étude (date de point), ou à la date de perte de vue du patient.

Ainsi, du fait de cette censure, les données de survie sont généralement incomplètes. La durée de survie est censurée à droite si l'individu n'a pas présenté l'évènement d'intérêt à sa dernière observation : les durées de survie ne sont pas toutes observées, mais on sait qu'elles sont supérieures à une certaine valeur, connue [14]. La censure à gauche correspond au cas où l'individu a déjà subi l'évènement avant la date d'origine : la date de l'évènement est inférieure à une certaine date connue. La censure par intervalle survient si, au lieu d'observer avec certitude la date précise de l'évènement, la seule information disponible est qu'il a eu lieu entre deux dates connues. L'analyse de survie doit pouvoir prendre en compte la censure ; ce n'est pas le cas de la régression logistique et de la régression linéaire. Beaucoup de modèles de survie font l'hypothèse que la censure est indépendante de l'évènement (censure non informative), exposant à un biais d'information si cela n'est pas vérifié. Les données de survie peuvent aussi être soumises à une troncature, signifiant qu'une partie des individus n'est pas observable, et exposant à un biais de sélection.

La méthode de Kaplan-Meier est généralement utilisée pour estimer la fonction de survie en présence de données censurées [14].

Le modèle à risques proportionnels de Cox est largement utilisé pour évaluer l'effet des variables sur la fonction de risque instantané, donc la relation entre les variables indépendantes et la survie [14]. Il s'exprime sous la forme : $h(t) = h_0(t)e^{(\beta_1 X_1 + \dots + \beta_k X_k)}$, avec $h(t)$ la fonction de risque instantané (taux de l'évènement d'intérêt observé entre t et $t+dt$ lorsque dt tend vers 0),

$h_0(t)$ la fonction de risque au temps 0, qui est une constante commune à tous les individus, X_1, \dots, X_k les variables indépendantes, et β_1, \dots, β_k les coefficients représentant l'effet de chaque variable sur la survenue de l'évènement [15]. Ce modèle est semi-paramétrique, car il ne fait aucune hypothèse sur la distribution de la fonction du risque de base $h_0(t)$. Il est proportionnel, car quels que soient 2 individus i et j , le rapport des risques instantanés ne varie pas au cours du temps. L'*hazard ratio* (HR) pour une variable donnée correspond à e^β ; c'est le rapport entre la probabilité instantanée de survenue de l'évènement d'intérêt dans un groupe et la probabilité dans un autre groupe pris comme référence, les groupes étant défini par les classes de la variable. L'HR est indépendant du temps, et se rapproche du risque relatif (RR) lorsque la prévalence de l'évènement est faible. L'effet des variables indépendantes sur la probabilité de survenue de l'évènement d'intérêt et leur effet sur l'incidence de l'évènement d'intérêt sont directement reliés [16].

Des alternatives paramétriques au modèle de Cox existent, qui font une hypothèse sur la distribution des durées de survie [14].

Le principal défaut du modèle de Cox est de faire l'hypothèse que la censure à droite est non informative; les patients censurés ne sont plus considérés comme susceptibles de présenter l'évènement d'intérêt et sortent de l'échantillon des patients à risque.

La survenue d'un risque compétitif, c'est-à-dire un évènement dont la survenue peut précéder celle de l'évènement d'intérêt, et qui empêche alors de développer ce dernier, est un cas également problématique de censure à droite. Dans les applications développées dans la partie 3.3.1, le décès de cause non cardiovasculaire est un évènement compétitif pour la survenue des MACE dans l'analyse de COPART, un patient décédé ne pouvant ensuite présenter l'évènement d'intérêt, et pour la même raison le décès toute cause est un évènement compétitif pour la survenue d'une récurrence d'ETV à l'arrêt du traitement anticoagulant dans l'analyse de RIETE présentée en partie 3.5.

L'estimation de la probabilité de l'évènement d'intérêt par la méthode de Kaplan-Meier est biaisée dans le sens d'une surestimation en cas de risque compétitif [17]. En effet, la somme des estimations de l'incidence de chacun des évènements sera supérieure à l'estimation de l'incidence du critère composite regroupant tous ces évènements. Le modèle de Cox avec estimation de la probabilité de l'évènement d'intérêt par la méthode de Kaplan-Meier aboutira également à un biais dans l'estimation de la magnitude de l'association entre les covariables et l'évènement d'intérêt. Une nouvelle méthode doit donc être utilisée: la fonction d'incidence

cumulée [18]. Elle traduit la probabilité de présenter un évènement avant un temps t et avant la survenue d'un autre type d'évènement. L'estimation des incidences par les fonctions d'incidence cumulée permet que la somme des incidences de chacun des évènements (dont les risques compétitifs) soit identique à l'incidence de l'évènement composite regroupant tous ces évènements. La fonction d'incidence cumulée tend non pas vers 1 comme l'estimateur de Kaplan-Meier, mais vers la proportion globale de l'évènement d'intérêt. L'estimateur d'Aalen-Johansen est un estimateur non paramétrique de la fonction d'incidence cumulée.

Deux types de fonctions de risque instantané sont utilisables en présence de risques compétitifs : la fonction de risque cause-spécifique (*cause-specific hazard function*) et la fonction de risque de la sous-distribution (*subdistribution hazard function*) [18]. La fonction de risque cause-spécifique représente le taux instantané de survenue d'un évènement donné chez les sujets qui n'ont encore présenté aucun des différents types d'évènements ; elle se rapproche de la fonction de risque instantané utilisée dans le modèle de Cox. La fonction de risque de la sous-distribution représente le taux instantané de survenue d'un évènement donné chez les sujets qui n'ont pas encore présenté cet évènement précis (mais qui peuvent déjà avoir présenté d'autres types d'évènements, donc les évènements compétitifs).

En cas de risque compétitif, 2 approches sont donc possibles : modéliser le risque cause-spécifique, ou modéliser le risque de la sous-distribution. Dans la première approche, on estime l'effet des covariables sur la fonction de risque cause-spécifique. La limite principale est l'absence de lien direct entre les HR ainsi calculés et l'incidence de l'évènement d'intérêt [17]. Dans la seconde approche, on estime l'effet des covariables sur la fonction de risque de la sous-distribution. Cette dernière méthode correspond au modèle de Fine et Gray.

2.4.2. Le modèle de Fine et Gray

Le modèle de Fine et Gray permet donc l'identification de facteurs pronostiques en présence de risques compétitifs [17]. Il prédit l'incidence cumulée de l'évènement d'intérêt, qui tend vers sa prévalence. La force de l'association entre un facteur pronostique et l'évènement d'intérêt est exprimée sous la forme d'un *sub-hazard ratio* (sHR), qui correspond au ratio des risques associés à la fonction d'incidence cumulée en présence et en l'absence du facteur

pronostique. Ainsi, si une covariable est associée à une augmentation du sHR, elle sera associée à une augmentation de l'incidence de l'évènement d'intérêt [16]. Ce n'est pas le cas pour les HR estimés par une fonction de risque cause-spécifique.

Un coefficient de régression du modèle de Fine et Gray traduit l'importance de la modification relative de la fonction de risque de la sous-distribution en cas de modification d'une unité de sa covariable associée [16]. Cela traduit la modification relative du taux instantané de survenue de l'évènement d'intérêt, chez les sujets qui n'ont pas encore présenté l'évènement ou qui ont présenté un évènement compétitif. L'importance de l'effet sur la fonction de risque de la sous-distribution est cependant différente de l'importance de l'effet sur le risque de survenue de l'évènement, bien que très proches si la probabilité de survenue de l'évènement d'intérêt est faible ($<0,20$). Les coefficients du modèle de Fine et Gray peuvent alors être interprétés comme des *odds ratios* (OR) pour la fonction d'incidence cumulée : devant un sHR de 2, on pourra dire qu'une augmentation d'une unité de la variable est associée à un doublement de la probabilité de survenue de l'évènement d'intérêt [16].

Enfin, si en présence de risque compétitif le modèle de Fine et Gray paraît plus adapté à un modèle à but pronostique, un modèle de risque cause-spécifique pourra être supérieur pour un modèle à but explicatif ou descriptif [16].

Des modèles de Cox et de Fine et Gray ont été appliquées aux cohortes COPART et RIETE dans la partie 3.

Quel que soit le type de modèle choisi, il faudra ensuite opérer un choix parmi les variables à inclure.

2.5. Sélection des variables à inclure

2.5.1. Raisons de la sélection des variables, et nombre maximal de variables explicatives

Les modèles statistiques sont des règles mathématiques simples dérivées de données empiriques, décrivant l'association entre un variable dépendante et plusieurs variables

explicatives [19]. Ils doivent être valides, c'est-à-dire produire des prédictions avec une précision acceptable, et pratiques, c'est-à-dire permettre des conclusions telles que « quelle est la modification dans la variable dépendante si une des variables indépendantes change d'une unité » [19]. Les études observationnelles, d'où est dérivée la majorité des modèles, comprennent habituellement un grand nombre de variables explicatives potentielles, et un ajustement sur des facteurs de confusion éventuels est également nécessaire. Le nombre de variables à inclure dans le modèle final peut donc être élevé. Cela peut compromettre la stabilité du modèle, mener à un surajustement avec une trop grande dépendance aux données ayant servi à dériver le modèle ce qui surestime sa performance et limite la validité externe, retenir par erreur une variable d'effet faible et non informative, rendre le modèle impossible à estimer du fait d'une multicolinéarité, et dans le cas d'un modèle à but prédictif, le rendre trop complexe et inutilisable en pratique clinique [1,2,20,21].

La nécessité d'inclure de termes d'interaction dans le modèle est débattue ; dans les modèles prédictifs, cela complexifie l'équation et n'améliore habituellement pas la capacité prédictive du modèle [4]. Si des termes d'interaction sont finalement inclus, il faut en limiter le nombre, et les sélectionner sur les connaissances préalables plutôt que de tous les tester et de ne garder que ceux avec l'effet le plus fort, ce qui conduirait à un surajustement [4].

Les variables quantitatives posent un problème spécifique. On préférera les garder inchangées dans le modèle, la transformation d'une variable quantitative en variable catégorielle faisant perdre de l'information. Mais l'hypothèse est alors faite que la relation entre la variable quantitative et l'évènement d'intérêt est linéaire. Souvent, cela n'est pas vérifié. Une solution pour identifier la non linéarité de la relation et transformer la variable pour la garder en tant que variable quantitative dans le modèle, est d'utiliser les polynômes fractionnaires [22].

Au final, un nombre de variables candidates de plus de 10 à 30 est considéré comme trop élevé, et une sélection des variables à garder dans le modèle final doit souvent être opérée [23]. De plus, les modèles multivariés sont très instables si le rapport entre le nombre d'occurrences de l'évènement d'intérêt et le nombre de covariables incluses est faible. Ce nombre d'évènements par variable (EPV : *events per variable*) quantifie le rapport entre la quantité d'information fournie par les données et le nombre de paramètres inconnus à estimer. L'EPV recommandé est de 5 à 15 pour un modèle non issu d'une sélection préalable, mais jusqu'à 50 pour le modèle initial sur lequel sera appliqué une sélection descendante par exemple, la sélection par les algorithmes ajoutant une source supplémentaire d'incertitude dans l'estimation. Au-delà de l'EPV, il faut aussi tenir compte de la force des effets et des

corrélations entre les variables [1]. En effet, une sélection sur des variables dont certaines sont corrélées changera l'interprétation des effets [1]. Pourtant, il ne faudra pas conserver dans le modèle final deux variables fortement corrélées : un choix devra être fait sur la variable à exclure [2].

Le choix des variables est donc lourd de conséquences. L'inclusion par erreur d'une variable dans le modèle augmente la variance de son coefficient β et de ceux des variables corrélées, modifiant la variance des prédictions et leurs erreurs quadratiques moyennes [1]. L'exclusion par erreur d'une variable augmente ou réduit la variance de son coefficient β et de ceux des variables corrélées, modifiant la variance des prédictions et leurs erreurs quadratiques moyennes, mais aussi biaise les coefficients β vers 0, ce qui introduit un biais dans les prédictions ainsi que leurs erreurs quadratiques moyennes.

Enfin, dans un modèle prédictif, les candidats prédicteurs doivent être mesurés avant ou à la date d'origine, et connus au moment où le modèle est censé être utilisé [4].

2.5.2. Sélection des variables par un algorithme à partir des données

Classiquement, la sélection des variables est réalisée à partir de la base de données, par un algorithme, en particulier de type sélection descendante. Dans ce dernier, l'analyse débute à partir d'un modèle plein, incluant toutes les variables candidates, puis celles-ci sont retirées une à une sur la base d'un critère prédéfini.

La sélection descendante a tendance à exclure les facteurs de confusion importants, car elle ignore les variables fortement associées à l'exposition si elles sont faiblement associées à l'évènement d'intérêt conditionnellement à l'exposition [19]. Elle conduit aussi à des modèles instables, de petites modifications dans les données pouvant conduire à des changements importants dans les variables conservées pour le modèle final, ainsi que dans la magnitude de leurs coefficients de régression [21,23]. Enfin, un algorithme de sélection des variables nécessite une base de données plus importante qu'une sélection a priori basée sur l'expertise.

Un test d'hypothèse est généralement utilisé comme critère de sélection des variables dans la sélection descendante. Des tests sont réalisés successivement sur les coefficients des

variables entre différents modèles, pour évaluer si leur association à l'évènement d'intérêt est significative. Les variables non significatives sont exclues une à une, jusqu'à ne garder que les variables significatives. Cela pose en théorie 2 problèmes : la multiplicité des tests conduit à une sous-estimation des *p-values*, et les *p-values* associées aux coefficients d'un modèle ne traduisent pas la significativité propre des variables, mais plutôt si les variables sont associées de manière significative à l'évènement d'intérêt dans le cadre d'un modèle spécifique [1]. La sélection des variables par un test de significativité produit de plus un biais de sélection : les variables dont l'effet est important (peut-être par hasard) auront plus de chance d'avoir un test significatif que celles dont l'effet est faible, et leur coefficient de régression correspondant sera surestimé [2]. Un autre risque est le surajustement, en particulier si la base de données est petite ou si les variables indépendantes ont des effets faibles, ce qui se traduira par une mauvaise calibration dans une autre base de données [2]. Enfin, le choix du seuil pour le test de significativité impacte également le nombre de variables conservées pour le modèle final, ce nombre augmentant avec l'augmentation de la *p-value* sélectionnée [2].

Le critère d'information d'Akaike (AIC : *Akaike information criterion*) est une alternative au test d'hypothèse. Il sélectionne un modèle parmi un ensemble de modèles plausibles, issus des diverses combinaisons possibles des variables. Il pénalise également le modèle en fonction de sa complexité. Il existe un lien entre l'AIC et les *p-values* des tests d'hypothèse sur les coefficients : le modèle final sélectionné sur l'AIC comprend en général des coefficients dont les *p-values* sont $<0,157$ [1]. En pratique, pour un modèle à but prédictif, une sélection descendante avec comme critère de sélection une *p-value* $<0,157$ et sans sélection préalable sur les résultats des analyses bivariées est un choix admissible [23].

Un critère de modification de l'estimation (*change-in-estimate*) a également été proposé. Il consiste à calculer le changement dans l'estimation des coefficients de régression des variables évalués dans deux modèles ; éliminer une variable importante conduira à une modification des coefficients des autres variables, ce qui ne sera pas le cas pour une variable d'importance négligeable. Ce critère peut être utilisé au cours d'une sélection descendante (sélection descendante augmentée) [19].

Outre la sélection descendante, de nombreux algorithmes ont été proposés pour la sélection automatique des variables à partir des données : sélection ascendante, sélection bivariée des variables (variables conservées dans le modèle selon leurs *p-values* dans l'analyse bivariée), régression LASSO (*least absolute shrinkage and selection operator*) [1]...

Des analyses de sensibilité et une étude de la stabilité du modèle sélectionné par l'algorithme sont généralement conseillées [1,21]. Les mesures de la stabilité sont :

- la fréquence d'inclusion de la variable, qui est la probabilité qu'une variable soit sélectionnée dans le modèle final, et qui est évaluée par sous-échantillonnage,
- la fréquence de sélection du modèle, qui est la probabilité qu'une combinaison spécifique de variables constitue le modèle final, et qui est évaluée par sous-échantillonnage,
- la fréquence à laquelle des paires de variables corrélées sont sélectionnées, et qui est évaluée par *bootstrap*,
- le biais conditionnel relatif, qui est le biais d'un coefficient de régression sélectionné relatif à sa vraie valeur, et qui est évalué par *bootstrap*,
- le ratio de l'erreur quadratique moyenne, qui est l'inflation de l'erreur-type d'un coefficient de régression induit par la procédure de sélection de variables, et qui est évalué par *bootstrap*.

A noter que le *bootstrap* peut aussi être utilisé lors de la sélection de variables, en ne gardant dans le modèle final que des variables dont la fréquence de sélection dans les échantillons *bootstrap* dépasse un seuil fixé.

2.5.3. Sélection des variables sur une connaissance préalable, formalisée par un graphe acyclique dirigé

A l'opposé des algorithmes de sélection des variables, une sélection sur les connaissances préalables issues notamment de la littérature scientifique et d'avis d'experts est préférée par de nombreux auteurs [23]. Un ensemble de variables comprenant les prédicteurs potentiels et les facteurs de confusion est ainsi sélectionné a priori, indépendamment des résultats des analyses bivariées dans la base de données, mais en tenant éventuellement compte de l'EPV. Ce mode de sélection des variables candidates est largement recommandé plutôt qu'une sélection sur la base des analyses bivariées. Une deuxième sélection est ensuite opérée sur ces variables, en fonction de la chronologie de la mesure de la variable, du coût, de la qualité

de la mesure, de sa disponibilité pour le clinicien qui utilisera le modèle. Un premier modèle multivarié est ensuite estimé (le modèle global) [1].

La sélection se poursuit ensuite à partir du modèle global. Une sélection automatique des variables par un algorithme peut alors être utilisée, aboutissant à une méthode mixte entre une sélection sur les connaissances préalables et une sélection automatique. Certaines variables peuvent aussi être forcées dans le modèle final indépendamment des résultats de la sélection automatique [24].

Mais cette nouvelle phase de sélection à partir du modèle global peut à nouveau être opérée à l'aide des connaissances préalables. Dans ce cadre, un graphe acyclique dirigé (DAG : *directed acyclic graph*) peut être utile pour formaliser les liens entre les variables. Un DAG est une représentation schématique non paramétrique du processus supposé de génération de données pour un ensemble de variables, dans un contexte spécifié ; il permet de représenter les liens de causalité supposés entre variables [25–27]. Il comprend un ensemble de nœuds, chaque nœud correspondant à une variable, et des flèches unidirectionnelles qui connectent ces nœuds et représentent les liens de causalité. Le graphe est dit acyclique car un nœud ne peut pas être causé par lui-même. Un DAG correspond en fait à un système d'équations, qui décrit la structure des variables, c'est-à-dire les liens de causalité potentiels.

Les données de la littérature et les avis d'experts sont fondamentaux dans le choix des variables du DAG, qui comprennent notamment celles participant à l'effet causal de l'exposition étudiée sur l'évènement d'intérêt, et les facteurs de confusion potentiels. Toute variable en influençant au moins 2 autres doit être représentée [26]. Les variables représentées peuvent ne pas être mesurées.

La relation entre 2 nœuds, représentée par une flèche, peut être positive ou négative, de forte ou de faible magnitude, linéaire ou non. Si on considère deux variables reliées par une flèche, la variable dont est issue la flèche est un ancêtre de la variable sur laquelle pointe la flèche, et cette dernière est un descendant de la variable précédente. Une absence de flèche entre 2 variables signifie l'absence de lien de causalité ; cette assertion constitue une hypothèse plus forte que l'existence d'un lien [26]. Une méthode est donc de partir d'un graphe saturé, toutes les variables étant reliées, puis de supprimer uniquement les connexions impossibles. Un chemin est l'ensemble des flèches qui connecte deux nœuds, quel que soit le sens de la flèche. Dans un chemin causal, toutes les flèches du chemin pointent dans le même sens. Un chemin

comprenant des flèches ne pointant pas dans le même sens est un chemin dérobé (*backdoor path*). Un chemin pourra être bloqué en ajustant sur une des variables le constituant.

Une collision est une variable sur laquelle pointent deux flèches ; un chemin comprenant une collision est naturellement bloqué. Il n'est donc pas nécessaire d'ajuster sur une collision si on veut étudier le lien entre les deux variables reliées par le chemin en question. Au contraire, ajuster sur une collision crée une association non causale (*spurious correlation*) entre les variables d'où sont issues les flèches pointant sur la collision, ce qui rouvre le chemin [27]. Il ne faut pas non plus ajuster sur un descendant d'une collision [25].

Un facteur de médiation se situe sur le chemin causal entre l'exposition et l'évènement d'intérêt : il est causé par l'exposition, et cause à son tour l'évènement d'intérêt. L'effet causal total d'une exposition sur l'évènement d'intérêt est l'effet total transmis au travers de tous les chemins causaux connectant l'exposition à l'évènement, et peut donc impliquer des médiateurs. A l'inverse, l'effet causal direct est un effet qui n'agit pas par un ou plusieurs médiateurs. L'ajustement sur un médiateur diminuera l'effet causal total de l'exposition sur l'évènement d'intérêt, et n'est pas conseillé ; il s'agit d'une situation de surajustement [25]. A noter que les différentes situations conduisant à un surajustement peuvent être identifiées à l'aide des DAG [20].

Un facteur de confusion cause à la fois l'exposition et l'évènement d'intérêt. Il est donc situé sur un chemin dérobé entre l'exposition et l'évènement d'intérêt, et introduira un biais de confusion dans l'estimation de l'effet de l'exposition sur l'évènement d'intérêt. Ce chemin devra être fermé, en ajustant sur ce facteur de confusion : ce dernier devra donc être conservé comme covariable dans le modèle multivarié final [25].

Un biais de sélection peut aussi être identifié par un DAG. Dans le cas d'un biais de sélection, l'association entre l'exposition et l'évènement d'intérêt diffère entre les sujets de la population d'origine et les sujets finalement inclus dans l'analyse finale. Le biais d'admission, ou biais de Berkson, est présent dans les études menées à partir d'une population hospitalière : si l'exposition et l'évènement d'intérêt causent tous deux une hospitalisation, l'hospitalisation apparaîtra comme une collision sur le chemin dérobé entre l'exposition et l'évènement d'intérêt, et ne prendre dans l'étude que des patients hospitalisés aura le même effet qu'ajuster sur l'hospitalisation, donc créera une association non causale entre l'exposition et l'évènement d'intérêt [25]. Les biais par sélection intentionnelle (différence du taux de perdu de vue, biais du survivant, biais d'autosélection) peuvent aussi être représentés dans un DAG.

Un DAG peut aussi identifier un biais d'information. Ce dernier s'interprète comme une plus grande probabilité, pour un groupe de patients donné, de se voir diagnostiquer un état précis par comparaison à un autre groupe, du fait d'une autre variable [25].

Un ensemble d'ajustements suffisant pour une estimation donnée, est l'ensemble des variables qui, une fois inclus dans un modèle, fournira une estimation non biaisée de l'effet de l'exposition sur l'évènement d'intérêt, en fermant tous les chemins non causaux et en ne laissant ouverts que les chemins causaux.

Des recommandations ont été édictées concernant la création d'un DAG ; malgré cela, le DAG reposant sur les connaissances disponibles, les structures causales réelles peuvent ne pas être représentées si une variable clé ou une association n'ont pas été identifiées [26,28]. Les autres limites sont l'absence de représentation de la magnitude de l'effet, notamment des biais, et l'absence de prise en compte de certains biais notamment liés à un échantillon de petite taille (biais lié à des données manquantes, biais lié à un facteur de confusion aléatoire qui ne peut pas être représenté dans le DAG, variable modifiant non pas l'estimation d'un paramètre mais sa variance) [29]. Enfin, l'environnement DAGitty a été développé spécifiquement pour faciliter la construction des DAG et l'identification de l'ensemble d'ajustements suffisant [30].

Dans la partie 3.3.1, la sélection des variables pour le modèle multivarié à but explicatif évaluant l'effet du diabète sur les évènements cardiovasculaires dans la cohorte COPART a été réalisée sur la base d'un DAG. Les covariables sélectionnées ont été l'âge, l'indice de masse corporelle, le sexe masculin et l'activité physique régulière, chacune de ces variables pouvant être un facteur de confusion biaisant l'évaluation du lien entre l'exposition (diabète) et l'évènement d'intérêt (évènements cardiovasculaires) : en effet, chacune de ces variables a un lien causal avec l'exposition, et avec l'évènement. A l'inverse, les facteurs de médiation tels qu'un antécédent de pathologie coronarienne ou vasculaire cérébrale, n'ont pas été inclus dans le modèle, car cela aurait induit une réduction de l'effet causal total du diabète sur les évènements cardiovasculaires. Dans le même travail, nous avons utilisé un modèle descriptif pour évaluer le lien entre plusieurs variables, dont le diabète, et le risque d'amputation majeure. Toutes les variables associées au risque d'amputation majeure avec une p-value <0,20 dans les analyses bivariées ont été incluses dans le modèle, puis une procédure de sélection descendante a été conduite, pour ne garder que les variables associées à l'évènement avec une p-value <0,05. La variable diabète a été forcée dans le modèle final, étant donné qu'il s'agissait de la variable d'intérêt.

Un autre modèle explicatif est illustré dans un second travail à partir de la cohorte COPART, en partie 3.3.2. Le but était d'étudier le lien entre la prescription d'un ARA2 et le risque de décès toute cause. A nouveau, un DAG a été utilisé pour identifier les covariables à inclure dans un modèle multivarié : centre, année de l'hospitalisation, âge, hypertension artérielle, diabète, pathologie coronarienne, insuffisance cardiaque, débit de filtration glomérulaire (DFG), aspirine, clopidogrel, statine, β -bloquant. La variable décrivant la présence d'un inhibiteur du système rénine angiotensine aldostérone (SRAA) était forcée dans le modèle. Il s'agissait d'une variable à 3 classes : ARA2, inhibiteur de l'enzyme de conversion de l'angiotensine (IEC), ou aucun de ces deux traitements.

2.5.4. Cas particulier de la sélection sur des données issues d'une imputation multiple

La sélection des variables par un algorithme doit être adaptée dans le cas de données manquantes corrigées par imputation multiple ; la méthode n'est cependant pas consensuelle.

Souvent, la sélection des variables se fait sur les seuls cas complets, ce qui revient à ne pas utiliser les jeux de données imputées ; la sélection est alors biaisée à moins que les données manquantes soient MCAR [31].

Parfois, la sélection est réalisée sur un seul des jeux de données imputées ; ce jeu de données est donc complet, mais cela revient à utiliser une imputation simple et non pas une imputation multiple. L'estimation des paramètres ne sera pas biaisée, mais les erreurs-type seront trop réduites, avec un risque d'inclure par erreur certaines variables dans le modèle final [31]. De plus, les résultats pourraient différer selon la base complète choisie pour la sélection parmi l'ensemble des bases imputées.

Une autre méthode de sélection est basée sur l'estimation des coefficients β de régression des variables et de leurs erreurs-type dans chacune des bases de données complètes (contenant les données imputées), combinées ensuite selon les règles de Rubin. Ces paramètres sont alors utilisés pour la sélection des variables, par exemple dans le cadre d'une sélection descendante. Même si elle est considérée comme la référence, cette approche selon les règles de Rubin peut être limitée par le temps de calcul nécessaire [31].

La sélection peut aussi être conduite sur chacun des jeux de données imputées, en conservant ensuite les variables pour le modèle final selon un des critères suivants :

- prendre les variables qui sont sélectionnées dans au moins une des bases de données imputées,
- prendre les variables qui sont sélectionnées dans au moins la moitié des bases de données imputées,
- prendre les variables qui sont sélectionnées dans toutes les bases de données imputées [31].

Contrairement à l'approche selon les règles de Rubin, la combinaison des modèles est réalisée seulement à la fin de la sélection, et non pas à chacune de ses étapes. Le risque du 1^{er} critère est de sélectionner trop de variables, et celui du 3^{ème} critère de ne pas en sélectionner assez. La méthode utilisant le 2nd critère peut être utile, mais ne prend pas en compte l'incertitude liée aux données manquantes.

Enfin, les bases de données complètes (imputées) peuvent être combinées en une seule base ; les paramètres estimés dans cette base ne seront pas biaisés, mais les erreurs-type seront trop réduites. Une pondération, calculée selon la proportion de valeurs manquantes, peut cependant être appliquée pour corriger les erreurs-type [31]. Même si l'approche de Rubin reste la meilleure, cette méthode peut donner des résultats acceptables.

Le problème de la sélection des variables sur données imputées s'est posé avec le travail réalisé sur la cohorte RIETE, dans les parties 3.5 et 3.6. Pour la partie 3.5, la sélection a été réalisée en deux temps, d'abord sur la base des connaissances préalables puis à l'aide d'une sélection descendante utilisant les coefficients β estimés selon les règles de Rubin et un seuil de *p-value* $< 0,157$. Pour la partie 3.6, la sélection a été réalisée sur les données de la littérature et les résultats de l'analyse de la partie 3.5.

2.6. Evaluation de la performance du modèle

2.6.1. Mesures de la performance du modèle

La performance d'un modèle prédictif doit toujours être vérifiée. Lors de la dérivation du modèle, on s'intéresse en particulier à la discrimination. Mais lors de sa validation, la calibration sera également fondamentale.

Dans un modèle de Cox, les prédictions sont obtenues à l'aide de l'indice pronostique, qui correspond à la somme des variables pondérées du modèle, la pondération de chaque variable étant réalisée à l'aide du coefficient β de régression [32]. Pour obtenir la probabilité prédite de l'évènement, outre l'indice pronostique, la fonction de survie de base est nécessaire (survie hypothétique d'un individu dont l'indice pronostique serait de 0) [33]. L'indice pronostique sera utile pour la calibration et la discrimination.

Le modèle de prédiction final devra, lui, être présenté sous la forme d'une équation comportant les coefficients β de régression [34]. En pratique, une version simplifiée est souvent présentée, où les coefficients β ont été convertis et arrondis à des nombres faciles à additionner, et qui sont liés au risque absolu de présenter l'évènement d'intérêt. Cette simplification conduisant à une perte d'information et donc à une diminution de la performance prédictive, il faudra aussi rapporter les estimateurs de la performance de ce modèle simplifié, pour permettre une comparaison avec le modèle initial.

2.6.1.1. Calibration

La calibration correspond à la concordance entre la probabilité, prédite par le modèle, de survenue de l'évènement d'intérêt au cours d'une période donnée, et sa fréquence effectivement observée. C'est la capacité du modèle à correctement prédire le risque absolu de survenue de l'évènement. Un modèle est bien calibré si la probabilité de survenue de l'évènement prédite pour chaque groupe de patients est proche de la proportion des patients du groupe ayant effectivement développé l'évènement. La calibration peut être reportée graphiquement, avec la probabilité prédite de la survenue de l'évènement sur l'axe des X, et la fréquence de l'évènement sur l'axe des Y [35,36]. L'intercept et la pente de cette droite peuvent être calculés ; pour un modèle parfaitement calibré, la pente est égale à 1 et l'intercept à 0. Un test d'hypothèse pour ces valeurs peut être réalisé. La calibration peut aussi être évaluée par un test statistique, tel le test de Hosmer et Lemeshow pour le modèle logistique, et les tests de

Nam-D'Agostino et de Gronnesby-Borgan pour les modèles de survie ; ces tests ne sont pas recommandés, car ils sont sensibles à la taille des échantillons, ont tendance à devenir significatifs avec l'augmentation du nombre de sujets, et ne donnent pas d'information en cas de mauvaise calibration [37]. Pour les analyses de survie, une méthode basée sur le modèle de Poisson a aussi été proposée [35].

Pour le travail réalisé sur la cohorte RIETE dans la partie 3.6, la méthode graphique avec calcul de la pente de la droite a été préférée, étant donné qu'elle a aussi été décrite dans le contexte des données imputées [38]. La pente de la droite a été calculée dans chacun des jeux de données imputées, puis combinée en un seul estimateur selon les règles de Rubin.

2.6.1.2.Discrimination

La discrimination correspond à la capacité du modèle à différencier les sujets qui développeront ou non l'évènement d'intérêt, donc à correctement classer les individus selon le niveau de risque, faible à élevé. Dans le contexte des analyses de survie, elle peut aussi être définie comme la capacité du modèle à séparer les individus avec une survie sans évènement élevée de ceux avec une survie sans évènement courte [39]. La discrimination est parfaite si les risques prédits pour tous les individus qui ont présenté l'évènement d'intérêt sont plus élevés que ceux des individus n'ayant pas présenté l'évènement. L'index de concordance (c-index) est classiquement utilisé dans ce contexte, traduisant la probabilité que, pour toute paire d'individus tirée au sort, un avec l'évènement d'intérêt et un sans, le modèle prédise un risque plus grand à l'individu qui a effectivement développé l'évènement d'intérêt. Le c-index correspond à l'aire sous la courbe (AUC : *area under the receiver-operating characteristic curve*) pour les modèles avec un évènement d'intérêt binaire. Le c-index est bien adapté aux modèles logistiques, mais pas aux modèles de survie, car il ne prend pas en compte l'information apportée par la durée de survie (en plus de l'information sur l'apparition ou non de l'évènement d'intérêt), et ne prend pas non plus en compte les sujets censurés avant la date de point [39]. Pour les modèles de survie, différents index de concordance ont été proposés : c-index de Harrell, Pencina et D'Agostino, dans sa version originale ou avec une estimation de l'AUC de Harrell selon la méthode de Uno, ces deux c-index étant les plus fiables, c-index de Chambless et Diao, indice

k de Gönen et Heller [36,39]. La statistique D de Royston et Sauerbrei (*prognostic separation D-statistic*) est également pertinente [40,41]. Son avantage est qu'elle peut être calculée dans des jeux de données imputées puis combinée en un seul estimateur selon les règles de Rubin [42]. Elle peut s'interpréter comme la séparation (log du HR) entre les distributions de survie de deux groupes pronostiques indépendants [43]. Plus la valeur de D est élevée, plus la séparation est importante et moins les distributions de survie se chevauchent. Et e^D peut s'interpréter comme le HR entre deux groupes à risque de mêmes effectifs, définis par une dichotomisation, selon la valeur médiane, de la distribution des indices pronostiques des patients.

Pour le travail réalisé sur la cohorte RIETE dans la partie 3.6, la statistique D de Royston et Sauerbrei a été préférée.

2.6.1.3. Autres paramètres estimant la performance du modèle

Les coefficients de détermination de type R^2 expriment la part de variance de la variable dépendante (événement d'intérêt) expliquée par le modèle de régression. Plusieurs R^2 existent [36,41]. Dans le cadre des modèles prédictifs, le R_{PM}^2 de Kent et O'Quigley et le R_D^2 de Royston et Sauerbrei, basé sur leur statistique D, semblent être les meilleurs dans les études de simulation pour quantifier la capacité à prédire [41]. Le R_{PM}^2 n'est pas affecté par la censure, de même que le R_D^2 si la distribution de l'indice pronostique du modèle est normale. De plus, ils sont facilement compris : par exemple, un R_{PM}^2 de 0,20 signifie que 20% de la variation dans l'évènement d'intérêt est expliqué par les variables du modèle. D'autres classes de mesures de type R^2 existent, pouvant exprimer, non pas la part de variance, mais la part de hasard expliquée par le modèle (ρ_{XO}^2, ρ_k^2) ou la précision de la prédiction (R_G^2, R_{SK}^2) [41,44].

Le score de Brier est un critère composite qui apporte des informations sur la calibration et la finesse des prédictions [45]. Il va de 0 à 1, avec 0 pour une prévision parfaite et 1 pour un mauvais modèle. Il permet aussi de comparer deux modèles.

L'erreur quadratique moyenne de la prédiction (MSPE : *mean squared prediction error*) évalue la dispersion, ou variabilité, des prédictions du modèle. Elle peut aussi être exprimée comme un pourcentage de la valeur moyenne des observations.

Tous ces indicateurs, ainsi que ceux évaluant la calibration et la discrimination, évaluent mal les conséquences cliniques de l'utilisation du modèle multivarié. De nouveaux indicateurs ont été développés pour cela : l'analyse de la courbe de décision, et la courbe d'utilité relative [46,47].

2.6.2. Validation interne

2.6.2.1.Méthodes habituelles

Le but d'une étude pronostique est de développer un modèle de classification à partir d'un jeu de données disponibles, et d'estimer la performance qu'il aurait sur de futures données indépendantes, c'est-à-dire sa performance prédictive. La performance d'un modèle multivarié évaluée sur les données ayant servies à le dériver, appelée performance apparente, est habituellement surestimée. En effet, le modèle a été développé pour s'adapter au mieux aux données de la cohorte de dérivation ; il y a donc un risque de surajustement, et d'optimisme dans l'évaluation de la performance apparente [34]. L'optimisme augmente lorsque le nombre d'évènements d'intérêt diminue et celui des prédicteurs augmente, donc lorsque l'EPV diminue [34]. L'optimisme augmente également lorsque aucune procédure de sélection des variables n'a été appliquée [34]. La performance devrait donc être estimée sur des données n'ayant pas servies à la dérivation du modèle [48]. Plusieurs approches sont proposées pour corriger le surajustement et l'optimisme, et estimer la performance d'un modèle sur des sujets indépendants [49].

Pour un modèle prédictif, la validation interne consiste à évaluer ses capacités prédictives sur la base de données initiales. Plusieurs méthodes existent : séparation de la base en une cohorte de dérivation et une cohorte de validation (*split-sample*), validation croisée (*cross-validation*), rééchantillonnage par *bootstrap*, sous-échantillonnage. Aucune de ces méthode n'est fiable sur de très petits échantillons ; en pratique, un échantillon avec au moins

100 évènements d'intérêt et 100 non-évènements est nécessaire pour une évaluation fiable de la performance d'un modèle prédictif [50].

La séparation de la base de données (*split-sample*) en une cohorte de dérivation, où sera développée le modèle multivarié, et une cohorte de validation, où sa performance sera estimée, est souvent utilisée. La performance est donc déterminée sur des données similaires à celles ayant permis la dérivation du modèle, mais indépendantes. La séparation des observations en deux groupes peut se faire de manière aléatoire ou non. Cette méthode est très critiquée, en particulier en cas de faible effectif, car elle aboutit à une perte de puissance pour le développement du score dans la cohorte de dérivation, et ne permet pas une évaluation correcte de sa performance dans la cohorte de validation avec souvent une estimation pessimiste et une variabilité importante [49,50].

La validation croisée, le rééchantillonnage par *bootstrap*, voire le sous-échantillonnage, apparaissent plus fiables pour la validation interne d'un score dérivé sur une base de données comprenant des effectifs faibles à modérés, ce qui est la situation la plus fréquente en médecine. En effet, dans les bases comportant plus de 100 000 évènements et moins de 100 prédicteurs potentiels, l'optimisme est négligeable, la validation interne consistant à estimer une performance moyenne n'a donc que peu d'intérêt : la quantification de l'hétérogénéité de la performance est préférable [50].

Dans la procédure de validation croisée (*cross-validation*), la population initiale est divisée en k échantillons, puis le modèle est dérivé sur l'effectif de $k-1$ échantillons et validé sur l'échantillon restant, puis la procédure est répétée k fois, afin que chaque sujet ait servi à tester le modèle. Les k estimations de la performance sont alors moyennées en un estimateur unique [49]. Avec 10 échantillons (donc une performance estimée 10 fois, à chaque fois sur 10% des données initiales), cette méthode a un biais et une variabilité faibles, mais n'est pas adaptée à toutes les mesures de la performance. Dans sa forme extrême, pour une base initiale contenant n observations, $k = n$ donc la validation est réalisée sur un seul sujet à la fois, ce qui est équivalent à une méthode *jackknife* [49]. La validation croisée produit cependant un estimateur biaisé ; ce biais peut être réduit en diminuant la taille des k échantillons, mais ce qui majore la variabilité.

Dans la procédure *bootstrap* classique, B échantillons *bootstrap* sont générés à partir d'un tirage au sort avec remplacement réalisé sur la base de données. La performance du modèle est estimée sur chacun des B échantillons, ainsi que dans la base initiale. La différence entre la

performance sur l'échantillon *bootstrap* et la performance sur la base initiale donne une estimation de l'optimisme. L'optimisme est ainsi évalué B fois (une fois dans chacun des B échantillons), puis ces B estimations sont moyennées pour fournir une estimation unique de l'optimisme. Cet estimateur permet alors de corriger la performance apparente du modèle, calculée sur la base initiale [4,49]. L'optimisme est aussi utilisé pour corriger les IC95% des paramètres de la performance. Deux méthodes de correction de l'IC95% ont été récemment proposées. La méthode « *location-shifted bootstrap confidence interval* » débute par une estimation par *bootstrap* de l'IC95% du paramètre de performance apparente du modèle (pente de la droite de calibration, D de Royston et Sauerbrei, R_D^2 de Royston et Sauerbrei...) sur la base de données initiale, puis on soustrait l'optimisme à la borne basse et à la borne haute de l'IC95%, pour obtenir un IC95% corrigé [51]. Dans la méthode « *two-stage bootstrap confidence interval* », les distributions *bootstrap* des paramètres corrigés par l'optimisme sont directement calculées : un nouveau rééchantillonnage *bootstrap* est réalisé dans chacun des échantillons *bootstrap* initiaux pour obtenir une distribution *bootstrap* des paramètres corrigés par l'optimisme, puis l'IC95% est calculé à partir de cette distribution [51]. Cette 2^{ème} méthode produirait des IC95% plus larges, rendant mieux compte de la variance des paramètres corrigés par l'optimisme, mais le temps de calcul nécessaire est exponentiel, limitant son utilisation. Le nombre B d'échantillons *bootstrap* nécessaire pour estimer correctement les paramètres de la performance est sujet à discussion, certaines méthodes ont été proposées permettant d'évaluer ce nombre [52]. En pratique, un nombre élevé de réplifications produit des estimations plus précises, mais nécessite un temps de calcul élevé, notamment si l'estimateur est complexe. Pour une procédure de *bootstrap* classique avec entre 5 et 20 EPV, 100 réplifications sont suffisantes, un nombre plus élevé n'améliorant que de manière marginale l'estimateur [49]. La taille de chaque échantillon *bootstrap* est aussi discuté, mais pourrait être équivalent à la taille de la base initiale (l'échantillon *bootstrap* ne reproduit cependant pas la base initiale, du fait du remplacement de l'observation tirée au sort) [49]. Enfin, concernant l'estimation des intervalle de confiance à 95% (IC95%) par les méthodes « *location-shifted bootstrap confidence interval* » et « *two-stage bootstrap confidence interval* », au moins 1000 échantillons *bootstrap* sont nécessaires [51]. Des variations de la méthode classique du *bootstrap* existent : les méthodes 0,632 et 0,632+ [49]. Le *bootstrap* semble la technique la plus fiable pour la validation interne. Il est recommandé de reproduire chacune des étapes de création du modèle, y compris l'étape de sélection des variables, dans chacun des B échantillons *bootstrap* ; cela peut en théorie conduire à B modèles multivariés différents, en plus du modèle développé à partir de la base initiale, avec pour chacun des valeurs différentes des estimateurs de la

performance (pente de la droite de calibration, c-index de Harrell...) [4,34]. Pour un estimateur donné, la moyenne des différences correspond à l'optimisme de la performance apparente du modèle développé dans les données initiales. L'avantage du *bootstrap* est de fournir, outre une estimation de l'optimisme qui permettra de corriger les estimateurs de la performance apparente, un facteur de contraction (*shrinkage factor*), qui pourra être utilisé pour tenir compte du surajustement dans l'estimation des coefficients β du modèle final [34].

2.6.2.2. Cas particulier de la validation interne sur des données issues d'une imputation multiple

La validation interne est complexe en cas de données manquantes. Pour rappel, l'imputation multiple des données manquantes est une technique de plus en plus utilisée, qui consiste à imputer de nombreuses fois les données manquantes en créant donc M jeux de données complets, de réaliser l'analyse statistique sur chacune des M bases, puis de combiner les paramètres estimés dans chaque base et leur variance en tenant compte de l'incertitude liée à l'imputation. Lorsque le paramètre d'intérêt est un estimateur de la performance prédictive, la méthode permettant d'associer cette validation interne à l'imputation multiple pour obtenir un estimateur non biaisé de la performance n'est pas claire [48].

L'imputation multiple des données manquantes, pour produire M bases de données complètes, suivie de la validation interne sur ces M bases par *bootstrap*, semble donner des estimations trop optimistes de la performance. La constitution de B échantillons *bootstrap* à partir de la base initiale, qui contiendront donc des données incomplètes, suivie d'une imputation multiple des données manquantes dans chacun des B échantillons *bootstrap* ainsi que dans le groupe des patients de la base initiale non présents dans les B échantillons, puis de la validation interne, est préférée par certains auteurs même si le temps de calcul peut être un obstacle [48]. La validation par *bootstrap* à partir des M bases de données imputées reste cependant valide [53]. Dans ce dernier cas, il serait préférable d'exclure l'évènement d'intérêt de l'équation d'imputation multiple, alors que la qualité de l'imputation est pourtant considérée meilleure en le conservant [38,48].

Les nombres d'échantillons *bootstrap* B et de bases imputées M nécessaires à une estimation correcte sont discutés, mais doivent être les plus grands possibles [48].

Enfin, pour les données imputées, les mesures de la performance du modèle dans chacun des jeux de données imputées peuvent être estimées à partir des prédictions du modèle puis combinées selon les règles de Rubin. Une autre méthode consiste à combiner selon les règles de Rubin les prédictions réalisées dans chacun des jeux de données imputées, puis d'estimer directement la performance du modèle en utilisant ces prédictions combinées ; les conséquences de cette méthode sont mal évaluées, mais la performance du modèle serait surestimée [38].

Pour le travail réalisé sur la cohorte RIETE, présenté dans la partie 3.6, pour des raisons de temps de calcul notamment, le choix a été fait de réaliser d'abord une imputation multiple par équations chaînées des données manquantes, présentée dans la partie 3.5, puis de tirer les échantillons *bootstrap* pour la validation interne. Ce sont les mesures de la performance qui ont été combinées avec les règles de Rubin, et pas les prédictions. L'évènement d'intérêt (récidive thromboembolique veineuse à l'arrêt du traitement anticoagulant) a été conservé dans l'équation d'imputation multiple.

2.6.3. Validation externe

La validité externe, ou généralisation, est la capacité du modèle multivarié à fournir des prédictions justes dans des échantillons de patients autres que celui où il a été dérivé [54]. Ses composantes sont :

- la reproductibilité, qui est la validité du modèle dans un échantillon de patients non inclus dans la cohorte de dérivation mais issu de la même population,
- la transposabilité, qui est la validité du modèle dans des échantillons de patients différents de ceux de la cohorte de dérivation (différence d'époque, de lieu, ou de stade de la maladie), mais reliés à ceux-ci, ou dont les données ont été recueillies par des méthodes différentes ou sur des durées de suivi différentes.

Comme pour la validation interne, la calibration et la discrimination seront les paramètres les plus importants à considérer [36].

La validation interne par *bootstrap* est une approche de la reproductibilité, mais pas de la transposabilité du fait du risque de sous-ajustement, qui correspond au cas où une variable n'a pas été incluse dans le modèle multivarié alors qu'elle apparaît importante dans un autre échantillon de patients [54]. Même si des techniques de séparation non randomisée de la base de dérivation ont été proposées (par exemple selon l'année d'inclusion pour étudier la transposabilité selon l'époque), l'étude de la transposabilité nécessite des données différentes de celles ayant servi à dériver le modèle [49,54].

La taille du nouvel échantillon de patients nécessaire pour tester la validité externe d'un modèle prédictif est discutée. Une règle classique est de sélectionner un échantillon de patients comprenant au moins 100 événements d'intérêt (voire au moins 200) et 100 non-événements [55]. Cette règle est remise en question, et des calculs d'effectifs basés sur une simulation, le R^2 du modèle, la variance de l'évènement d'intérêt et sa proportion dans le nouvel échantillon, l'erreur-type des paramètres étudiés, ou la courbe de calibration sont proposées [56–58].

L'interprétation d'une étude de validation externe peut être complexe, mais une méthode a été développée reposant sur 3 étapes : la comparaison de la cohorte de dérivation et de la cohorte de validation, l'évaluation de la performance du modèle dans la cohorte de dérivation, puis l'interprétation des résultats de l'étude de validation en combinant les résultats des deux étapes précédentes [59].

Un problème fréquent lors de la validation externe, est un taux d'incidence de l'évènement d'intérêt différent entre la cohorte initiale et le nouvel échantillon de patients testé, ce qui altère la calibration [55]. Le modèle pourra cependant être corrigé en ajustant le risque de base ou en modifiant l'intercept [55]. Outre une différence entre les taux d'incidence, une hétérogénéité dans les effets des prédicteurs (variables incluses dans le modèle) et une différence entre les distributions des valeurs d'un prédicteur entre la cohorte de dérivation et la cohorte de validation expliqueront les performances différentes du modèle lors de l'étude de validité externe [55]. Plutôt que de rejeter le modèle, des méthodes existent pour le corriger, reposant sur l'ajustement du risque de base ou des poids des prédicteurs, la réestimation de l'effet des prédicteurs et la modification du jeu des prédicteurs inclus dans le modèle [60].

Pour une maladie donnée, les incidences cumulées des événements d'intérêt peuvent être très différents entre les essais cliniques contrôlés randomisés et les cohortes. Cela est illustré dans la partie 3.4, qui explore la validité externe des résultats des études COMPASS et VOYAGER-PAD.

2.6.4. Evaluation de l'impact de l'implémentation du modèle dans la pratique clinique

L'étude d'impact du modèle consiste à quantifier le bénéfice de son utilisation (sur l'amélioration de l'état de santé, ou sur le rapport coût/efficacité) dans la pratique clinique. Cette étape est rarement réalisée. Elle repose sur la comparaison de deux groupes, idéalement dans le cadre d'une étude randomisée, un groupe intégrant le modèle dans la prise en charge, et l'autre géré sans l'aide du modèle, selon les standards habituels [60].

3. Applications

3.1. Cohortes utilisées

3.1.1. L'artériopathie athéromateuse des membres inférieurs et la cohorte COPART

L'AAMI est définie par une obstruction, provoquée par l'athérome, d'une ou plusieurs artères des membres inférieurs, responsable d'un défaut de perfusion du territoire en aval. Le diagnostic actuel repose sur une diminution de l'indice de pression systolique (IPS) à la cheville ($<0,90$), qui est le rapport entre la pression systolique mesurée sur les artères de la cheville et la pression systolique humérale [61]. L'AAMI peut se présenter sous 3 stades : asymptomatique, ischémie d'effort, ischémie permanente. Au cours de l'ischémie d'effort ou claudication artérielle intermittente, le patient décrit des douleurs du membre lors de la marche, cédant au repos. L'ischémie permanente est caractérisée par la présence de douleurs au repos et/ou de troubles trophiques (ulcère, nécrose, gangrène). Le stade d'ischémie critique correspond à un sous-type de l'ischémie permanente, pour lequel le risque d'amputation de membre est majeur, et est généralement défini par la présence de douleurs de décubitus ou de troubles trophiques depuis au moins 15 jours avec une pression artérielle systolique inférieure à 50 mmHg à la cheville ou inférieure à 30 mmHg à l'orteil, ou une pression d'oxygène mesurée à travers la peau ($TcPO_2$) inférieure à 30 mmHg [62].

Les patients atteints d'AAMI présentent un risque élevé de MACE, qui regroupent les décès de cause cardiovasculaire, les syndromes coronariens aigus (SCA) et les accidents vasculaires cérébraux (AVC) ischémiques. Malgré les traitements médicamenteux, l'incidence cumulée à 1 an de ces événements est au moins de 12%, et jusqu'à 16,8% après revascularisation d'un membre en ischémie critique [63,64]. Le risque de complication pour le membre inférieur (MALE : *major adverse limb events*), comprenant notamment la nécessité d'une revascularisation, l'ischémie aiguë et l'amputation, est important : au cours d'un suivi de 5 ans, 21% des patients au stade d'ischémie d'effort vont évoluer vers une ischémie critique, avec pour 4 à 27% une amputation [65].

Afin de réduire le risque d'événements cardiovasculaires et de complication du membre, les différentes sociétés savantes recommandent d'associer un antiplaquettaire, une statine et un

inhibiteur du SRAA, en particulier un IEC, en plus de la prise en charge habituelle des facteurs de risque cardiovasculaire [66,67].

La revascularisation du membre, si elle est techniquement réalisable, est toujours indiquée en cas d'ischémie permanente ou d'ischémie aiguë, et se discute au stade d'ischémie d'effort avec une réduction notable de la distance de marche sans amélioration après une rééducation vasculaire à la marche. Elle fait appel à différentes techniques : l'angioplastie associée ou non à la pose d'un stent, qui consiste en une dilatation au ballon de la zone sténosée de l'artère par voie endovasculaire, la recanalisation, également par voie endovasculaire, qui permet de contourner une zone occluse en créant une dissection localisée dans la paroi en regard, l'endartériectomie, qui nécessite une ouverture chirurgicale de l'artère pour une exérèse de la plaque d'athérome, et le pontage chirurgical, soit par une veine du patient (saphène) voire une allogreffe, soit par une prothèse, qui permet de dériver une occlusion en général longue. La thrombectomie par sonde de Fogarty, par voie endovasculaire, est la technique de choix en cas d'ischémie aiguë de membre, au cours de laquelle l'artère est brutalement occluse notamment par un embolie à point de départ cardiaque (complication d'une fibrillation atriale).

L'amputation peut être nécessaire, en cas d'ischémie permanente sans possibilité de revascularisation, et notamment en présence d'un trouble trophique, ou bien associée à la revascularisation si l'état local est trop dégradé (nécrose, gangrène). On distingue 4 niveaux classiques d'amputation : orteil, trans-métatarsien (avant-pied), jambe (environ 20 cm sous le genou) et cuisse (environ 20 cm sous la hanche). L'amputation est dite majeure si elle concerne la jambe ou la cuisse, et mineure si elle concerne l'orteil ou l'avant-pied.

COPART est une cohorte hospitalière multicentrique, prospective, incluant de manière consécutive tous les patients hospitalisés dans les services de médecine vasculaire de 4 centres hospitaliers universitaires (CHU) français pour la prise en charge d'une AAMI. L'étude a été initiée en juin 2004 par le service de médecine vasculaire de l'hôpital Rangueil (CHU de Toulouse). La cohorte a été étendue en octobre 2006 au service de chirurgie thoracique et cardiovasculaire – médecine vasculaire de l'hôpital Dupuytren (CHU de Limoges) et au service de médecine interne et vasculaire de l'hôpital Saint-André (CHU de Bordeaux). Enfin, le service de médecine vasculaire de l'hôpital européen Georges Pompidou (assistance publique – hôpitaux de Paris) a été inclus en 2012. Les caractéristiques de la cohorte, en particulier les critères d'inclusion et de non inclusion, ont déjà été publiées [68]. Tous les patients inclus présentent une AAMI symptomatique, confirmée par une imagerie et par les paramètres hémodynamiques. La prise en charge des patients suit les recommandations internationales. Après l'hospitalisation initiale, les patients sont suivis au moins un an. Les données sont

collectées de manière prospective, notamment par des attachés et infirmières de recherche clinique spécifiquement formés. Les patients, leurs familles et/ou leurs médecins sont systématiquement contactés à 1 an de la sortie d'hospitalisation. La cohorte COPART est la seule cohorte française de patients présentant une AAMI. Elle ne s'adresse cependant qu'aux patients hospitalisés, donc potentiellement les plus graves.

Un score pronostique pour le risque de décès toute cause, de SCA et d'AVC à 1 an de la sortie d'hospitalisation a déjà été dérivé à partir de la cohorte COPART [69]. Ce score a bénéficié d'une validation externe sur 129 patients dans une étude prospective, avec des résultats encourageants [70]. Il n'est cependant pas utilisé en pratique. Les données concernant le pronostic du membre inférieur sont plus rares et hétérogènes, en particulier concernant les facteurs pronostiques et l'impact des thérapeutiques non chirurgicales. Le score WIFI (*Wound, Ischemia, and foot Infection*) est le plus utilisé concernant l'évaluation du risque pour le membre inférieur [71]. Il résulte uniquement d'un consensus d'experts, avec des résultats parfois discordants en pratique [72,73].

3.1.2. La maladie thromboembolique veineuse et la cohorte RIETE

La maladie thromboembolique veineuse (MTEV) regroupe l'embolie pulmonaire (EP), la thrombose veineuse profonde (TVP) du membre inférieur, les thromboses de sites inhabituels (TVP du membre supérieur dont la thrombose sur voie veineuse centrale, thromboses splanchniques, thrombophlébite cérébrale, thrombose de veine rétinienne, thrombose de veines gonadiques...) et la thrombose veineuse superficielle (TVS). L'EP expose à un risque de décès par choc cardiogénique. Un modèle pronostique a été proposé pour l'évaluation de la gravité de l'embolie pulmonaire : le score PESI (*pulmonary embolism severity index*) [74]. Une version simplifiée de ce score a également été validée [75]. Pour le membre inférieur, les TVP proximales désignent les thromboses dont la tête du thrombus est située dans la veine poplitée ou plus haut (veines fémorale superficielle et commune, iliaque externe et commune, cave inférieure). Les TVP distales désignent ainsi les thromboses à l'étage infrapoplité (veines fibulaires, tibiales postérieures, tibiales antérieures, gastrocnémiennes, soléaires). Le risque d'EP des TVP distales est réputé plus faible, de même que leur risque de récurrence, leur association aux facteurs de risque biologiques de thrombose et au cancer occulte. Les TVP

distales sont peu ou pas représentées dans les études randomisées contrôlées sur la MTEV. Pour le membre supérieur, le réseau veineux profond est représenté par les veines jugulaire interne, cave supérieure, brachiocéphalique, sous-clavière, axillaire, humérale, radiale et ulnaire. Cependant la majorité des études s'intéresse uniquement aux veines sous-clavière, axillaire et humérale (\pm veine jugulaire interne). La durée minimale du traitement anticoagulant pour une EP ou une TVP proximale est de 3 mois [76,77]. Le risque de récurrence à l'arrêt du traitement anticoagulant dépend avant tout des facteurs favorisant de l'épisode initial. On distingue ainsi des évènements :

- provoqués par un facteur de risque majeur transitoire, dans les 3 mois précédant le diagnostic : chirurgie avec anesthésie générale de plus de 30 minutes, alitement d'au moins 3 jours à l'hôpital pour une pathologie aiguë, accouchement par césarienne,
- provoqués par un facteur de risque mineur transitoire, dans les 2 mois précédant le diagnostic : chirurgie avec anesthésie générale de moins de 30 minutes, alitement de moins de 3 jours à l'hôpital pour une pathologie aiguë, alitement d'au moins 3 jours en dehors de l'hôpital pour une pathologie aiguë, traitement par œstrogènes, grossesse ou accouchement, traumatisme d'un membre inférieur associé à une réduction de la mobilisation depuis au moins 3 jours,
- provoqués par un facteur de risque persistant : cancer actif, maladie inflammatoire chronique intestinale, syndrome des anticorps antiphospholipides,
- non provoqués : ces évènements peuvent être favorisés ou non par un facteur de risque biologique héréditaire de thrombose (déficit en antithrombine, en protéine C ou en protéine S, polymorphisme du facteur V Leiden ou du facteur II G20210A), par le sexe masculin, l'obésité, l'âge, les antécédents familiaux d'ETE [78].

Le taux d'incidence de la récurrence au cours de la 1^{ère} année suivant l'arrêt d'une anticoagulation d'au moins 3 mois pour un premier ETEV (EP ou TVP proximale) varie de 3,1 /100 patients-année (IC95%, 2,0-4,2) pour un évènement provoqué par un facteur transitoire à au moins 7,9 /100 patients-année (IC95%, 2,0-4,2) pour un évènement non provoqué [79]. Ce taux d'incidence varie également en fonction du type d'évènement provoqué, de 1,0 /100 patients-année (IC95%, 0,0-2,3) pour un évènement provoqué par la chirurgie à 5,8 /100 patients-année (IC95%, 3,2-8,3) pour un évènement provoqué hors chirurgie. Après un 1^{er} évènement non provoqué, l'incidence cumulée de la récurrence à 10 ans de l'arrêt du traitement anticoagulant varie de 36,1% (IC95%, 27,8-45,0) à 52,6% (IC95%, 45,6-59,5) [80,81].

Le cancer est un des principaux facteurs de risque persistant ; il augmente le risque de 1^{er} évènement d'un facteur 4 à 6, avec des taux d'incidence variant entre 1,3 /100 patients-année (IC95%, 0,7-2,3) et 6,8 /100 patients-année (IC95%, 4,8-9,6) [78,82,83]. L'incidence cumulée de la récurrence à l'arrêt du traitement est d'environ 9% à 6 mois et 15% à 1 an [76,84,85].

La dérivation de modèles multivariés prédictifs du risque de récurrence à l'arrêt du traitement anticoagulant est un objectif ancien. Pour un 1^{er} évènement non provoqué, 3 principaux modèles existent : les scores HERDOO2 (*hyperpigmentation, edema, or redness in either leg, D-dimer level $\geq 250 \mu\text{g/L}$, obesity with body mass index $\geq 30 \text{ kg/m}^2$, or older age, ≥ 65 years*), de Vienne, et DASH (*D-dimer, age, sex, hormonal therapy*) [86–88]. Un 4^{ème} modèle a été dérivé plus récemment, dans le cadre de l'étude PROSPERO [89]. Seul le score HERDOO2 a bénéficié d'une étude de validation externe prospective [90]. Dans le cadre des évènements thromboemboliques veineux associés au cancer, le score d'Ottawa a été proposé pour le risque de récurrence mais sous traitement anticoagulant, et aucun modèle à l'arrêt du traitement [91].

La cohorte RIETE est une cohorte prospective multicentrique internationale, incluant de manière consécutive tous les patients diagnostiqués avec une TVP, EP ou TVS symptomatique confirmées par des examens objectifs [92]. Les patients sont suivis au moins 3 mois. Il s'agit de la plus importante cohorte internationale sur cette pathologie. Elle a donné lieu à de très nombreuses publications. En particulier, deux modèles ont été dérivés concernant la prédiction de la mortalité chez les patients présentant une thrombose associée au cancer [93,94]. Le cancer apparaissait également comme facteur associé à un mauvais pronostic dans un autre modèle dérivé à partir de cette cohorte [95]. Récemment, des méthodes d'apprentissage automatique (arbres de décision, algorithme des k plus proches voisins, machine à vecteurs de support, réseaux de neurones) ont été appliqués à la prédiction du devenir des patients arrêtant de manière prématurée le traitement anticoagulant pour une EP [96].

3.2. Logiciel utilisé

Les analyses ont été réalisées à l'aide du logiciel statistique STATA (StataCorp LLC®, College Station, Texas), version STATA/SE 14.2 pour les premières analyses et version STATA/SE 17.1 pour les analyses les plus récentes.

3.3. Différence entre modèle explicatif et modèle descriptif, et exemples de graphes acycliques dirigés

3.3.1. Article “Mortality, Cardiovascular and Limb events in Patients with Symptomatic Lower Extremity Artery Disease and Diabetes”

3.3.1.1. Manuscrit

Mortality, Cardiovascular and Limb events in Patients with Symptomatic Lower Extremity Artery Disease and Diabetes

François-Xavier Lapébie, MD^{1,2}, Vanina Bongard, MD, PhD^{2,3,4}, Philippe Lacroix, MD^{5,6}, Victor Aboyans, MD, PhD^{6,7}, Joël Constans, MD, PhD^{8,9}, Carine Boulon, MD⁸, Emmanuel Messas, MD, PhD^{10,11}, Florence Thomas-Delecourt, PharmD¹², David Rosenbaum, MD, PhD¹², Jean Ferrières, MD, PhD^{2,3,4}, Alessandra Bura-Rivière, MD, PhD^{1,13}

¹Department of Vascular Medicine, Toulouse University Hospital, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France

²UMR 1295 INSERM, CERPOP, Toulouse III – Paul Sabatier University, 37 allée Jules Guesde, 31073 Toulouse Cedex, France

³Department of Epidemiology, Toulouse University Hospital, 37 allée Jules Guesde, 31073 Toulouse Cedex, France

⁴Federation of Cardiology, Toulouse University Hospital, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France

⁵Department of Cardiovascular and Thoracic Surgery – Vascular Medicine, Limoges University Hospital, 2 avenue Martin Luther King, 87042 Limoges Cedex, France

⁶UMR 1094 INSERM & IRD, Limoges University, 2 rue du Dr Marcland, 87025 Limoges Cedex, France

⁷Department of Cardiology, Limoges University Hospital, Limoges, 2 avenue Martin Luther King, 87042 Limoges Cedex, France

⁸Department of Vascular Medicine, Bordeaux University Hospital, 1 rue Jean Burguet, 33076 Bordeaux Cedex, France

⁹Bordeaux University, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France

¹⁰Department of Vascular Medicine, Assistance Publique – Hôpitaux de Paris, 20 rue Leblanc, 75908 Paris Cedex 15, France

¹¹UMR 970 INSERM, Paris Descartes University, 56 rue Leblanc, 75015 Paris, France

¹²AstraZeneca, Tour Carpe Diem, 31 place des Corolles, 92400 Courbevoie, France

¹³UMR 1031 INSERM, StromaLab, Toulouse III – Paul Sabatier University, 4 bis avenue Hubert Curien, 31100 Toulouse, France

Corresponding author

François-Xavier Lapébie

Department of vascular medicine, Toulouse university hospital, Toulouse, France

Hôpital Rangueil, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France

Phone number: +33 5 61 32 24 38

Fax number: +33 5 61 32 26 34

E-mail: lapebie.fx@chu-toulouse.fr; fx.lapebie@gmail.com

E-mail addresses of co-authors

Vanina Bongard: vanina.bongard@univ-tlse3.fr

Philippe Lacroix: philippe.lacroix@unilim.fr

Victor Aboyans: victor.aboyans@chu-limoges.fr

Joël Constans: joel.constans@chu-bordeaux.fr

Carine Boulon: carine.boulon@chu-bordeaux.fr

Emmanuel Messas: emmanuel.messas@aphp.fr

Florence Thomas-Delecourt: florence.thomas@astrazeneca.com

David Rosenbaum: david.rosenbaum@astrazeneca.com

Jean Ferrières: jean.ferrieres@univ-tlse3.fr

Alessandra Bura-Rivière: bura-riviere.a@chu-toulouse.fr

Notice of grant support

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: AstraZeneca France funded the statistical analysis [grant number D1843R00286]. AstraZeneca was involved in the study design. AstraZeneca was not involved in the collection, analysis, and interpretation of data, in the writing of the report, and in the decision to submit the article for publication.

Abstract

The aim of this study was to compare the prognosis of patients according to diabetes status, during a 1-year follow-up after hospital admission for lower extremity artery disease, in the prospective COPART (COhorte de Patients ARTériopathes) registry. Inclusion criteria were intermittent claudication, ischemic rest pain, tissue loss, or acute limb ischemia, with radiological and hemodynamic confirmation. Among 2 494 patients, 1 235 (49.5%) had diabetes. Incidence rates for major adverse cardiovascular events (MACE) were 18.0/100 person-years (95% confidence interval [CI], 15.4-21.0) for the diabetes group and 11.1/100 person-years (95%CI, 9.2-13.4) for the non-diabetes group. Incidence rate of all-cause mortality were 29.8/100 person-years (95%CI, 26.5-33.4) for the diabetes group and 19.7/100 person-years (95%CI, 17.2-22.7) for the non-diabetes group. Incidence rates of major limb amputation were 24.2/100 person-years (95%CI, 21.1-27.8) for the diabetes group and 11.6/100 person-years (95%CI, 9.6-14.0) for the non-diabetes group. Diabetes was associated with MACE, adjusted HR 1.60 (95%CI, 1.16-2.22), and all-cause mortality, unadjusted HR 1.49 (95%CI, 1.24-1.78). In the multivariate analysis, diabetes was no longer associated with major amputation, adjusted HR 1.15 (95%CI, 0.87-1.51). Patients hospitalized for lower extremity artery disease with diabetes had a higher risk of MACE than those without diabetes.

Keywords: peripheral artery disease, diabetes mellitus, cohort studies, major adverse cardiovascular events, mortality, amputation.

Running title

Lower extremity artery disease and diabetes

Introduction

Patients with diabetes are at high risk for cardiovascular (CV) events, including CV-death, myocardial infarction (MI) and ischemic stroke (IS) [97–99]. Diabetes is especially associated with a risk of lower extremity artery disease (LEAD), with an odds ratio (OR) of 1.68 (95% confidence interval [CI], 1.53-1.84) [100]. In patients with type 2 diabetes, the risk of LEAD was higher in the presence of high HbA1c, microvascular disease, and other CV risk factors [101,102].

Furthermore, diabetes worsens LEAD prognosis, with a higher risk of all-cause mortality, CV death, MI, IS, acute limb ischemia, amputation and revascularization [103–105]. The magnitude of this risk varies between published data: event rates vary between patients diagnosed using only the ankle-brachial index (ABI), who may be asymptomatic, patients with intermittent claudication (IC), and those with chronic limb threatening ischemia (CLTI) [106]. Besides, results may differ between randomized-controlled trials (RCT) and cohort studies and large contemporary studies are lacking [106]. Indeed, the actual best medical therapy along with recent sodium glucose transporter-2 (SGLT2) inhibitors or glucagon peptide-1 (GLP-1) receptor agonists, by reducing the likelihood of adverse CV events, could modified the impact of diabetes on these outcomes [66]. However, in a subgroup of the EUCLID (Examining Use of Ticagrelor in Peripheral Artery Disease) trial, a large RCT with 13 885 LEAD patients (but 19% asymptomatic), patients with diabetes remained at higher risk for CV and limb ischemic events, despite contemporary risk reduction treatment [107].

Our objective was to assess the real-life contemporary risk of major adverse CV events (MACE), all-cause death, and major amputation in patients with LEAD and diabetes compared with those with LEAD alone in the COPART (COhorte de Patients ARTériopathes) registry, during a 1-year follow-up after hospital admission for symptomatic LEAD. Because COPART only includes patients with symptomatic LEAD hospitalized for its management, we hypothesized that the incidence of CV and limb events was higher than in others cohort studies. Additionally, we investigated the impact and causal link of diabetes on MACE and major amputation.

Methods

We used the COPART registry, an ongoing multicentre, observational, cohort study, since 2006 prospectively collecting data on consecutive patients hospitalized for symptomatic LEAD in the vascular medicine departments of 4 French academic centers (Bordeaux, Limoges, Paris and Toulouse). Details on the study protocol have been published elsewhere [68]. Patient care was provided according to the international guidelines [66,67,108,109]. After the initial hospitalization, patients were followed for at least 12 months. Data were prospectively collected by trained medical COPART investigators, using a computerized standardized case report form, and using all records available during hospitalization and follow-up. Patients, their families and/or their physicians were systematically contacted 1 year after hospital discharge. The patients gave their informed consent to participate. The study complies with the declaration of Helsinki and was approved by the Ethic Committee of Toulouse University Hospital. The current report is based on a database locked in October 2017, including patients recruited from 2006 to 2015. This database was recently used for 2 other studies, for different purposes and results [110,111].

COPART registry inclusion and exclusion criteria are listed in Supplemental Table 1, and have been published elsewhere [68]. All the patients included in the present analysis were symptomatic and underwent imaging of the lower limb arteries (duplex ultrasound, angiography computerized tomography, or angiography), confirming arterial stenosis or occlusion, and had a low ABI, toe-brachial index (TBI), or transcutaneous oxygen pressure (TcPO₂), confirming limb ischemia.

For this analysis, the main outcome was a composite of CV death, MI and IS, which is often referred as MACE, during 1-year following hospital admission. The secondary outcomes were all-cause mortality, CV mortality, non-CV mortality, MI, IS, major amputation, adverse limb events (ALE: a composite of leg revascularization and minor or major amputation), leg revascularization and minor and major amputations. Revascularization could be achieved by endovascular therapy, including percutaneous angioplasty (with or without stenting), subintimal recanalization and embolectomy, open surgery, including endarterectomy and surgical vein or prosthetic bypass, or hybrid surgery, which associates endovascular therapy and open surgical treatment. Amputation above the ankle level was considered major, and minor when below.

Details regarding baseline variables collected have already been published elsewhere [68]. In particular, diabetes mellitus was defined by documented medical history, the use of a hypoglycemic drug treatment or fasting plasma glucose levels ≥ 1.26 g/L. Type of diabetes was not recorded, but it was mainly type 2. Physical activity was defined by moderate to high levels of physical activity at least twice a week. Dyslipidemia was defined by a medical history, use of lipid-lowering agents, fasting low-density lipoprotein (LDL)-cholesterol ≥ 160 mg/dL or fasting triglycerides ≥ 200 mg/dL. Metabolic syndrome was defined in the presence of any 3 of 5 risks factors: waist circumference ≥ 94 cm for men and ≥ 80 for women or body mass index (BMI) ≥ 25 kg/m² if waist circumference is lacking, fasting triglycerides ≥ 150 mg/dL or drug treatment for elevated triglycerides, fasting high-density lipoprotein (HDL)-cholesterol < 40 mg/dL in men or < 50 mg/dL in women or drug treatment for reduced HDL-cholesterol, systolic blood pressure ≥ 130 mmHg and/or diastolic blood pressure ≥ 85 mmHg or antihypertensive drug treatment for hypertension, fasting glucose ≥ 1.0 g/L or drug treatment of elevated glucose [112]. CLTI was defined by LEAD category 4, 5 or 6 (Rutherford classification) and ankle pressure < 50 mmHg or systolic toe pressure < 30 mmHg or TcPO₂ < 30 mmHg [62]. Estimated glomerular filtration rate (eGFR) was reported according to the Chronic Kidney Disease-Epidemiology collaboration (CKD-EPI) equation, and chronic kidney disease was defined when eGFR < 60 mL/min/1.73m², with eGFR categories as follows: moderately reduced when eGFR < 60 mL/min/1.73 m² and ≥ 30 mL/min/1.73 m², and severely reduced when eGFR < 30 mL/min/1.73 m² but without dialysis [113,114]. Anemia was diagnosed when the hemoglobin value measured at admission was < 13 g/dL in men and < 12 g/dL in women. Biological inflammatory syndrome was diagnosed when C-reactive protein (CRP) ≥ 10 mg/L.

Statistical analysis

Categorical variables were presented as numbers and percentages. Continuous variables were presented as means and standard deviations (SD). Comparisons were made using Chi² test for categorical variables or Fisher exact test when appropriate, while Student t test and Mann-Whitney test were used for continuous variables. The index date was the date of admission of the index hospitalization. The Kaplan-Meier method was used to estimate cumulative incidence of events and differences between patients with diabetes and patients without diabetes were analyzed with the log-rank test. The incidence rate for each event occurring during the follow-up was estimated from the number of patients developing the considered event during the corresponding total number of patient-years of follow-up and were expressed per 100 person-years. Patients were censored at the date of the first considered event.

If they did not develop the event during the year of follow-up, they were censored at the date of the 1-year final record completion or at the date of loss of follow-up.

Unadjusted and multivariate analyses were performed using Cox proportional-hazard models, after checking for the proportional-hazard assumption and the log-linearity assumption for continuous independent variables. As a sensitivity analysis, unadjusted Fine and Gray proportional-hazard models were carried out for the major amputation outcome, with all-cause mortality as the competing event. For the first multivariate model, in order to specifically evaluate the relationship between diabetes and MACE, a directed acyclic graph (DAG) was compiled prior to analysis to determine the causal link. Variables were chosen based on knowledge regarding these associations from the literature review, regardless the p-values in the unadjusted models. Only the variables known as being confounding factors (factors that influence both diabetes onset and MACE) were included. Other variables were not considered. Especially, coronary artery disease (CAD), cerebrovascular disease and heart failure were not kept in this model because these are intermediate variables, i.e. variables on a causal path from exposure (diabetes) to outcome (MACE); this over-adjustment can increase bias due to a reduction in the total causal effect. The following independent variables were included in the first model: diabetes, age, sex, physical activity, BMI as a categorical variable. Hazard ratios (HR) were estimated with a 95%CI. For the second multivariate model, we wanted to assess predictors of major amputation. Variables associated with major amputation with a $p < 0.20$ in an unadjusted Cox analysis were introduced in a multivariate Cox model. These variables were: gender, age, diabetes, dyslipidemia, hypertension, current smoking, metabolic syndrome, family history of CAD, cerebrovascular disease, heart failure, atrial fibrillation, LEAD classification, prior limb revascularization, prior limb amputation, CLTI, gangrene, eGFR, anemia, inflammatory syndrome, aspirin or clopidogrel, vitamin K antagonist (VKA), statin, angiotensin-converting-enzyme inhibitors (ACEI) and betablocker. Then a step-by-step backward procedure was applied until only the variables significantly and independently associated with the risk of complication remained. Diabetes was kept in the final model regardless its p-value because it was a variable of interest. No imputation was made on missing data. A 2-sided $p < 0.05$ was considered significant. Analyses were performed on STATA Statistical software (release 14.2, StataCorp LLC®, College Station, Texas).

Results

At the latest data export, the COPART registry consisted of 2 520 patients. Among them, 26 were excluded from analyses for the following reasons: unknown inclusion date (n=2), unknown LEAD classification (n=1), asymptomatic LEAD (n=21), diabetes status unknown (n=2). Eventually, 2 494 patients were eligible for this study: 1 235 patients (49.5%) had diabetes (type 2 or type 1) and 1 259 patients (50.5%) did not. The mean duration of the follow-up was 287 ± 134 days for the whole cohort, with a mean of 281 ± 136 days for the diabetes group and 293 ± 132 days for the non-diabetes group. During the 1-year follow-up, 483 patients died (138 during index hospitalization including 78 with diabetes, 345 after hospital discharge including 205 with diabetes), 1 755 patients were alive at 365 days and completed the follow-up, and 256 (115 with diabetes) were lost to follow-up before 365 days with a mean duration of follow-up reaching 80 ± 121 days.

At admission, hypertension, obesity, CV diseases and renal insufficiency were more frequent in the diabetes group than in the non-diabetes group (Table 1). Prior limb amputation was encountered in more diabetic than non-diabetic patients. LEAD at hospital admission was often severe, with Rutherford category 5 or 6, CLTI and gangrene being more frequent in the diabetes group than in the non-diabetes group; furthermore, 19.5% of patients had been transferred from other hospitals to COPART tertiary centers for LEAD management. Renin-angiotensin system inhibitors (RASi) and beta-blockers were more frequent in the diabetes group than in the non-diabetes group. Only 59.5% of patients received a statin at hospital admission, without difference between diabetic and non-diabetic patients. This proportion increased with the year of inclusion; up to 68.4% for patients included in the last year. Statins were also introduced during the hospitalization, with 73.8% of patients on treatment at hospital discharge (89.3% for patients included in the last year).

Incidence rate of MACE was 18.0/100 person-years (95%CI, 15.4-21.0) for the diabetes group and 11.1/100 person-years (95%CI, 9.2-13.4) for the non-diabetes group (Figure 1). Cumulative incidence of MACE at 1 year was 15.5% (95%CI, 13.4-17.8) for the diabetes group and 10.0% (8.4-11.9) for the non-diabetes group (Figure 2), $p<0.001$ for the log-rank test. Incidence rate of all-cause mortality was 29.8/100 person-years (95%CI, 26.5-33.4) for the diabetes group and 19.7/100 person-years (95%CI, 17.2-22.7) for the non-diabetes group (Figure 1). Cumulative incidence of all-cause mortality at 1 year was 24.6% (95%CI, 22.2-27.2) for the diabetes group and 17.1% (15.1-19.4) for the non-diabetes group (Supplemental Figure 1), $p<0.001$ for the log-rank test. Incidence rate of major limb amputation was 24.2/100 person-

years (95%CI, 21.1-27.8) for the diabetes group and 11.6/100 person-years (95%CI, 9.6-14.0) for the non-diabetes group (Figure 1). Cumulative incidence of major amputation at 1 year was 18.5% (95%CI, 16.3-20.9) for the diabetes group and 9.9% (95%CI, 8.3-11.8) for the non-diabetes group (Figure 3), $p < 0.001$ for the log-rank test. In the unadjusted Cox analysis, diabetes was associated with MACE, HR 1.58 (95%CI, 1.24-2.01), with all-cause mortality, HR 1.49 (95%CI, 1.24-1.78), and with major limb amputation, HR 1.97 (95%CI, 1.56-2.49) (Figure 4). There was no noticeable difference between unadjusted Cox and Fine and Gray analysis for the major amputation outcome (data not shown). In the multivariate Cox model, diabetes remained associated with MACE, HR 1.60 (95%CI, 1.16-2.22) (Supplemental Table 2). However, in the multivariate Cox model built to assess predictors of major limb amputation, diabetes was no more associated with an increased risk of major amputation (Table 2, Supplemental Table 3).

The impact of LEAD presentation (Rutherford classification) on the results was explored in subgroup analyses (Supplemental Tables 4, 5 and 6). The risks of MACE, all-cause death and major amputation tended to be higher in the Rutherford 5 or 6 groups than in the Rutherford 1, 2 or 3 groups. The association of diabetes with MACE tended to lose statistical significance, at least in part, due to a lack of power in the Rutherford 1, 2, 3 or 4 groups. Diabetes remained associated with all-cause death in the Rutherford 1, 2 or 3 groups, and with major amputation in the Rutherford 5 or 6 groups.

Discussion

Our study highlights that a high proportion of patients hospitalized for LEAD have diabetes: 49.5% (95%CI, 47.6-51.5). This prevalence is lower in previous studies, ranging from 19% in a health-care database comprising in-patients and outpatients, to 34.7% in a cohort of patients hospitalized for LEAD [115,116]. Our results are closer to that of a meta-analysis of studies that enrolled adult patients with severe limb ischemia, emphasizing that the proportion of diabetic patients is higher in the most severe stages of LEAD [117,118]. Diabetes accelerates LEAD progression and worsens disease severity, so a high frequency is expected in patients at the end stage of LEAD [119]. Because the COPART registry only includes patients hospitalized in tertiary centers for the management of symptomatic LEAD, ischemic rest pain, tissue loss, and acute ischemia were particularly common at admission, concerning more than half of the patients.

As expected, diabetes was associated with MACE. Of note, these association tended to lose statistical significance in subgroup analyses according to the stage of LEAD, probably due to a lack of power, at least in the Rutherford 1, 2, 3 and 4 groups. Nevertheless, the association of diabetes with MACE in the entire cohort is close to that reported in the EUCLID trial, with an adjusted HR of 1.43 (95%CI, 1.28-1.61) [107]. Whereas in several cohort studies, insulin prescription and high HbA1c level were associated with an increased risk for MACE or mortality, this association was not found in the COPART registry (data not shown) [120]. Mostly, incidences of MACE for diabetes and non-diabetes group in COPART were much higher than previously published. In a subgroup analysis of the EUCLID trial, incidences of MACE were only 5.68 per 100 person-years (95%CI, 5.28-6.11) for the diabetes group and 3.65 (95%CI, 3.40-3.92) for the non-diabetes group [107]. Even after a first major adverse limb event (MALE; the composite of major amputation, acute limb ischemia, and severe limb ischemia leading to a vascular intervention), the incidence of MACE for participants with LEAD in a secondary analysis of the recent COMPASS (Cardiovascular Outcomes for People Using Anticoagulation Strategies) trial was 5.4 per 100 person-years, therefore much lower than in the present study [121]. Patients samples in RCTs, including those focused on LEAD, are highly selected with a lower risk profile than real-world populations, and are not broadly representative of patients encountered in daily practice [5,63,122]. Indeed, compared to the EUCLID and COMPASS trials, COPART patients were older and had more severe LEAD at admission, which is associated with MACE [118].

Again, the incidence of all-cause mortality in COPART was much higher than both the EUCLID and COMPASS trials: only 4.21 per 100 person-years (95%CI, 3.87-4.57) for the diabetes group and 3.19 per 100 person-years (95%CI, 2.96-3.43) for the non-diabetes group in the EUCLID trial, and 10.5 per 100 person-years for the COMPASS trial [107,121]. Incidence of all-cause mortality for symptomatic LEAD in cohort studies was higher than these RCTs, 6.25 per 100 person-years (95%CI, 5.12-7.37) in a meta-analysis, but remained much lower than that observed in COPART [65]. This could be related to the high proportion of CLTI patients in COPART. Indeed, in subgroup analyses, the risk of all-cause death tended to be higher in the end-stages of LEAD. However, in a meta-analysis, the incidence of all-cause mortality for the CLTI patients subgroup was lower than in our study, 18.3 per 100 person-years (95%CI, 2.5-46.8) [65,106]. Assessing causes of death can be challenging, but while COPART patients are at very high CV risk, less than half of deaths were CV-related. This proportion is close to that of LEAD patients at lower CV risk than COPART patients [104,123]. In our study, the other causes of death were infections (13.3% of deaths), cancers (6.0%), respiratory failure (3.1%), hemorrhage (1.7%), kidney failure (1.2%), liver failure (1.2%), suicide (0.2%) and other non-CV death (35.4%). Of note, a recent study showed, for the first time, that an abnormal ABI is also associated with cancer mortality, HR 1.85 (95%CI, 1.09-3.15), but data on the site of cancer was missing [124]. Previously published studies also found a higher cancer mortality in symptomatic LEAD group compared with controls [125,126]. CV disease and cancer possess possible interactions. Chronic inflammation, which is notably induced by diabetes and others CV risk factors, is involved in the pathogenesis and progression of both CV disease and cancer [127]. These conditions share common risk factors, such as tobacco use [127]. Nowadays, improvements in diabetes care reduced diabetes-related mortality, so these patients are exposed to developing cancer [128]. In our study, diabetes was also associated with non-CV death.

Surprisingly, diabetes was no longer associated with major amputation in the multivariate Cox analysis. However, lower limb arterial disease in diabetic patients has specific features that should increase the risk of amputation. Diabetes affects limb arteries by promoting atherosclerotic occlusive disease, but also arterial wall stiffening related to medial calcinosis [129,130]. Unlike others CV risk factors, diabetes typically affects distal arteries, as dorsalis pedis artery, medial plantar artery and lateral plantar artery, which is associated both with CLTI, OR 13.25 (95%CI, 1.69-104.16), and have fewer revascularization possibilities [129,130]. Furthermore, the causes of diabetic foot ulcers are multiple with, in addition to LEAD, the role

of peripheral neuropathy, local infection and soft tissue and bony deformity, cannot be improved by revascularization [131]. In particular, in subgroup analyses, diabetes was associated with major amputation in the presence of a wound (Rutherford 5 or 6 group). In fact, adjusting for variables such as LEAD classification, CLTI and gangrene, which do not fulfil the definition of a confounder (they are not causes of both diabetes and amputation), but act as a mediator between diabetes and amputation, results in overadjustment with an attenuation of the true causal effect of diabetes on the risk of amputation. This was also found in an analysis of 2 RCTs, where the adjustment on LEAD classification and medial calcinosis led to a loss of the statistical significance of the association between diabetes and the risk of major amputation [132]. This association is therefore explained by a more severe LEAD at admission.

The incidence of major amputation in COPART was much higher than in the previously published data, where it ranged from less than 1 per 100 person-years to 10.0 per 100 person-years (95%CI, 1.0-23.2), although more common in diabetic patients than non-diabetic, and in case of CLTI [106,107]. The incidence of major and minor amputations in the secondary analysis of the COMPASS trial was slightly closer than our: 22.0 per 100 person-years [121]. In this secondary analysis, all patients were included following a first MACE, so they approached the COPART patients, who were all included during an initial hospitalization for symptomatic LEAD.

To summarize, as expected, diabetes was associated with MACE, and with more severe LEAD presentation at admission, which led to a high risk of major amputation. However, the incidences of MACE, all-cause death and amputations were much higher than in the previously published data. It is widely accepted that patients enrolled in CV trials are on average at lower risk because of younger age and less comorbidities [5,63,122]. Yet, our results also appear to be very different from other cohort studies, probably because our cohort only included patients hospitalized for symptomatic LEAD, with a high proportion of diabetic patients. These very high CV risk patients are likely to be among those who will benefit the most from contemporary risk reduction therapies, including SGLT2 inhibitors, GLP-1 receptor agonists, and low-dose rivaroxaban plus aspirin [133–135]. Concerns have been raised about the risk of amputation, mainly minor, with SGLT2 inhibitors, especially canagliflozin, but this has not been confirmed in 2 recent meta-analysis [136,137].

Our analysis has some limitations. First, information about type and onset of diabetes were not recorded. Second, since COPART only includes patients with symptomatic LEAD hospitalized in tertiary centers, we cannot generalize the results to all of the LEAD patients;

however, COPART provides important findings in this subgroup of LEAD patients. Third, because of the observational nature of the study, and despite multivariate analysis, we cannot exclude the possibility of non-controlled variables that could differ between the 2 groups; these could affect the outcomes.

Conclusion

Patients hospitalized for LEAD with diabetes had a higher risk of the composite outcome of MACE and its components (CV death, MI and IS), all-cause mortality, and major amputation than those without diabetes. Incidence rates of mortality, CV-related or not, and major amputation were particularly higher in both groups than in previously published data. Because these outcomes remain poor in patients requiring hospitalization due to LEAD, this population should be clearly individualized in LEAD studies and should benefit from the best risk reduction therapies.

Acknowledgements

We thank the research-assistants team: Delphine Baudoin, Clémence Labetoulle, Emmanuelle Lorian, Marion Mangin, Béatrice Montalègre, Monique Pailhol, Brigitte Pelvet, and Irina Smoliy for their technical help and data collection. We thank Katja Berg, Dr Narinder Bhalla, Dr Björn Karlson, Dr Marco Kuster, and Eva Lesén for their critical reading of a previous version of the manuscript.

Conflicts of interest

Dr Lapébie, Dr Lacroix and Dr Messas have nothing to disclose. Dr Bongard reports personal fees from Sanofi outside the submitted work. Dr Constans and Dr Boulon reports grants from AstraZeneca outside the submitted work. Dr Aboyans reports personal fees from Bayer Healthcare, BMS/Pfizer, Novartis, and Sanofi, outside the submitted work. Dr Thomas-Delecourt and Dr David Rosenbaum are employees at AstraZeneca. Dr Ferrières reports personal fees from Amgen, Sanofi, MSD, and Servier, outside the submitted work. Dr Bura-Rivière reports grants, personal fees, non-financial support and other from Bayer Healthcare and BMS/Pfizer, grants and non-financial support from Astra-Zeneca, outside the submitted work.

Author contribution

All authors contributed to: (1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data, (2) drafting the article or revising it critically for important intellectual content, and, (3) final approval of the version to be published.

Tables

Table 1: Demographic and clinical characteristics of the study population at baseline (inclusion in COPART registry, i.e. hospital admission).

Characteristics	All (n = 2,494)	Diabetes (n = 1,235)	No diabetes (n= 1,259)	p
<i>Demographic data</i>				
Men	1868/2485 (75.2%)	934/1233 (75.8%)	934/1252 (74.6%)	0.507
Age – years	70.6 ± 12.8	71.8 ± 12.8	69.3 ± 13.8	<0.001
Regular physical activity	639/1762 (36.3%)	289/847 (34.1%)	350/914 (38.3%)	0.069
<i>Cardiovascular risk factors</i>				
Dyslipidemia	1822/2491 (73.1%)	955/1232 (77.5%)	867/1259 (68.9%)	<0.001
Hypertension	2137/2493 (85.7%)	1118/1234 (90.6%)	1019/1259 (80.9%)	<0.001
Current smokers	624/2461 (25.4%)	192/1218 (15.8%)	432/1243 (34.8%)	<0.001
Obesity	462/2300 (20.1%)	333/1138 (29.3%)	129/1162 (11.1%)	<0.001
Metabolic syndrome	1640/2226 (73.7%)	1038/1143 (90.8%)	602/1083 (55.6%)	<0.001
Family history of CAD	282/2151 (13.1%)	121/1045 (11.6%)	161/1106 (14.6%)	0.041
<i>Medical history</i>				
CAD	911/2478 (36.8%)	503/1225 (41.1%)	408/1253 (32.6%)	<0.001
Ischemic stroke	358/2457 (14.6%)	193/1214 (15.9%)	165/1243 (13.3%)	0.065
Heart failure	297/2453 (12.1%)	177/1207 (14.7%)	120/1246 (9.6%)	<0.001
Atrial fibrillation	501/2452 (20.4%)	277/1205 (23.0%)	224/1247 (18.0%)	0.002
COPD or asthma	454/2418 (18.8%)	223/1191 (18.7%)	231/1227 (18.8%)	0.948
<i>LEAD at admission</i>				
Newly diagnosed LEAD	711/2486 (28.6%)	354/1232 (28.7%)	357/1254 (28.5%)	0.884
Prior revascularization	849/2468 (34.4%)	400/1221 (32.8%)	449/1247 (36.0%)	0.090
Prior amputation	333/2468 (13.5%)	239/1221 (19.6%)	94/1247 (7.5%)	<0.001
LEAD classification				<0.001
Rutherford 1, 2 or 3	724/2494 (29.0%)	239/1235 (19.4%)	485/1259 (38.5%)	
Rutherford 4	236/2494 (9.5%)	87/1235 (7.0%)	149/1259 (11.8%)	

Rutherford 5 or 6	1322/2494 (53.0%)	835/1235 (67.6%)	487/1259 (38.7%)	
Acute limb ischemia	212/2494 (8.5%)	74/1235 (6.0%)	138/1259 (11.0%)	
CLTI	812/2164 (37.5%)	479/1040 (46.1%)	333/1124 (29.6%)	<0.001
Gangrene	306/2432 (12.6%)	200/1207 (16.6%)	106/1225 (8.7%)	<0.001
<i>Laboratory tests at admission</i>				
eGFR – mL/min/1.73 m ²	66.0 ± 31.2	59.0 ± 31.9	72.8 ± 28.8	<0.001
Chronic kidney disease				<0.001
None	1431/2440 (58.7%)	603/1212 (49.8%)	828/1228 (67.4%)	
Moderate	676/2440 (27.7%)	370/1212 (30.5%)	306/1228 (24.9%)	
Severe	177/2440 (7.3%)	129/1212 (10.6%)	48/1228 (3.9%)	
Dialysis	156/2440 (6.4%)	110/1212 (9.1%)	46/1228 (3.8%)	
Anemia	1246/2375 (52.5%)	724/1180 (61.4%)	522/1195 (43.7%)	<0.001
Inflammatory syndrome	1253/2246 (55.8%)	717/1130 (63.5%)	536/1116 (48.0%)	<0.001
HbA _{1c} – %		7.5 ± 1.5		
<i>Treatment at admission</i>				
Aspirin or clopidogrel or oral anticoagulant	1826/2172 (84.1%)	910/1082 (84.1%)	916/1090 (84.0%)	0.966
Aspirin or clopidogrel	1571/2123 (74.0%)	773/1066 (72.5%)	798/1057 (75.5%)	0.117
Aspirin	1175/2118 (55.5%)	583/1064 (54.8%)	592/1054 (56.2%)	0.525
Clopidogrel	624/2114 (29.5%)	317/1061 (29.9%)	307/1053 (29.2%)	0.716
Oral anticoagulant	336/2468 (13.6%)	178/1227 (14.5%)	158/1241 (12.7%)	0.199
Statin	1258/2115 (59.5%)	644/1059 (60.8%)	614/1056 (58.1%)	0.211
ACEI or ARB	1208/2115 (57.1%)	673/1062 (63.4%)	535/1053 (50.8%)	<0.001
ACEI	683/2108 (32.4%)	364/1056 (34.5%)	319/1052 (30.3%)	0.042
ARB	543/2107 (25.8%)	321/1058 (30.3%)	222/1049 (21.2%)	<0.001
Beta-blockers	720/2111 (34.1%)	393/1060 (37.1%)	327/1051 (31.1%)	0.004
Antidiabetic medications				

No treatment		226/1227 (18.4%)		
OAD only		358/1227 (29.2%)		
Insulin only		474/1227 (38.6%)		
OAD and insulin		169/1227 (13.8%)		

Data are presented as n (%) or mean \pm standard deviation (SD). ACEI: angiotensin-converting-enzyme inhibitors. ARB: angiotensin-receptor blockers. CAD: coronary artery disease. CLTI: chronic limb threatening ischemia. COPD: chronic obstructive pulmonary disease. eGFR: estimated glomerular filtration rate. LEAD: lower extremity artery disease. OAD: oral antidiabetic drugs.

Table 2: Factors associated with major amputation during the 1-year follow-up in a multivariate Cox regression analysis (with diabetes kept in the model regardless its p-value).

Factors	Hazard ratio	95%CI	p
Diabetes mellitus	1.15	0.87-1.51	0.339
Prior amputation	1.74	1.30-2.34	<0.001
LEAD classification			
Rutherford 1, 2 or 3	Ref.	Ref.	Ref.
Rutherford 4	20.75	2.63-163.54	0.004
Rutherford 5 or 6	51.75	7.04-380.16	<0.001
Acute limb ischemia	113.72	15.53-832.82	<0.001
Chronic limb threatening ischemia	2.20	1.54-3.14	<0.001
Gangrene	1.62	1.20-2.20	0.002
Anemia	1.77	1.31-2.40	<0.001

95%CI: confidence interval. LEAD: lower extremity artery disease. Ref.: reference.

Supplemental Table 1: COPART registry inclusion and exclusion criteria.

Inclusion criteria	Consent to participate into the study	
	Age >18 years	
	<p>Hospitalization specifically for management (including revascularization, wound care, and disease assessment) of symptomatic LEAD of atherosclerotic origin. The clinical presentation should be one of the following:</p>	<ul style="list-style-type: none"> a. an intermittent claudication, associated with: <ul style="list-style-type: none"> i. an ABI <0.90 or >1.30, <li style="text-align: center;">OR ii. in the case of normal ABI at rest, a positive treadmill test (Strandness protocol, 3.2 km/h on a 10% slope), <li style="text-align: center;">OR iii. in the case of normal ABI at rest, an arterial stenosis >50% on duplex ultrasound or angiography computerized tomography or angiography; b. an ischemic rest pain, ulceration or gangrene, related to a documented LEAD with significant arterial stenosis; c. an acute lower-limb ischemia related to a documented LEAD with significant arterial stenosis.
Exclusion criteria	Impossible follow-up (i.e. patients who cannot be contacted after hospitalization, follow-up carried out in a foreign country)	
	Arterial occlusive disease not related to atherosclerosis (i.e. thromboangiitis obliterans, popliteal artery entrapment syndrome, iliac artery endofibrosis, vasculitis)	
	Acute ischemia without LEAD (i.e. embolism caused by atrial fibrillation)	

Supplemental Table 2: Association between diabetes and major adverse cardiovascular events during the 1-year follow-up in a multivariate Cox regression analysis.

Factors	Hazard ratio	95%CI	p
Diabetes mellitus	1.60	1.16-2.22	0.005
Age (for an increase of one year)	1.05	1.03-1.06	<0.001
Men	1.11	0.77-1.59	0.573
Body mass index			
BMI ≥ 18.5 and < 25 kg/m ²	Ref.	Ref.	Ref.
BMI < 16.5 kg/m ²	0.81	0.20-3.32	0.772
BMI ≥ 16.5 and < 18.5 kg/m ²	0.67	0.27-1.67	0.392
BMI ≥ 25 and < 30 kg/m ²	0.74	0.51-1.07	0.106
BMI > 30 kg/m ²	0.73	0.47-1.12	0.151
Regular physical activity	0.63	0.45-0.89	0.009

95%CI: 95% confidence interval. BMI: body mass index. Ref.: reference.

Supplemental Table 3: Unadjusted hazard ratios for major amputation during the 1-year follow-up (unadjusted Cox analysis).

Factors	Hazard ratio	95%CI	p
Demographic data			
Men	1.21	0.92-1.60	0.165
Age in years (for an increase of 1 year)	1.02	1.01-1.02	0.001
Regular physical activity	0.58	0.43-0.78	<0.001
Cardiovascular risk factors			
Diabetes	1.97	1.56-2.49	<0.001
Dyslipidemia	0.63	0.50-0.79	<0.001
Hypertension	0.78	0.58-1.05	0.103
Current smokers	0.59	0.44-0.79	<0.001
Obesity	1.08	0.81-1.44	0.596
Metabolic syndrome	1.26	0.95-1.68	0.104
Family history of CAD	0.77	0.52-1.14	0.195
Medical history			
CAD	0.95	0.75-1.20	0.677
Ischemic stroke	1.44	1.08-1.92	0.012
Heart failure	1.82	1.35-2.44	<0.001
Atrial fibrillation	1.57	1.21-2.02	0.001
COPD or asthma	1.18	0.90-1.56	0.237
LEAD at admittance			
Newly diagnosed LEAD	0.94	0.74-1.21	0.647
Prior revascularization	1.38	1.10-1.73	0.006
Prior amputation	3.06	2.40-3.91	<0.001
LEAD classification			
Rutherford 1, 2 or 3	Ref.	Ref.	Ref.
Rutherford 4	45.73	6.04-346.17	<0.001
Rutherford 5 or 6	163.90	23.00-1168.13	<0.001
Acute limb ischemia	161.60	22.23-1174.86	<0.001
CLTI	3.68	2.84-4.78	<0.001
Gangrene	2.74§	2.11-3.55	<0.001
Laboratory exams at admittance			
eGFR (for an increase of 1 mL/min/1.73 m ²)	0.99†	0.99-1.00	0.004
Chronic kidney disease			
None	Ref.	Ref.	Ref.
Moderate	1.11	0.86-1.45	0.424
Severe	1.77	1.20-2.62	0.004
Dialysis	1.78	1.20-2.64	0.004
Anemia	3.42	2.61-4.50	<0.001

Inflammatory syndrome	5.21	3.78-7.18	<0.001
Treatment at admittance			
Aspirin or clopidogrel or oral anticoagulant	0.67	0.51-0.89	0.006
Aspirin or clopidogrel	0.59	0.46-0.75	<0.001
Aspirin	0.71	0.56-0.89	0.003
Clopidogrel	0.80	0.61-1.04	0.093
Oral anticoagulant	1.85	1.41-2.43	<0.001
Statin	0.68	0.54-0.85	0.001
ACEI or ARB	0.88§	0.70-1.12	0.299
ACEI	0.82§	0.63-1.05	0.117
ARB	1.05	0.80-1.36	0.736
Beta-blockers	0.72	0.55-0.93	0.011
Antidiabetic medications (only for the diabetes group)			
No treatment	Ref.	Ref.	Ref.
OAD only	0.85	0.56-1.31	0.468
Insulin only	1.23	0.83-1.81	0.295
OAD and insulin	0.62	0.35-1.09	0.095

95%CI: 95% confidence interval. ACEI: angiotensin-converting-enzyme inhibitors. ARB: angiotensin-receptor blockers. CAD: coronary artery disease. CLTI: chronic limb threatening ischemia. COPD: chronic obstructive pulmonary disease. eGFR: estimated glomerular filtration rate. LEAD: lower extremity artery disease. OAD: oral antidiabetic drugs. Ref.: reference. SD: standard deviation. §Proportional hazard assumption not respected. †Log-linearity assumption not respected.

Supplemental Table 4: MACE in patients with diabetes and control subjects, according to the Rutherford classification of LEAD.

	Incidence rate, per 100 person-years (95%CI)		One-year cumulative incidence, % (95%CI)		p of log-rank test	HR (95%CI)
	Diabetes	No diabetes	Diabetes	No diabetes		
Rutherford 1,2 or 3	8.0 (4.8-13.2)	4.3 (2.7-6.8)	7.2 (4.4-11.7)	4.2 (2.7-6.6)	0.075	1.85 (0.93-3.66)§
Rutherford 4	14.7 (8.1-26.6)	7.1 (3.7-13.7)	13.1 (7.5-22.5)	6.6 (3.5-12.3)	0.102	2.05 (0.85-4.96)
Rutherford 5 or 6	21.7 (18.3-25.8)	21.1 (16.7-26.7)	18.3 (15.6-21.4)	17.4 (14.0-21.5)	0.842	1.03 (0.77-1.38)
Acute limb ischemia	15.6 (7.8-31.3)	11.9 (6.9-20.4)	13.0 (6.6-24.4)	10.0 (5.9-16.6)	0.643	1.23 (0.51-2.97)

95%CI: 95% confidence interval. HR: hazard ratio. LEAD: lower extremity artery disease.

MACE: major adverse cardiovascular events. §Proportional hazard assumption not respected.

Supplemental Table 5: All-cause death in patients with diabetes and control subjects, according to the Rutherford classification of LEAD.

	Incidence rate, per 100 person-years (95%CI)		One-year cumulative incidence, % (95%CI)		p of log-rank test	HR (95%CI)
	Diabetes	No diabetes	Diabetes	No diabetes		
Rutherford 1,2 or 3	8.7 (5.4-13.9)	4.0 (2.5-6.4)	8.2 (5.2-12.8)	3.9 (2.5-6.2)	0.021	2.17 (1.11-4.25)
Rutherford 4	14.1 (7.8-25.5)	13.3 (8.3-21.4)	12.7 (7.3-21.8)	12.3 (7.8-19.0)	0.874	1.06 (0.50-2.27)
Rutherford 5 or 6	37.5 (33.0-42.6)	39.5 (33.4-46.8)	29.7 (26.6-33.0)	30.5 (26.5-35.1)	0.639	0.95 (0.77-1.17)
Acute limb ischemia	40.1 (26.1-61.5)	26.3 (18.2-37.8)	30.8 (21.2-43.2)	21.5 (15.5-29.4)	0.185	1.46 (0.83-2.56)

95%CI: 95% confidence interval. HR: hazard ratio. LEAD: lower extremity artery disease.

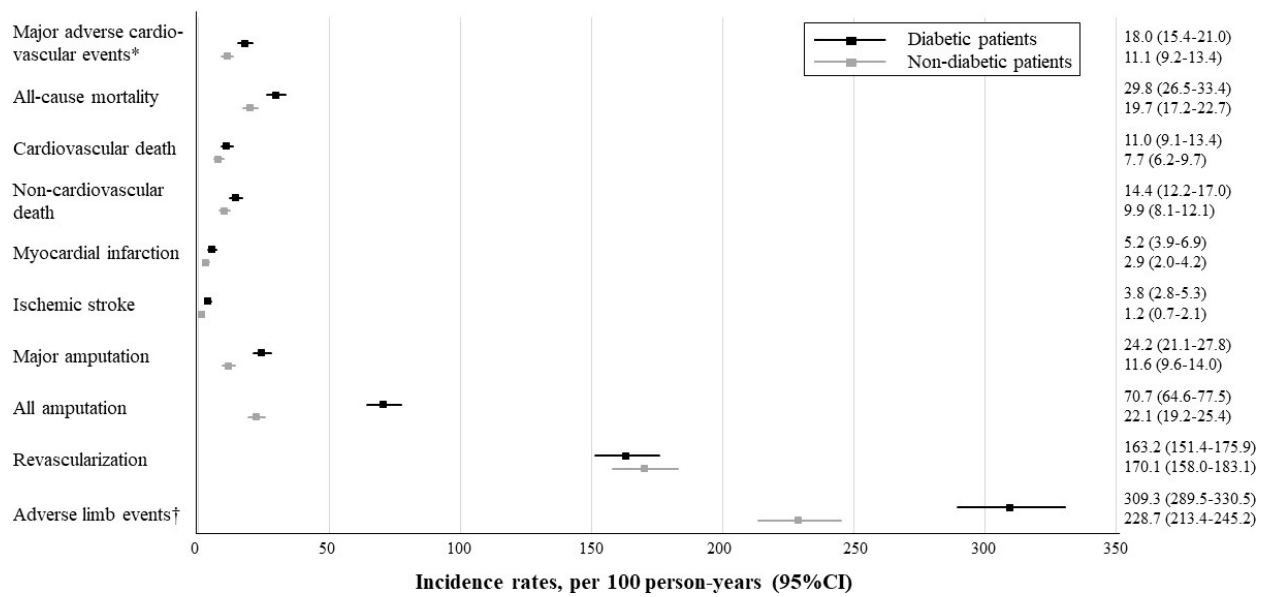
Supplemental Table 6: Major amputation in patients with diabetes and control subjects, according to the Rutherford classification of LEAD.

	Incidence rate, per 100 person-years (95%CI)		One-year cumulative incidence, % (95%CI)		p of log-rank test	HR (95%CI)
	Diabetes	No diabetes	Diabetes	No diabetes		
Rutherford 1,2 or 3	0	0.2 (0.0-1.7)	0	0.2 (0.0-1.7)	0.497	/
Rutherford 4	8.1 (3.7-18.1)	7.4 (3.9-14.3)	7.3 (3.4-15.6)	6.8 (3.6-12.6)	0.841	1.11 (0.40-3.12)
Rutherford 5 or 6	35.3 (30.5-40.9)	24.0 (19.0-30.2)	24.6 (21.6-27.9)	18.2 (14.7-22.4)	0.008	1.45 (1.10-1.90)
Acute limb ischemia	30.7 (18.2-51.8)	28.7 (19.7-41.9)	23.4 (14.6-36.5)	21.7 (15.4-30.1)	0.977	1.01 (0.53-1.93)

95%CI: 95% confidence interval. HR: hazard ratio. LEAD: lower extremity artery disease.

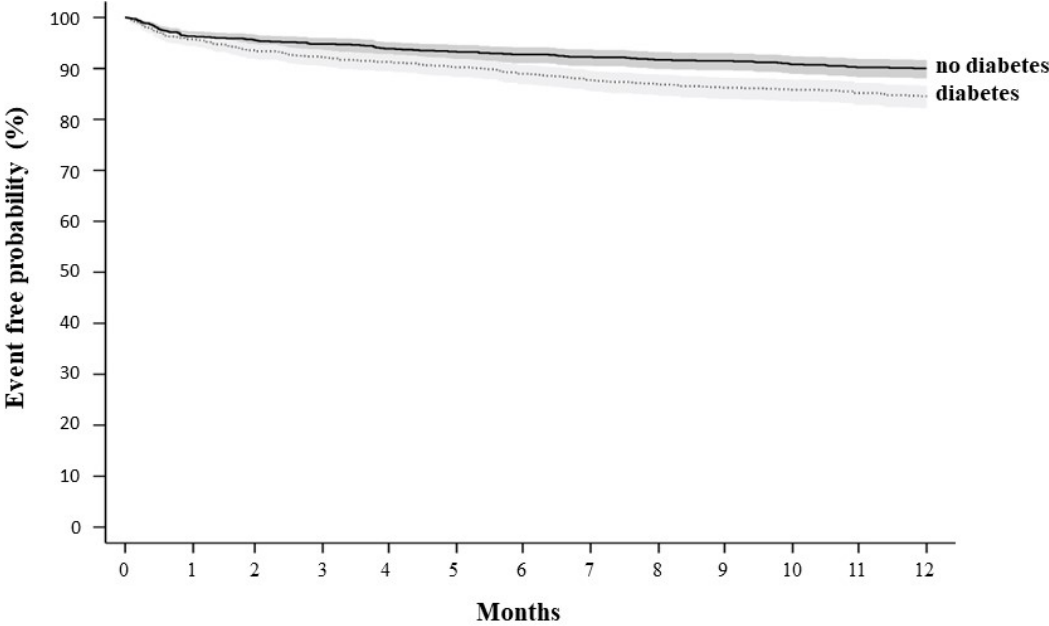
Figures

Figure 1: Incidence rates of events (per 100 person-years).



*Major adverse cardiovascular events are a composite of cardiovascular death, myocardial infarction and ischemic stroke. †Adverse limb events are a composite of leg revascularization and minor or major amputation. 95%CI: 95% confidence interval.

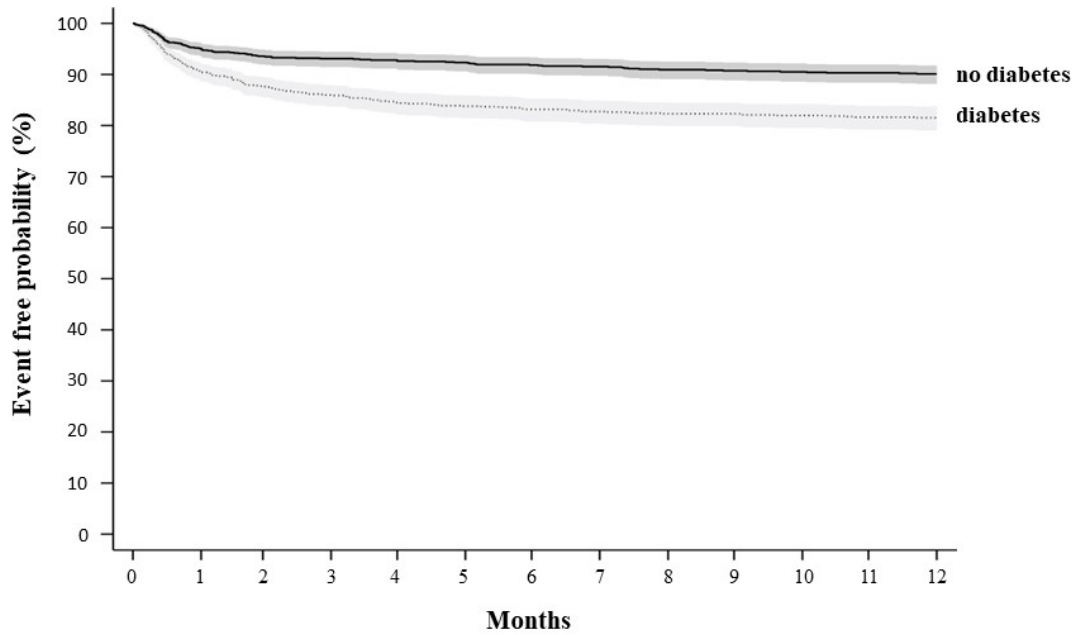
Figure 2: Kaplan-Meier analysis of major adverse cardiovascular events between the diabetes group and the non-diabetes group.



No. at Risk

Diabetes	1233	978	892	838	789
No diabetes	1254	1035	972	932	890

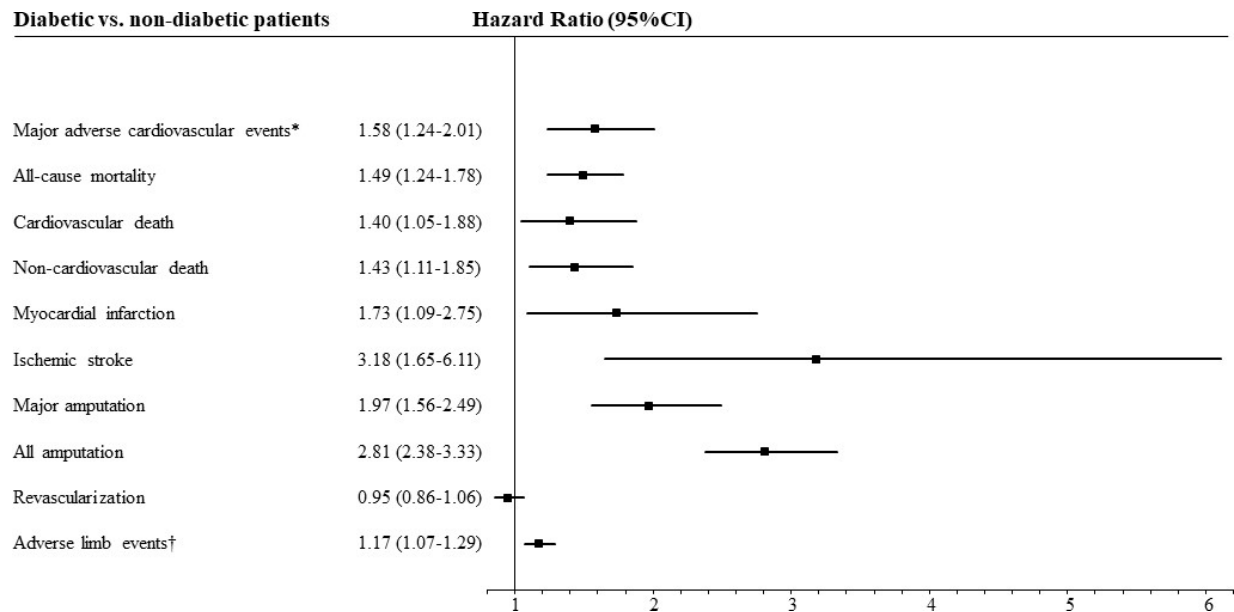
Figure 3: Kaplan-Meier analysis of major amputation between the diabetes group and the non-diabetes group.



No. at Risk

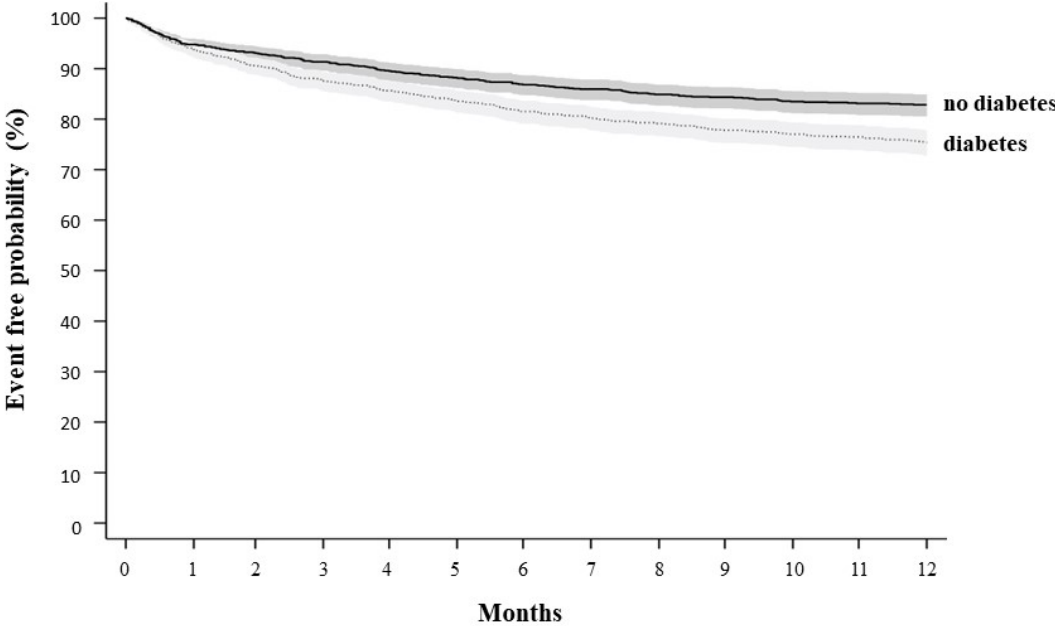
Diabetes	1233	878	789	746	701
No diabetes	1254	990	923	878	839

Figure 4: Comparison of outcomes between the diabetes group and the non-diabetes group by unadjusted Cox analysis.



Non-diabetes group used as the reference category. *Major adverse cardiovascular events are a composite of cardiovascular death, myocardial infarction and ischemic stroke. †Adverse limb events are a composite of leg revascularization and minor or major amputation. 95%CI: 95% confidence interval.

Supplemental Figure 1: Kaplan-Meier analysis of all-cause mortality between the diabetes group and the non-diabetes group.



No. at Risk

Diabetes	1233	1003	927	880	837
No diabetes	1254	1053	993	956	918

3.3.1.2.Résumé et remarques

L'objectif de cette étude réalisée à partir de la cohorte COPART était de comparer le pronostic des patients diabétiques dans l'année suivant une hospitalisation pour prise en charge d'un AAMI, à celui des patients non diabétiques. Parmi les 2494 patients inclus, 1235 (49,5%) présentaient un diabète. Les taux d'incidence des MACE, du décès toute cause et de l'amputation majeure étaient respectivement de 18,0 /100 patients-année (IC95%, 15,4-21,0), 29,8 /100 patients-année (IC95%, 26,5-33,4) et 24,2/100 patients-année (IC95%, 21,1-27,8) pour les patients diabétiques, et 11,1 /100 patients-année (IC95%, 9,2-13,4), 19,7 /100 patients-année (IC95%, 17,2-22,7) et 11,6 /100 personnes-année (IC95%, 9,6-14,0) pour les patients non diabétiques.

Ce travail illustre deux types de modèles multivariés. Afin d'étudier spécifiquement le lien de causalité entre le diabète et les MACE, un premier modèle à but explicatif a été construit, sur la base d'un DAG. Dans ce modèle, on confirmait un lien avec un HR de 1,49 (IC95%, 1,24-1,78). Concernant le risque d'amputation majeure, un second modèle à but descriptif a été proposé, avec une sélection des variables opérée par une méthode descendante. La variable diabète a été forcée dans le modèle final. Contrairement au résultat de l'analyse bivariée, le lien entre le diabète et le risque d'amputation majeure n'était plus statistiquement significatif, avec un HR de 1,15 (IC95%, 0,87-1,51). En pratique, le risque d'amputation était principalement associé à la présentation initiale de l'AAMI, dont on sait qu'elle est influencée par le diabète ; cet ajustement sur une variable de médiation a donc pu réduire l'effet causal total de l'exposition (diabète) sur l'évènement d'intérêt (amputation majeure).

Cet article a été accepté pour publication dans *Angiology* (en 2020, impact factor [IF] 3,619, rang C pour « Peripheral vascular disease »).

3.3.2. Article “Impact of angiotensin receptor blockers on mortality after hospitalization for symptomatic lower extremity artery disease”

3.3.2.1.Manuscrit

Erratum : Dans le manuscrit suivant, le risque de survenue des évènements d'intérêt est exprimé en incidence cumulée à 1 an et non pas en taux d'incidence. Par exemple, pour la mortalité dans le groupe ARA2, il faut comprendre une incidence cumulée à 1 an de 12,6% (IC95%, 9,7-16,1), et non pas un taux d'incidence de 12,6 /100 personnes-année (IC95%, 9,7-16,1) calculé pour la 1^{ère} année du suivi.

Impact of angiotensin receptor blockers on mortality after hospitalization for symptomatic lower extremity artery disease

François-Xavier Lapébie, MD^{1,2,*}, Alessandra Bura-Rivière, MD, PhD^{1,3}, Philippe Lacroix, MD^{4,5}, Joël Constans, MD, PhD^{6,7}, Carine Boulon, MD⁶, Emmanuel Messas, MD, PhD^{8,9}, Victor Aboyans, MD, PhD^{5,10}, Jean Ferrières, MD, PhD^{2,11,12}, Vanina Bongard, MD, PhD^{2,11,12}

¹*Department of vascular medicine, Toulouse University Hospital, Toulouse, France*

²*UMR 1027 INSERM, Toulouse III – Paul Sabatier University, Toulouse, France*

³*UMR 1031 INSERM, StromaLab, Toulouse III – Paul Sabatier University, Toulouse, France*

⁴*Department of cardiovascular and thoracic surgery – vascular medicine, Limoges University hospital, Limoges, France*

⁵*UMR 1094 INSERM, Limoges University, Limoges, France*

⁶*Department of vascular medicine, Bordeaux, Bordeaux University Hospital, France*

⁷*Bordeaux University, Bordeaux, France*

⁸*Department of vascular medicine, Assistance Publique – Hôpitaux de Paris, Paris, France*

⁹*UMR 970 INSERM, Paris Descartes University, Paris, France*

¹⁰*Department of cardiology, Limoges University Hospital, Limoges, France*

¹¹*Department of epidemiology, Toulouse University Hospital, Toulouse, France*

¹²*Federation of cardiology, Toulouse University Hospital, Toulouse, France*

* Joint first author

Corresponding author

François-Xavier Lapébie

Department of vascular medicine, Toulouse university hospital, Toulouse, France

Hôpital Rangueil, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France

Phone number: +33 5 61 32 24 38

Fax number: +33 5 61 32 26 34

E-mail: lapebie.fx@chu-toulouse.fr; fx.lapebie@gmail.com

Abstract

Aims: The objective was to assess the association between angiotensin-converting-enzyme inhibitors (ACEI) or angiotensin receptor blockers (ARB) prescription at discharge in patients hospitalized for symptomatic lower extremity artery disease (LEAD) and one-year mortality.

Methods and results: The COPART registry is a multicentre, prospective, observational, cohort study which includes consecutive patients hospitalized for symptomatic LEAD in 4 French academic centres. All-cause mortality during a one-year follow-up after hospital discharge was compared between patients with ARB, patients with ACEI and patients without ARB or ACEI. Analyses were performed using Cox models. As a sensitivity analysis, a propensity score (PS)-matching analysis was carried out. Among 1,981 patients, 421 had ARB (21.3%), 766 ACEI (38.7%), and 794 no ACEI/ARB (40.1%) at discharge. During the one-year follow-up, incidence rates for mortality were 12.6/100 person-years (95% confidence interval, 9.7-16.1) for patients with ARB, 15.8/100 person-years (95%CI, 13.4-18.6) for patients with ACEI and 19.8/100 person-years for patients without ACEI/ARB (95%CI, 17.2-22.8). In a multivariate Cox model, ARB at discharge was associated with decreased mortality compared with no ACEI/ARB, hazard ratio (HR) 0.68 (95%CI, 0.49-0.95), and with ACEI, HR 0.69 (95%CI, 0.49-0.97). These results are consistent with those obtained by the Cox analyses in the PS-matched sample: HR 0.68 (95%CI, 0.47-0.98) for patients with ARB compared with no ARB.

Conclusion: ARB at discharge after hospitalization for symptomatic LEAD is associated with a better survival compared with ACEI or no ACEI/ARB.

Keywords: peripheral artery disease, angiotensin receptor blockers, angiotensin-converting enzyme inhibitors, cohort studies, mortality.

Introduction

In 2015, the global prevalence of lower extremity artery disease (LEAD) in individuals ages 25 years and older was 5.56% (95%CI, 3.79-8.55), with an increase in the number of cases by 17.1% compared to 2010 [138]. LEAD is associated with an increased risk of mortality, mostly cardiovascular related. In a meta-analysis, the 5-year cumulative incidence of all-cause mortality was as twice as higher than individuals without LEAD [65]. Renin-angiotensin system (RAS) activation has an important role in the risk of cardiovascular event. In the Heart Outcomes Prevention Evaluation (HOPE) study, ramipril, an angiotensin-converting-enzyme inhibitor (ACEI), reduced rates of death from cardiovascular causes, myocardial infarction (MI) and stroke by 22% as compared with placebo [139]. In this study 44% of patients had LEAD and only 47% had hypertension. Angiotensin-receptor blockers (ARB) have been proposed as alternatives to ACEI especially in case of intolerance to the latter. In 2016, the American College of Cardiology and the American Heart Association suggested ACEI or ARB for patients with LEAD to reduce the risk of cardiovascular events [67]. In 2017, the European Society of Cardiology did not suggested ACEI or ARB in all patients with LEAD but as treatment for hypertension [66]. Randomized controlled trials (RCT) did not consistently exhibit a mortality benefit for ARB compared with placebo in patients at risk of cardiovascular events but without heart failure [140]. However, no head-to-head comparison trials of ARB versus ACEI in patients without heart failure has documented a mortality advantage of ACEI [140]. Eventually, trials focusing on LEAD are lacking, even if some studies have evaluated the benefit of ACEI on walking distance [141]. Our objective was to assess the association between ACEI and ARB prescription at discharge and one-year mortality in patients hospitalized for LEAD. We hypothesized that ARB are associated with at least the same decrease in one-year mortality after hospital discharge than ACEI.

Methods

We used the COPART (COhorte de Patients ARTériopathes) registry, an ongoing multicentre, prospective, observational, cohort study, collecting prospectively data on consecutive patients hospitalized for symptomatic LEAD in the vascular medicine departments of 4 French academic centres (Bordeaux, Limoges, Paris, Toulouse) since 2006. Details on the study protocol have been published elsewhere [68]. Patient care was provided according to usual practice. After the initial hospitalization all the patients were followed for at least 12 months. Data were prospectively collected by trained medical COPART investigators, following a computerized standardized case report form, and using all records available during hospitalization and follow-up. For the assessment of vital status, patients, their families and/or their physicians were systematically contacted one year after hospital discharge. The patients gave their informed consent to participate. The study complies with the declaration of Helsinki and was approved by the Ethic Committee of Toulouse University Hospital. The current report is based on a database locked on October 2017, including patients recruited from 2006 to 2015, with at least one year of follow-up in the registry.

Each patient required the following criteria to be included in COPART registry: age ≥ 18 years, consent to participate, and specifically hospitalized for treatment and/or management of symptomatic LEAD of atherosclerotic origin. Clinical presentations included either an intermittent claudication, associated with abnormal ankle-brachial index (ABI) < 0.90 or > 1.30 , or, in the case of normal ABI at rest, a positive treadmill test (Strandness protocol), arterial stenosis $> 50\%$ on duplex ultrasound or angiography computerized tomography or angiography, or ischemic rest pain, ulceration or gangrene or an acute lower-limb ischemia related to a documented LEAD with significant arterial stenosis.

Patients for whom follow-up was improbable, those with arterial occlusive disease not related to atherosclerosis, those with acute ischemia without LEAD and those refusing to participate were not included.

The main outcome was the occurrence of all-cause mortality during the one-year following hospital discharge. The secondary outcomes were myocardial infarction (MI), stroke, a combination of 1-year mortality, MI and stroke, revascularization, amputation and leg deterioration requiring a revascularization or an amputation. Revascularization could be achieved by angioplasty (with or without stenting), subintimal recanalization, endarterectomy, embolectomy, surgical vein or prosthetic bypass. Amputation above the ankle level was

considered major, and minor if below. Details regarding baseline variables collected have already been published elsewhere [68]. Few definitions have been adapted. Dyslipidaemia was defined by a medical history, use of lipid-lowering agents, fasting low-density lipoprotein (LDL)-cholesterol ≥ 70 mg/dL or fasting triglyceridaemia ≥ 200 mg/dL. Critical limb ischemia was defined by LEAD category 4, 5 or 6 (Rutherford classification) and ankle pressure < 50 mmHg or systolic toe pressure < 30 mmHg or transcutaneous oxygen pressure < 30 mmHg. Estimated glomerular filtration rate (eGFR) was reported according to the Chronic Kidney Disease-Epidemiology collaboration (CKD-EPI) equation, and renal failure was defined when eGFR < 60 mL/min/1.73 m², which is moderate when eGFR < 60 mL/min/1.73 m² and ≥ 30 mL/min/1.73 m², and severe when eGFR < 30 mL/min/1.73 m² but without dialysis. Anaemia was diagnosed when the haemoglobin value measured at admission was less than 13 g/dL in men and less than 12 g/dL in women. Biological inflammatory syndrome was diagnosed when CRP ≥ 10 mg/L.

Categorical variables were presented as numbers and percentages. Continuous variables were presented as means and standard deviations (SD). Comparisons were made using Chi2 test for categorical variables or Fisher exact test when appropriate, while Student t test, Mann-Whitney test, one-way analysis of variance or Kruskal-Wallis test were used for continuous variables. The index date was the date of discharge from the hospitalization. The Kaplan-Meier method was used to estimate event rates. The incidence rates for complications occurring during the follow-up were estimated as the number of patients developing at least one complication for 100 person-years. Patients were censored at the date of the first complication. If they did not develop a complication during the year of follow-up, they were censored at the date of the one-year final record completion or at the date of loss of follow-up. Unadjusted and multivariate analyses were performed using Cox proportional-hazard models, after checking for the proportional-hazard assumption and the log-linearity assumption for continuous variables. For the multivariate model, prior to analysis, a directed acyclic graph (DAG) was compiled to determine the causal link between ACEI/ARB and death. Variables were chosen based on knowledge regarding these associations from the literature review, regardless the p-values in the unadjusted models. In particular, variables known as being confounding factors (factors that influence both ACEI or ARB prescription and death) were included. The following variables were included in the model: RAS inhibitors (RASi) at discharge (ACEI, ARB or none), centre, year of the index hospitalization, age, hypertension, diabetes, coronary artery disease (CAD), heart failure, eGFR, aspirin at discharge, clopidogrel at discharge, statin at discharge, beta-

blockers at discharge. HR were estimated with a 95%CI. As a sensitivity analysis, a propensity score (PS)-matching analysis was carried out. PS was defined as the probability of treatment assignment conditionally to the measured baseline variables. The goal of the PS matching was to find pairs of patients receiving ARB or no ARB but sharing the same predicted probabilities of receiving ARB. For the estimation of PS, we used a logistic regression model in which treatment status (ARB or no ARB) was regressed on the following baseline characteristics: centre, year of the index hospitalization, age, hypertension, diabetes, CAD, heart failure, renal failure, LEAD clinical presentation (Rutherford classification, acute limb ischemia), leg revascularization (angioplasty, bypass), limb amputation (major, minor), iloprost, treatment at discharge (aspirin, clopidogrel, oral anticoagulant, statin, beta-blockers). Balancing property of the PS was tested. The calibration of the PS was checked by the Hosmer and Lemeshow test and its discrimination was assessed by the receiver operating characteristic (ROC) curve and the area under the curve (AUC). We then performed a 1-to-1 matching without replacement by nearest neighbour method without calliper width. We included in an unadjusted Cox model all matched pairs. A p-value <0.05 was considered to be statistically significant. Analyses were performed on STATA Statistical software (release 14.2, StataCorp LLC®).

Results

At the latest data export, the COPART registry comprised 2,520 patients. Among them, 344 were excluded from analyses for the following reasons: unknown inclusion date ($n = 1$), asymptomatic LEAD ($n = 6$), death during index hospitalization ($n = 109$), a combined prescription of ACEI and ARB at discharge ($n = 15$), lack of data about RASI prescription at discharge ($n = 195$), or no follow-up (<1 day) after hospital discharge ($n = 213$). We kept 1,981 patients for analyses: 766 (38.7%) with ACEI at hospital discharge (ACEI group), 421 (21.3%) with ARB (ARB group) and 794 (40.1%) without RASI (no-ACEI/ARB group). The mean duration of the follow-up was 319 ± 98 days for the whole cohort, with a mean of 322 ± 93 for ACEI group, 332 ± 85 days for ARB group and 309 ± 107 for no-ACEI/ARB group ($p = 0.012$ for the comparison between the 3 groups). After hospital discharge, among these 1,981 patients, 326 died within 365 days of follow-up, 1,108 were alive at 365 days and have completed the follow-up, and 547 were lost to follow-up before 365 days with a mean duration of follow-up reaching 336 ± 67 days.

At admission, cardiovascular risk factors and cardiovascular diseases were more frequent among ACEI and ARB groups than among no-ACEI/ARB group (Table 1). LEAD at hospital admission was often severe, with Rutherford category 5 or 6 being more frequent among ACEI and no-ACEI/ARB groups than among ARB group. Renal failure was more frequent among ARB group than among ACEI and no-ACEI/ARB groups.

During hospitalization, major amputation was more frequent among no-ACEI/ARB group than among ACEI and ARB groups (Table 2). At hospital discharge, ACEI and ARB groups were more often on antiplatelet therapy, statin and beta-blockers than no-ACEI/ARB group (Table 3).

In the unadjusted Cox model, ACEI and ARB were associated with a reduce in all-cause mortality during the one-year follow-up after hospital discharge, as compared with no-ACE/ARB (Figure 1, Table 4, Table 5). There was a trend towards a decrease in mortality with ARB compared with ACEI, but this was not statistically significant. Other factors associated with all-cause mortality in the unadjusted Cox model are exposed in supplementary material table 6. In the multivariate Cox model, ARB was associated with a decreased one-year risk of all-cause mortality compared with no-ACEI/ARB, HR 0.68 (95%CI, 0.49-0.95), but also compared with ACEI, HR 0.69 (95%CI, 0.49-0.97), and ACEI was no longer associated with lower mortality (Figure 2, and supplementary material table 7). In a first sensitivity analysis,

interventions during hospitalization (revascularization, amputation) were included in a second multivariate Cox model with the variables selected for the first multivariate Cox model, but HR were not modified (data not shown). In a second sensitivity analysis, 1,433 patients with data on RASI prescription both at hospital discharge and at the end of the follow-up were analyzed. There was no significant difference for treatment modification between ACEI, ARB and no-ACEI/ARB groups (25.8%, 29.7% and 22.5% respectively, $p=0.064$, supplementary material table 8). The multivariate Cox analysis was then reran on the 1,070 patients who were under the same RASI treatment (ACEI, ARB, no-ACEI/ARB) at hospital discharge and at the end of the follow-up. Again, ARB was associated with a decreased one-year risk of all-cause mortality compared with no-ACEI/ARB, HR 0.23 (95%CI, 0.08-0.66), but also compared with ACEI, HR 0.34 (95%CI, 0.11-0.98), while ACEI was no longer associated with lower mortality, HR 0.69 (95%CI, 0.39-1.21) compared with no-ACEI/ARB.

These results are consistent with those obtained by the Cox regression analyses in the PS-matched sample: 302 patients treated by ARB at hospital discharge and without missing data were matched with 302 patients not treated by ARB at hospital discharge (153 treated by ACEI, 149 without RASI, supplementary material table 9). ARB was associated with a reduced mortality during the one-year follow-up, HR 0.68 (95%CI, 0.47-0.98), $p = 0.039$, by comparison with no ARB (Figure 2).

Discussion

We compared the prognosis impact of ARB and ACEI after hospitalization for symptomatic LEAD using COPART, a multicentre prospective registry in France. The results showed that treatment with ARB was associated with a better 1-year survival compared with patients without RASI or with ACEI.

A previous study also documented a mortality benefit of ARB by comparison of ACEI in patients with or at risk of cardiovascular disease. In a propensity score matching analysis using the Reduction of Atherothrombosis for Continued Health (REACH) registry, a multicentre prospective observational registry of 69,055 outpatients with high cardiovascular risk (12% with LEAD) and a 4-years follow-up, ARB use was associated with a lower rate of all-cause mortality compared with ACEI, HR 0.84 (95%CI, 0.75-0.94) [142,143]. In a subgroup analysis, ARB was associated with lower rates of all-cause mortality in secondary prevention but not in primary prevention patients.

However, most of the RCTs comparing ACEI vs placebo in patients without heart failure were positive, whereas most of the ARB vs placebo RCT were negative. In a meta-analysis of 20 RCT comparing ARB with placebo in patients at high cardiovascular risk but without heart failure, ARB was not associated with a decrease in all-cause mortality, with a relative risk of 1.01 (95%CI, 0.96-1.06) [140]. This may be explained by a higher event rate in the placebo group in the ACEI trials than in the ARB trials. Moreover, in most ACEI trials the majority of patients were naïve of RASI before randomization, and after randomization other RASI were rarely used, neither in the active arm nor in the placebo one. Conversely, in most ARB trials the majority of patients were receiving RASI before randomization, and after randomization many patients received non-study RASI, either in the active arm or in the placebo arm [140,144]. Thus, a meta-analysis of 7 head-to-head RCT of ARB vs ACEI with 22,422 participants without heart failure revealed no mortality difference [140]. Rates of outcomes were also similar between ACEI and ARB in RCTs or meta-analysis of RCTs in hypertension, diabetes, chronic kidney disease, cerebrovascular disease, CAD and heart failure [145–148].

Several reasons could explain the difference between ARB and ACEI in our study. First, patients with LEAD have greater long-term risk of death than patients with other atherosclerosis related diseases [149]. To our knowledge, our study is the first to compare ARB and ACEI in LEAD. Because most of ACEI and ARB studies were conducted in lower cardiovascular risk populations, the difference between ARB and ACEI could have been underestimated. Second,

a significant proportion of patients do not tolerate ACEI because of cough or the occurrence of angioedema; as a result, there is less drug withdrawal with ARB than with ACEI [140]. However, in our study there did not seem to be significant difference for treatment modification between ACEI, ARB and no-ACEI/ARB groups. Third, even if most of data on ACEI and ARB doses were not systematically collected in the COPART registry, the proportion of patients with the highest treatment dose could be different between ACEI and ARB groups. Indeed, dose data at the end of the follow-up were available for 163 patients: the highest dose of ACEI was prescribed for 20 patients among 102 treated by ACEI (19.6%), and the highest dose of ARB was prescribed for 31 patients among 61 treated by ARB (50.8%), $p < 0.001$ for the comparison between the two groups. Because most of the RCTs assessing ACEI and ARB in patients without heart failure have tested doses at the upper limit of the antihypertensive dose ranges, it is not certain that lower doses provide identical benefits. For example, in the Ongoing Telmisartan Alone and in Combination with Ramipril Global Endpoint Trial (ONTARGET) study, ramipril was given at 10 mg per day and telmisartan at 80 mg per day [150]. Finally, effect of ACEI and ARB on improvement of endothelial dysfunction could be different. Angiotensin II contributes to endothelial dysfunction and promotes both hypertrophy and intimal hyperplasia. It exerts its primary effects by binding to the AT1 subtype of the angiotensin II receptor (AT1R), expressed at increased levels in the intima in response to vascular injury. Enhancement of endothelial function has been related to improved cardiovascular prognosis. While ACEI causes a subtotal inhibition of angiotensin II production, ARB provides a complete blockade of the AT1R [151]. However RCT with a head-to-head comparison of ARB and ACEI on endothelial function are lacking, and in one of the only studies, there was no statistically significant difference between the effects of telmisartan and ramipril on improving renal endothelial dysfunction. Yet, telmisartan produced a significant increase in renal plasma flow and a decrease in renal vascular resistance compared with baseline, whereas ramipril had no effect on renal hemodynamics [152].

Our analysis has some limitations. First, most of the patients with ACEI or ARB at hospital discharge used these treatments before hospitalization. Because the baseline characteristics of prevalent users may be affected by the drug itself, a bias may have been introduced. Second, while it is known that the beneficial effect of RASI is obtained with higher doses, we could not record most of the data regarding doses of ACEI and ARB used, thus we could not take into account differences between optimal doses within the same drug class. Third, some of the data were lacking regarding drug withdrawal during the one-year follow-up.

Fourth, because of the observational nature of the study, and despite the use of propensity score, we cannot exclude the possibility of non-controlled variables differing in the 3 groups with could affect the outcome. Finally, because COPART only includes inpatients with LEAD, the cardiovascular risk of the cohort could be greater than the risk of non-hospitalized patients and we cannot generalize the results to all of the LEAD patients.

Despite these limitations, ARB might be considered over ACEI in LEAD patients to reduce mortality.

Funding

None.

Acknowledgements

We thank the research-assistants team: Delphine Baudoin, Clémence Labetoulle, Emmanuelle Lorian, Marion Mangin, Béatrice Montalègre, Monique Pailhol, Brigitte Pelvet, Irina Smoliy.

Conflicts of interest

Dr Lapébie, Dr Lacroix, Dr Constans, Dr Boulon and Dr Messas have nothing to disclose. Dr Bura-Rivière reports grants from Ministère de la recherche, grants, personal fees, non-financial support and other from Bayer Healthcare and BMS/Pfizer, grants and non-financial support from Astra-Zeneca, outside the submitted work. Dr Aboyans reports personal fees from Bayer Healthcare, BMS/Pfizer, Novartis, and Sanofi, outside the submitted work. Dr Ferrières reports personal fees from Amgen, Sanofi, MSD, and Servier, outside the submitted work. Dr Bongard reports personal fees from Sanofi outside the submitted work.

Tables

Table 1: Characteristics at hospital admission of patients with or without ACEI or ARB at hospital discharge

	Patients with ACEI (n = 766 [%])	Patients with ARB (n = 421 [%])	Patients without ACEI/ARB (n= 794 [%])	p
Demographic data				
Men	591 /764 (77.4)	308 /418 (73.7)	598 /791 (75.6)	0.360
Age in years, mean ±SD	69.6 ±12.4	71.1 ±11.5	70.6 ±13.4	0.065
Cardiovascular risks factors				
Diabetes mellitus	376 /766 (49.1)	238 /421 (56.5)	367 /793 (46.3)	0.003
Dyslipidaemia	732 /766 (95.6)	404 /421 (96.0)	747 /794 (94.1)	0.252
Hypertension	683 /766 (89.2)	414 /421 (98.3)	625 /794 (78.7)	<0.001
Current smokers	213 /758 (28.1)	89 /419 (21.2)	203 /789 (25.7)	0.036
Obesity	157 /715 (22.0)	110 /395 (27.9)	127 /742 (17.1)	<0.001
Family history of CAD	99 /653 (15.2)	57 /376 (15.2)	65 /684 (9.5)	0.003
Medical history				
Coronary artery disease	351 /761 (46.1)	146 /420 (34.8)	264 /791 (33.4)	<0.001
Cerebrovascular disease	106 /758 (14.0)	69 /417 (16.6)	114 /780 (14.6)	0.489
Heart failure	90 /748 (12.0)	33 /416 (7.9)	101 /783 (12.9)	0.032
Atrial fibrillation	145 /747 (19.4)	84 /416 (20.2)	172 /784 (21.9)	0.461
COPD or asthma	124 /744 (16.7)	80 /405 (19.8)	152 /779 (19.5)	0.271
LEAD at admittance				
Newly diagnosed LEAD	235 /764 (30.8)	149 /419 (35.6)	265 /792 (33.5)	0.218
Prior revascularization	287 /761 (37.7)	122 /418 (29.2)	251 /786 (31.9)	0.006
Prior amputation	98 /757 (13.0)	47 /419 (11.2)	124 /788 (15.7)	0.070
LEAD classification				0.003
Rutherford 1,2 or 3	234 /766 (30.6)	140 /421 (33.3)	197 /794 (24.8)	/
Rutherford 4	78 /766 (10.2)	56 /421 (13.3)	76 /794 (9.6)	/
Rutherford 5 or 6	398 /766 (52.0)	194 /421 (46.1)	448 /794 (56.4)	/
Acute limb ischemia	56 /766 (7.3)	31 /421 (7.4)	73 /794 (9.2)	/
Critical limb ischemia	297 /706 (42.1)	143 /386 (37.1)	288 /709 (40.6)	0.268
Gangrene	86 /742 (11.6)	41 /412 (10.0)	89 /773 (11.5)	0.659

Laboratory exams at admittance				
eGFR, mean \pm SD	70.8 \pm 28.2	63.4 \pm 28.1	64.8 \pm 34.0	<0.001
Renal failure				
None	491 /756 (65.0)	228 /419 (54.4)	450 /779 (57.8)	<0.001
Moderate	205 /756 (27.1)	141 /419 (33.7)	194 /779 (24.9)	/
Severe	32 /756 (4.2)	31 /419 (7.4)	60 /779 (7.7)	/
Dialysis	28 /756 (3.7)	19 /419 (4.5)	75 /779 (9.6)	/
Anaemia	336 /729 (46.1)	212 /410 (51.7)	415 /740 (56.1)	0.001
Inflammatory syndrome	359 /692 (51.9)	185 /384 (48.2)	431 /694 (62.1)	<0.001

Table 2: Intervention during index hospitalization

	Patients with ACEI (n = 766 [%])	Patients with ARB (n = 421 [%])	Patients without ACEI/ARB (n= 794 [%])	p
Revascularization	381 /763 (49.9)	205 /421 (48.7)	392 /794 (49.4)	0.919
Angioplasty	270 /763 (35.4)	153 /421 (36.3)	226 /793 (28.5)	0.003
Embolectomy	28 /760 (3.7)	18 /420 (4.3)	40 /790 (5.1)	0.412
Endarterectomy	47 /758 (6.2)	20 /417 (4.8)	60 /791 (7.6)	0.161
Bypass	81 /763 (10.6)	37 /421 (8.8)	130 /794 (16.4)	<0.001
Amputation	122 /766 (15.9)	63 /421 (15.0)	162 /792 (20.5)	0.019
Major amputation	44 /766 (5.7)	24 /421 (5.7)	82 /792 (10.4)	0.001
Minor amputation	83 /766 (10.8)	42 /421 (10.0)	91 /792 (11.5)	0.720

Table 3: Comparison of cardiovascular treatment at discharge of index hospitalization between ACEI group, ARB group and no-ACEI/ARB group

	Patients with ACEI (n = 766 [%])	Patients with ARB (n = 421 [%])	Patients without ACEI/ARB (n= 794 [%])	p
Aspirin	561 /766 (73.2)	277 /421 (65.8)	508 /794 (64.0)	<0.001
Clopidogrel	179 /766 (23.4)	117 /421 (27.8)	163 /794 (20.5)	0.017
Aspirin and/or clopidogrel	644 /766 (84.1)	353 /421 (83.9)	614 /794 (77.3)	0.001
Oral anticoagulant	118 /763 (15.5)	71 /421 (16.9)	125 /793 (15.8)	0.814
Aspirin or clopidogrel or oral anticoagulant	705 /765 (92.2)	394 /421 (93.6)	702 /794 (88.4)	0.004
Statin	637 /766 (83.2)	328 /421 (77.9)	535 /793 (67.5)	<0.001
Beta-blockers	266 /766 (34.7)	144 /421 (34.2)	227 /793 (28.6)	0.022

Table 4: Incidence rates of overall death, myocardial infarction, stroke, leg revascularization and leg amputation during the one-year follow-up (per 100 person-years)

	Patients with ACEI (% [95%CI])	Patients with ARB (% [95%CI])	Patients without ACEI/ARB (% [95%CI])
Overall mortality	15.8 (13.4-18.6)	12.6 (9.7-16.1)	19.8 (17.2-22.8)
Mortality, myocardial infarction or stroke	18.7 (16.1-21.7)	16.6 (13.4-20.6)	22.6 (19.9-25.8)
Myocardial infarction	3.2 (2.1-4.8)	3.5 (2.1-5.9)	2.9 (1.9-4.4)
Stroke	2.2 (1.3-3.7)	2.8 (1.6-5.1)	1.7 (1.0-3.0)
Poor leg outcome	36.7 (33.2-40.4)	31.4 (27.1-36.1)	37.1 (33.6-40.7)
Revascularization	26.0 (22.9-29.4)	21.0 (17.3-25.3)	24.5 (21.5-27.8)
Amputation	18.2 (15.6-21.3)	17.0 (13.7-21.1)	19.6 (16.9-22.7)
Major amputation	9.6 (7.6-12.0)	7.5 (5.3-10.5)	10.3 (8.3-12.8)

Table 5: Comparison of outcomes between patients with ACEI at discharge, patients with ARB at discharge and patients without ACEI or ARB at discharge by unadjusted Cox models

	ACEI Vs no ACEI/ARB HR (95%CI)	ARB Vs no ACEI/ARB HR (95%CI)	ARB Vs ACEI HR (95%CI)
Overall mortality	0.77 (0.60-0.97)	0.60 (0.44-0.82)	0.78 (0.56-1.08)
Mortality, myocardial infarction or stroke	0.80 (0.64-0.99)	0.70 (0.53-0.92)	0.88 (0.66-1.17)
Myocardial infarction	1.10 (0.60-2.02)	1.25 (0.63-2.48)	1.14 (0.58-2.22)
Stroke	1.24 (0.58-2.66)	1.62 (0.71-3.67)	1.30 (0.60-2.83)
Poor leg outcome	0.98 (0.82-1.16)	0.82 (0.66-1.01)	0.83 (0.68-1.03)
Revascularization	1.08 (0.88-1.32)	0.86 (0.66-1.11)	0.80 (0.62-1.03)
Amputation	0.91 (0.72-1.16)	0.85 (0.63-1.13)	0.93 (0.69-1.25)
Major amputation	0.92 (0.66-1.28)	0.72 (0.47-1.10)	0.78 (0.51-1.20)

Table 6: Factors associated with all-cause mortality during the one-year follow-up (unadjusted Cox models)

	Hazard ratio	95%CI	p
Demographic data			
Men	0.80	0.63-1.01	0.062
Age (for an increase of 1 year)	1.06	1.05-1.07	<0.001
Cardiovascular risks factors			
Diabetes mellitus	1.54	1.24-1.91	<0.001
Dyslipidaemia	0.60	0.40-0.89	0.012
Hypertension	1.10	0.80-1.51	0.547
Current smokers	0.34	0.24-0.48	<0.001
Obesity	0.67	0.49-0.92	0.014
Family history of CAD	0.61	0.41-0.93	0.020
Medical history			
Coronary artery disease	1.12	0.90-1.39	0.308
Cerebrovascular disease	1.37	1.04-1.81	0.025
Heart failure	3.33	2.61-4.23	<0.001
Atrial fibrillation	2.29	1.83-2.87	<0.001
COPD or asthma	1.04	0.79-1.37	0.776
LEAD at admittance			
Newly diagnosed LEAD	1.01	0.81-1.27	0.908
Prior revascularization	0.86	0.68-1.08	0.194
Prior amputation	1.58	1.20-2.06	0.001
LEAD classification			
Rutherford 1,2 or 3	Ref.	Ref.	Ref.
Rutherford 4	2.44	1.41-4.22	0.001
Rutherford 5 or 6	5.92*	2.30-6.45	<0.001
Acute limb ischemia	3.86*	2.30-6.45	<0.001
Critical limb ischemia	1.98	1.58-2.50	<0.001
Gangrene	1.87	1.42-2.46	<0.001
Laboratory exams at admittance			
eGFR (for an increase of 1 mL/min/1.73 m ²)	0.98	0.98-0.99	<0.001
Renal failure			
None	Ref.	Ref.	Ref.
Moderate	2.41	1.90-3.06	<0.001
Severe	3.55	2.52-4.99	<0.001
Dialysis	2.37	1.58-3.54	<0.001
Anaemia	2.10	1.67-2.66	<0.001
Inflammatory syndrome	2.85	2.24-3.61	<0.001
Intervention during index hospitalization			
Revascularization	0.67	0.54-0.83	<0.001

Angioplasty	0.69*	0.54-0.87	0.002
Embolectomy	0.85	0.48-1.51	0.576
Endarterectomy	0.93	0.58-1.47	0.749
Bypass	0.87	0.62-1.23	0.433
Amputation	1.86	1.46-2.37	<0.001
Major amputation	1.99	1.43-2.75	<0.001
Minor amputation	1.68	1.25-2.24	<0.001
Iloprost infusion	1.37	0.93-2.02	0.110
Cardiovascular treatment at discharge			
Aspirin or clopidogrel or oral anticoagulant	0.43*	0.32-0.58	<0.001
Aspirin and/or clopidogrel	0.51	0.40-0.65	<0.001
Aspirin	0.74	0.59-0.92	0.007
Clopidogrel	0.50	0.36-0.68	<0.001
Oral anticoagulant	1.19	0.91-1.57	0.202
Statin	0.48*	0.38-0.60	<0.001
Beta-blocker	0.86	0.68-1.09	0.214
ACEI or ARB	0.70	0.56-0.87	0.001
ACEI	0.89	0.71-1.11	0.296
ARB	0.67	0.50-0.91	0.009

*Proportional hazard assumption not respected

Table 7: Association between ACEI, ARB, no ACEI or ARB and all-cause death during the one-year follow-up in a multivariate Cox regression analysis

	Hazard ratio	95%CI	p
RASI treatment at discharge			
ACEI versus no ACEI/ARB	1.00	0.77-1.29	0.983
ARB versus no ACEI/ARB	0.68	0.49-0.95	0.023
ARB versus ACEI	0.69	0.49-0.97	0.031
Centre			
Toulouse	Ref.	Ref.	Ref.
Limoges	1.36	1.02-1.80	0.033
Bordeaux	1.40	0.92-2.12	0.112
Paris*	/	/	/
Year of the index hospitalization			
2006	Ref.	Ref.	Ref.
2007	2.71	0.84-8.68	0.094
2008	2.45	0.76-7.89	0.133
2009	2.03	0.62-6.65	0.244
2010	2.78	0.84-9.16	0.093
2011	1.59	0.46-5.49	0.464
2012	2.35	0.72-7.70	0.158
2013	2.12	0.65-6.91	0.213
2014	1.62	0.49-5.35	0.427
2015	0.65	0.13-3.23	0.594
Age (for an increase of 1 year)	1.05	1.03-1.06	<0.001
Hypertension	0.72	0.51-1.02	0.065
Diabetes	1.25	1.00-1.58	0.054
Coronary artery disease	1.06	0.83-1.36	0.637
Heart failure	2.07	1.58-2.71	<0.001
eGFR (for an increase of 1 mL/min/1.73 m ²)	0.99	0.99-1.00	<0.001
Aspirin at discharge	0.91	0.71-1.16	0.423
Clopidogrel at discharge	0.62	0.43-0.88	0.008
Statin at discharge	0.76	0.60-0.97	0.029
Beta-blocker at discharge	0.81	0.62-1.05	0.112

*Centre of Paris omitted because of collinearity

Table 8: RASI treatment at the end of the follow-up

	Patients with ACEI at discharge (n = 574 [%])	Patients with ARB at discharge (n = 303 [%])	Patients without ACEI/ARB at discharge (n= 556 [%])	p
RASI treatment at the end of the follow-up				
ACEI	426 /574 (74.2)	21 /303 (6.9)	80 /556 (14.4)	/
ARB	18 /574 (3.1)	213 /303 (70.3)	42 /556 (7.6)	/
ACEI and ARB	2 /574 (0.4)	7 /303 (2.3)	3 /556 (0.5)	/
No ACEI/ARB	128 /574 (22.3)	62 /303 (20.5)	431 /556 (77.5)	/
Treatment modification at the end of the follow-up	148 /574 (25.8)	90 /303 (29.7)	125 /556 (22.5)	0.064

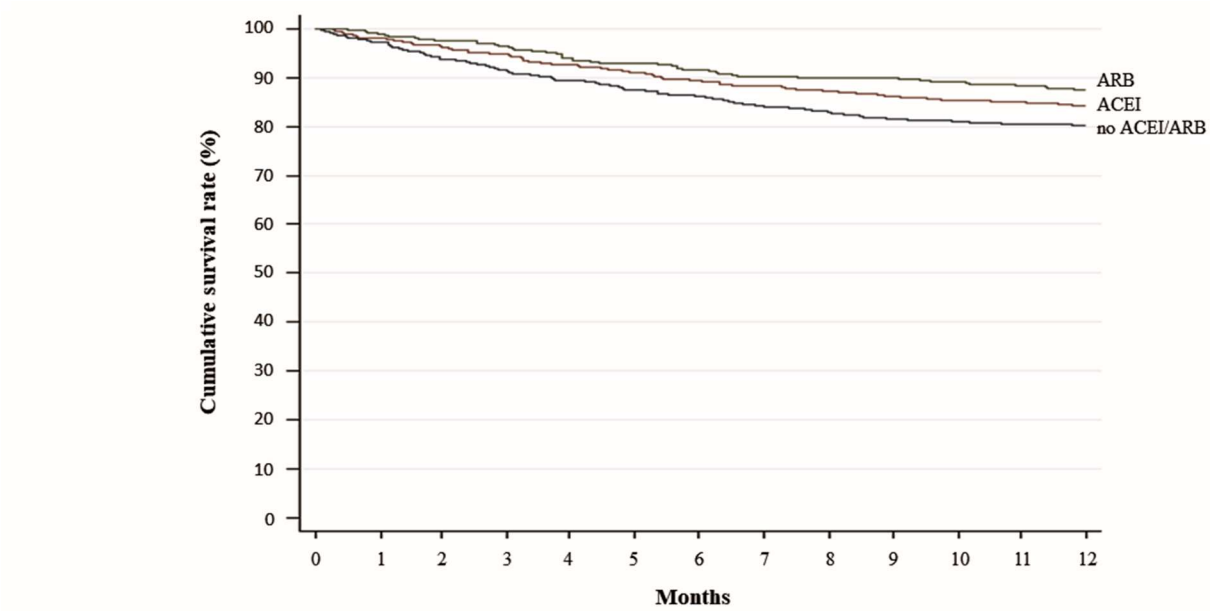
Table 9: Baseline characteristics of patients in the propensity matched cohorts

	All patients (n = 604 [%])	Patients with ARB (n = 302 [%])	Patients without ARB (n= 302 [%])	p
Demographic data				
Men	446 /603 (74.0)	216 /301 (71.8)	230 /302 (76.2)	0.218
Age in years, mean ±SD	71.6 ±11.8	71.6 ±11.8	71.7 ±11.9	0.987
Cardiovascular risks factors				
Diabetes mellitus	350 /604 (58.0)	179 /302 (59.3)	171 /302 (56.6)	0.510
Dyslipidaemia	584 /604 (96.7)	292 /302 (96.7)	292 /302 (96.7)	1.000
Hypertension	594 /604 (98.3)	298 /302 (98.7)	296 /302 (98.0)	0.524
Current smokers	137 /600 (22.8)	65 /300 (21.7)	72 /300 (24.0)	0.496
Obesity	140 /562 (24.9)	74 /281 (26.3)	66 /281 (23.5)	0.435
Family history of CAD	70 /515 (13.6)	39 /267 (14.6)	31 /248 (12.5)	0.486
Medical history				
Coronary artery disease	201 /604 (33.3)	106 /302 (35.1)	95 /302 (31.5)	0.342
Cerebrovascular disease	88 /599 (14.7)	51 /299 (17.1)	37 /300 (12.3)	0.103
Heart failure	56 /604 (9.3)	29 /302 (9.6)	27 /302 (8.9)	0.779
Atrial fibrillation	132 /602 (21.9)	60 /300 (20.0)	72 /302 (23.8)	0.255
COPD or asthma	115 /586 (19.6)	56 /289 (19.4)	59 /297 (19.9)	0.882
LEAD at admittance				
Newly diagnosed LEAD	400 /603 (66.3)	192 /301 (63.8)	208 /302 (68.9)	0.186
Prior revascularization	194 /598 (32.4)	84 /299 (28.1)	110 /299 (36.8)	0.023
Prior amputation	73 /601 (12.2)	32 /301 (10.6)	41 /300 (13.7)	0.255
LEAD classification				0.579
Rutherford 1,2 or 3	192 /604 (31.8)	102 /302 (33.8)	90 /302 (29.8)	/
Rutherford 4	86 /604 (14.2)	38 /302 (12.6)	48 /302 (15.9)	/
Rutherford 5 or 6	279 /604 (46.2)	138 /302 (45.7)	141 /302 (46.7)	/
Acute limb ischemia	47 /604 (7.8)	24 /302 (8.0)	23 /302 (7.6)	/
Critical limb ischemia	207 /546 (37.9)	100 /272 (36.8)	107 /274 (39.1)	0.582
Gangrene	55 /588 (9.4)	27 /297 (9.1)	28 /291 (9.6)	0.825
Laboratory exams at admittance				
eGFR, mean ±SD	62.7 ±29.9	61.4 ±28.4	64.0 ±31.4	0.286
Renal failure				0.093

None	320 /604 (53.0)	156 /302 (51.7)	164 /302 (54.3)	/
Moderate	198 /604 (32.8)	104 /302 (34.4)	94 /302 (31.1)	/
Severe	44 /604 (7.3)	27 /302 (8.9)	17 /302 (5.6)	/
Dialysis	42 /604 (7.0)	15 /302 (5.0)	27 /302 (8.9)	/
Anaemia	292 /573 (51.0)	152 /292 (52.1)	140 /281 (49.8)	0.593
Inflammatory syndrome	221 /543 (40.7)	105 /271 (38.8)	116 /272 (42.7)	0.355
Intervention during index hospitalization				
Revascularization	288 /604 (47.7)	146 /302 (48.3)	142 /302 (47.0)	0.745
Angioplasty	202 /604 (33.4)	101 /302 (33.4)	101 /302 (33.4)	1.000
Embolectomy	30 /601 (5.0)	15 /301 (5.0)	15 /300 (5.0)	0.993
Endarterectomy	38 /601 (6.3)	16 /301 (5.3)	22 /300 (7.3)	0.310
Bypass	59 /604 (9.8)	29 /302 (9.6)	30 /302 (9.9)	0.891
Amputation	87 /604 (14.4)	47 /302 (15.6)	40 /302 (13.3)	0.417
Major amputation	34 /604 (5.6)	18 /302 (6.0)	16 /302 (5.3)	0.724
Minor amputation	59 /604 (9.8)	32 /302 (10.6)	27 /302 (8.9)	0.493
Iloprost infusion	38 /604 (6.3)	18 /302 (6.0)	20 /302 (6.6)	0.738
Cardiovascular treatment at discharge				
Aspirin or clopidogrel or oral anticoagulant	551 /604 (91.2)	279 /302 (92.4)	272 /302 (90.1)	0.314
Aspirin and/or clopidogrel	485 /604 (80.3)	247 /302 (81.8)	238 /302 (78.8)	0.357
Aspirin	391 /604 (64.7)	201 /302 (66.6)	190 /302 (62.9)	0.349
Clopidogrel	155 /604 (25.7)	77 /302 (25.5)	78 /302 (25.8)	0.926
Oral anticoagulant	100 /604 (16.6)	50 /302 (16.6)	50 /302 (16.6)	1.000
Statin	461 /604 (76.3)	234 /302 (77.5)	227 /302 (75.2)	0.503
Beta-blocker	203 /604 (33.6)	108 /302 (35.8)	95 /302 (31.5)	0.263
ACEI	153 /604 (25.3)	0 /302 (0.0)	153 /302 (50.7)	<0.001

Figures

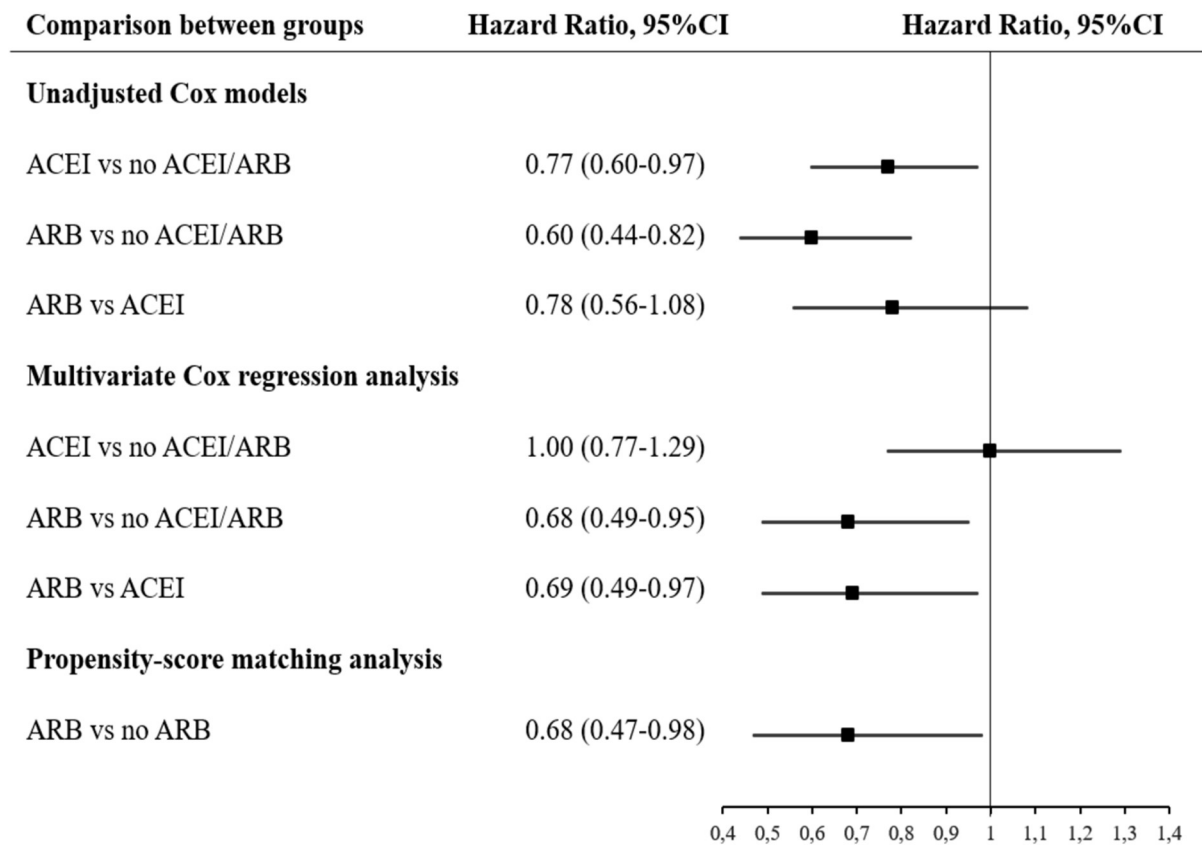
Figure 1: Kaplan-Meier cumulative survival rate between patients with ACEI at discharge, patients with ARB at discharge and patients without ACEI or ARB at discharge.



No. at Risk

ACEI	766	718	676	648	390
ARB	421	401	381	374	255
no ACEI/ARB	794	715	670	633	463

Figure 2: Comparison of all-cause mortality between patients with ACEI at discharge, patients with ARB at discharge and patients without ACEI or ARB at discharge by unadjusted Cox models, multivariate Cox regression analysis and propensity-score matching analysis. For the multivariate Cox regression analysis, only one model was made with RASI treatment at discharge included as a 3-class categorical variable (“ACEI”, “ARB”, and “no ACEI/ARB”) and HR were calculated with “no ACEI/ARB” then “ACEI” as reference class.



3.3.2.2. Résumé et remarques

Dans ce travail, toujours dans la cohorte COPART, on s'intéressait à l'association entre la prescription d'un IEC ou d'un ARA2 à la sortie d'hospitalisation pour prise en charge d'une AAMI symptomatique, et la mortalité toute cause dans la première année du suivi. Sur les 1981 patients inclus dans l'analyse, 421 (21,3%) bénéficiaient d'un ARA2, 766 (38,7%) d'un IEC, et 794 (40,1%) d'aucun de ces traitements. Les incidences cumulées à 1 an des décès dans ces 3 groupes étaient respectivement de 12,6% (IC95%, 9,7-16,1), 15,8% (IC95%, 13,4-18,6) et 19,8% (IC95%, 17,2-22,8).

Un modèle multivarié à but explicatif a été développé afin d'estimer le lien de causalité entre la prescription d'un inhibiteur du SRAA et le décès, avec une sélection des variables opérée par un DAG. Ce modèle a mis en évidence une association entre la prescription d'un ARA2 et une réduction de la mortalité par comparaison avec l'absence d'inhibiteur du SRAA, avec un HR de 0,68 (IC95%, 0,49-0,95), et par comparaison avec un IEC, avec un HR de 0,69 (IC95%, 0,49-0,97).

A noter qu'un modèle de Cox non ajusté a également été réalisé dans un échantillon de patients appariés selon un score de propension, comme analyse de sensibilité, confirmant ces résultats avec un HR de 0,68 (IC95%, 0,47-0,98) pour les patients traités par ARA2 par comparaison aux patients non traités par ARA2. Le score de propension est défini comme la probabilité pour un patient de recevoir une intervention expérimentale (traitement ou autre) conditionnellement à ses caractéristiques initiales [153]. Dans ce travail, il a donc été utilisé pour définir la probabilité pour un patient de recevoir un ARA2. Un appariement a ensuite été réalisé, de type 1 pour 1, entre un patient ayant effectivement reçu un ARA2 et un patient n'ayant pas reçu un ARA2, ces deux patients ayant pourtant des scores de propension comparables. Cette méthode permet d'améliorer la comparabilité des groupes dans un étude observationnelle, où l'allocation du traitement n'a pas été réalisée par randomisation, en prenant en compte plusieurs facteurs de confusion potentiels. Cependant, elle ne permet pas de prendre en compte des facteurs de confusion non mesurés [153].

Cet article a été publié dans *European Heart Journal – Cardiovascular pharmacotherapy* (en 2020, IF 6,617, rang A pour « Cardiac & cardiovascular systems » et « Pharmacology & pharmacy ») [110].

3.4. Validation externe des résultats d’essais contrôlés randomisés dans une cohorte : article “External applicability of the COMPASS and VOYAGER-PAD trials on patients with symptomatic lower extremity artery disease in France: the COPART registry”

3.4.1. Manuscrit

External applicability of the COMPASS and VOYAGER-PAD trials on patients with symptomatic lower extremity artery disease in France: the COPART registry

François-Xavier Lapébie, MD^{1,2}, Victor Aboyans, MD, PhD^{3,4}, Philippe Lacroix, MD^{4,5}, Joël Constans, MD, PhD^{6,7}, Carine Boulon, MD⁶, Emmanuel Messas, MD, PhD^{8,9}, Jean Ferrières, MD, PhD^{2,10,11}, Vanina Bongard, MD, PhD^{2,10,11}, Alessandra Bura-Rivière, MD, PhD^{1,12}

¹*Department of vascular medicine, Toulouse University Hospital, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France*

²*UMR 1295 INSERM, CERPOP, Toulouse III – Paul Sabatier University, 37 allée Jules Guesde, 31073 Toulouse Cedex, France*

³*Department of cardiology, Limoges University Hospital, 2 avenue Martin Luther King, 87042 Limoges Cedex, France*

⁴*UMR 1094 INSERM & IRD, Limoges University, 2 rue du Dr Marcland, 87025 Limoges Cedex, France*

⁵*Department of cardiovascular and thoracic surgery – vascular medicine, Limoges University hospital, 2 avenue Martin Luther King, 87042 Limoges Cedex, France*

⁶*Department of vascular medicine, Bordeaux University Hospital, 1 rue Jean Burguet, 33076 Bordeaux Cedex, France*

⁷*Bordeaux University, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France*

⁸*Department of vascular medicine, Assistance Publique – Hôpitaux de Paris, 20 rue Leblanc, 75908 Paris Cedex 15, France*

⁹*UMR 970 INSERM, Paris Descartes University, 56 rue Leblanc, 75015 Paris, France*

¹⁰*Department of epidemiology, Toulouse University Hospital, 37 allée Jules Guesde, 31073 Toulouse Cedex, France*

¹¹*Federation of cardiology, Toulouse University Hospital, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France*

¹²UMR 1031 INSERM, StromaLab, Toulouse III – Paul Sabatier University, 4 bis avenue Hubert Curien, 31100 Toulouse, France

Short title

Applicability of COMPASS and VOYAGER-PAD trials

Category of the manuscript

Original article

Word count

Abstract: 290

Main text: 3 404

References: 24

Tables/figures: 4/5 (color should be used for figures)

Supplementary material: 8 tables

Corresponding author and post-publication corresponding author

François-Xavier Lapébie

Department of vascular medicine, Toulouse university hospital, Toulouse, France

Hôpital Rangueil, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France

Phone number: +33 5 61 32 24 38

Fax number: +33 5 61 32 26 34

E-mail: lapebie.fx@chu-toulouse.fr; fx.lapebie@gmail.com

What does this study add to the existing literature and how will it influence future clinical practice?

Only 30.1% of patients hospitalized for symptomatic lower extremity artery disease are eligible to rivaroxaban 2.5 mg bid plus aspirin, based on the COMPASS or VOYAGER-PAD criteria. These patients have higher rates of ischaemic events compared with those in these trials. The one-year cumulative incidences were 6.0% (95%CI, 4.3-8.1) in the COMPASS-eligible subset vs 3.5% (95%CI, 2.9-4.3) in the COMPASS control arm for major adverse cardiovascular events, and 27.9% (95%CI, 19.9-38.3) in the VOYAGER-PAD-eligible subset vs 6.0% (95%CI, 5.3-6.9) in the VOYAGER-PAD control arm for major adverse limb events. This regimen may have greater benefits in the real-world population.

Abstract

Objectives: The aim of this study is to examine the external applicability of the COMPASS and the VOYAGER-PAD trials among patients with lower extremity artery disease (LEAD) in the real-world.

Design: It is a multicenter, retrospective analysis of prospectively collected COPART data, a French multicentre registry of patients hospitalized for symptomatic LEAD.

Materials and Methods: The proportion of patients eligible to the combination of rivaroxaban 2.5 mg bid plus aspirin based on either COMPASS or VOYAGER-PAD criteria is reported. The one-year cumulative incidence of outcomes between eligible vs non-eligible patients, as well as eligible patients vs control arms of the COMPASS (LEAD patients subgroup) and the VOYAGER-PAD trials are compared. Analyses were performed using Cox models.

Results: Among 2 259 evaluable patients, only 679 (30.1%) were eligible for low-dose rivaroxaban plus aspirin regimen. Others were non-eligible because of the need for anticoagulant (respectively 48.5% and 38.9% of patients meeting COMPASS and VOYAGER-PAD exclusion criteria) or dual antiplatelet therapy use (15.7% and 16.5%), high-bleeding risk (14.4% and 11.6%), malignancy (26.1% and 21.0%), history of ischaemic/haemorrhagic stroke (21.1% and 19.8%), and severe renal failure (13.2% and 10.5%). COMPASS and VOYAGER-PAD eligible and ineligible patients were at higher risk of ischaemic events than participants in these trials. The one-year cumulative incidences were 6.0% (95%CI, 4.3-8.1) in the COMPASS-eligible subset vs 3.5% (95%CI, 2.9-4.3) in the COMPASS control arm for MACE, and 27.9% (95%CI, 19.9-38.3) in the VOYAGER-PAD-eligible subset vs 6.0% (95%CI, 5.3-6.9) in the VOYAGER-PAD control arm for major adverse limb events.

Conclusion: Many patients hospitalized for symptomatic LEAD in France are not eligible for the low-dose rivaroxaban plus aspirin combination. In turn, those eligible may potentially have greater absolute benefit because of higher risk as compared to those enrolled in those trials.

Keywords: peripheral artery disease, rivaroxaban, cohort studies, mortality, amputation

Introduction

Patients with lower extremity artery disease (LEAD) are at heightened risk of major adverse cardiovascular events (MACE), including cardiovascular (CV) death, myocardial infarction (MI), and stroke, and are also at increased risk of all-cause mortality. Despite antithrombotic therapy, the one-year cumulative incidence of MACE is at least 12%, and is as high as 16.8% after revascularization for chronic limb-threatening ischemia (CLTI) [63,64]. The limb-related events rates are also high. In a meta-analysis, during a 5-years follow up, 21% of intermittent claudication (IC) patients were diagnosed as having CLTI, with 4-27% undergoing amputation [65].

To improve both the general CV and limb outcome, guidelines recommend a medical therapy including antiplatelet medication, as a pillar of CV prevention [66,67]. However, even when this therapy is well implemented, the residual risk in patients with LEAD is notable. Hence, new antithrombotic strategies have been recently assessed with promising results [134,154].

The COMPASS trial (Cardiovascular Outcomes for People Using Anticoagulation Strategies) tested the combination of rivaroxaban 2.5 mg bid plus aspirin for long term management of patients with stable atherosclerotic disease, and showed a significant reduction of 24% of MACE with this combination as compared with the aspirin alone arm, with consistent and significant results in the subgroup of patients with symptomatic LEAD [135,154].

The VOYAGER-PAD trial (Vascular Outcomes Study of Acetylsalicylic acid Along with Rivaroxaban in Endovascular or Surgical Limb Revascularization for Peripheral Artery Disease) tested the same combination and doses of rivaroxaban and aspirin in selected patients with symptomatic LEAD immediately after a successful revascularization procedure, a setting with high risk for ischemic and bleeding complications. This trial showed a significant reduction of 15% of the composite outcome of acute limb ischemia (ALI), major amputation for vascular causes and MACE in the low-dose rivaroxaban plus aspirin arm as compared with aspirin plus placebo arm [134]. In both trials, while an excess in bleeding was reported, the benefit/risk ratio was overall in favour of the use of this combined antithrombotic strategy.

However, it is widely accepted that the results obtained in trials need further validation in real-world clinical practice, as patients enrolled in CV trials are on average at lower risk because of younger age and less comorbidities [5,63,122]. Regarding patients with LEAD, the number eligible for the low-dose rivaroxaban plus aspirin combination is uncertain. It is also

unclear whether the patients selected for treatment with the combination of low-dose rivaroxaban plus aspirin will have similar risk in clinical practice than those enrolled in the COMPASS and the VOYAGER-PAD trials.

Therefore, we examined the external applicability of the COMPASS and the VOYAGER-PAD trials results by describing the proportion of COMPASS and VOYAGER-PAD eligible patients in a cohort of consecutive patients hospitalized for symptomatic LEAD in the vascular medicine departments in France. We also compared outcome rates between eligible vs non-eligible patients for COMPASS and VOYAGER-PAD trials, and between these eligible patients vs those randomized in the reference arm (i.e. aspirin arm) in those two trials.

Material and Methods

The COMPASS trial

The COMPASS (Cardiovascular Outcomes for People Using Anticoagulation Strategies) trial was a phase-3, double-blind randomized controlled trial (RCT), which randomly assigned 27 395 patients with stable atherosclerotic vascular disease (coronary artery disease [CAD], LEAD, or carotid artery disease) to receive rivaroxaban (2.5 mg bid) plus aspirin (100 mg once daily), rivaroxaban (5 mg bid) plus placebo, or aspirin (100 mg once daily) plus placebo [154]. The detailed COMPASS inclusion and exclusion criteria are described in Supplementary Material Table 1.

The VOYAGER-PAD trial

The VOYAGER-PAD (Vascular Outcomes Study of Acetylsalicylic Acid along with Rivaroxaban in Endovascular or Surgical Limb Revascularization for Peripheral Artery Disease) trial was a phase 3, double-blind RCT, which randomly assigned 6 564 patients with LEAD who had undergone revascularization to receive rivaroxaban (2.5 mg bid) plus aspirin (100 mg once daily), or aspirin (100 mg once daily) plus placebo [134]. Clopidogrel could be additionally prescribed at the discretion of the investigator, but not for more than 6 months. The detailed VOYAGER-PAD inclusion and exclusion criteria are described in Supplementary Material Table 3.

The COPART registry

The COPART (COhorte de Patients ARTériopathes) registry is an ongoing multicentre, observational, cohort study, since 2006 prospectively collecting data on consecutive patients hospitalized for symptomatic LEAD in the vascular medicine departments of 4 French academic centres (Bordeaux, Limoges, Paris, and Toulouse). Patients may have been hospitalized for scheduled surgery following a consultation, or for emergency care, or transferred from other hospitals or from others departments of the same hospital. Patients requiring only outpatient care for LEAD are not included in the COPART registry. Details on the study protocol have been published elsewhere [68]. Patient care was provided according to usual practice. After the initial hospitalization, patients were followed for at least 12 months. Data were prospectively collected by trained medical COPART investigators, following a

computerized standardized case report form, and using all records available during hospitalization and follow-up. Patients, their families and/or their physicians were systematically contacted one year after hospital discharge. The patients gave their informed consent to participate. The study complies with the declaration of Helsinki and was approved by the Ethic Committee of Toulouse University Hospital. The current report is based on a database locked in October 2017, including patients recruited from 2006 to 2015.

Each patient required the following criteria to be included in COPART registry: age ≥ 18 years, consent to participate, and specifically hospitalized for treatment and/or management of symptomatic LEAD of atherosclerotic origin. Clinical presentations included either an intermittent claudication, associated with abnormal ABI < 0.90 or > 1.30 , or, in the case of normal ABI at rest, a positive treadmill test (Strandness protocol, 3.2 km/h on a 10% slope), arterial stenosis $> 50\%$ on duplex ultrasound or angiography computerized tomography or angiography, or ischemic rest pain, ulceration or gangrene or an ALI related to a documented LEAD with significant arterial stenosis.

Patients for whom follow-up was improbable (i.e. patients who cannot be contacted after hospitalization, follow-up carried out in a foreign country), those with arterial occlusive disease not related to atherosclerosis, those with ALI without LEAD and those refusing to participate were not included.

“COMPASS-eligible” and “VOYAGER-PAD-eligible” study populations

Patients with missing data regarding eligibility criteria for COMPASS or VOYAGER-PAD were excluded. Then, the analysis started with COMPASS inclusion and exclusion criteria (Supplementary Material Table 1). A detailed list of adjustments made in these criteria for the analysis of the COPART registry is provided in Supplementary Material Table 2. First, patients meeting any adapted COMPASS exclusion criteria were excluded; the main reasons for exclusion were reported. Second, remaining patients were included in the “COMPASS-eligible” subset if they fulfilled adapted COMPASS inclusion criteria. The analysis continued with VOYAGER-PAD inclusion and exclusion criteria (Supplementary Material Table 3). A detailed list of adjustments made in these criteria for the analysis of the COPART registry is provided in Supplementary Material Table 4. First, patients not undergoing limb revascularization were excluded. Second, patients meeting any adapted VOYAGER-PAD exclusion criteria were excluded; the main reasons for these exclusions were reported. Third,

remaining patients were included in the “VOYAGER-PAD-eligible” subset if they fulfilled adapted VOYAGER-PAD inclusion criteria. Patients excluded from VOYAGER-PAD and those of the “VOYAGER-PAD-eligible” subset were included in the “COMPASS-eligible” subset if they fulfilled adapted COMPASS inclusion criteria without meeting any adapted COMPASS exclusion criteria.

Primary and secondary outcomes

For this study, the primary outcome is MACE, a composite of CV death, MI, and stroke, similar to the primary outcome of the COMPASS trial. The primary outcome of VOYAGER-PAD trial was the composite of MACE, ALI, and major amputation for vascular causes. ALI during the follow-up is not specifically recorded in COPART registry, however major amputation and limb revascularization are recorded, so ALI is included in the composite of limb revascularization or major amputation. Secondary outcomes also available in the COPART, COMPASS and VOYAGER-PAD databases were also analysed: all-cause death, CV death, non-CV death, MI, stroke, major amputation, and major adverse limb events (MALE), a composite of major amputation and limb revascularization. Revascularization could be achieved by percutaneous angioplasty (with or without stenting), subintimal recanalization, endarterectomy, embolectomy, surgical vein or prosthetic bypass. Amputation above the ankle level was considered major, and minor when below.

Statistical analysis

Baseline characteristics of the following subgroups are described using mean \pm standard deviation or median \pm interquartile range for continuous variables and frequencies and percentages for categorical variables:

- COMPASS-eligible subset: COPART patients with all the COMPASS inclusion criteria and none of the COMPASS exclusion criteria,
- COMPASS-non eligible subset: COPART patients without all the COMPASS inclusion criteria and/or with at least one COMPASS exclusion criteria,
- COMPASS participants with symptomatic LEAD of the aspirin alone arm: patients with symptomatic LEAD randomized in the aspirin alone group of the COMPASS trial [135],

- VOYAGER-PAD-eligible subset: COPART patients with limb revascularization, with all the VOYAGER-PAD inclusion criteria and none of the VOYAGER-PAD exclusion criteria,
- VOYAGER-PAD-non eligible subset: COPART patients with limb revascularization, without all the VOYAGER-PAD inclusion criteria and/or with at least one VOYAGER-PAD exclusion criteria,
- VOYAGER-PAD participants of the placebo arm: patients randomized in the aspirin group of the VOYAGER-PAD trial, with clopidogrel which can be additionally prescribed up to 6 months of limb revascularization [134].

Proportions of COMPASS exclusion criteria and VOYAGER-PAD exclusion criteria were reported.

Since in COMPASS trial rivaroxaban was given to patients with stable LEAD, the index date for COMPASS-eligible and non-eligible subsets is the date of discharge from the index hospitalization. For VOYAGER-PAD-eligible and non-eligible subsets, the index date is the date of the qualifying revascularization procedure.

Comparisons between subgroups are made using Chi-2 test for categorical variables or Fisher exact test when appropriate, while Student's t-test and Mann-Whitney test are used for continuous variables. Z-test was used to compare continuous variables between COMPASS-eligible subset and COMPASS participants with symptomatic LEAD of the aspirin arm. The Kaplan-Meier method is used to estimate cumulative incidence of outcomes at one year, with 95% confidence interval (95%CI). Patients were censored at the date of the first considered outcome. If they did not develop the outcome during the year of follow-up, they were censored at the date of the one-year final record completion or at the date of loss of follow-up. In order to allow statistical comparisons of outcomes between COMPASS participants with symptomatic LEAD and COMPASS-eligible subset, and between VOYAGER-PAD participants and VOYAGER-PAD-eligible subset, estimations of p-value based on 95%CI are given according to Krzywinski and Altman graphs [155].

No imputation was made on missing data. A p-value <0.05 was considered as statistically significant. Analyses were performed on STATA Statistical software (release 14.2, StataCorp LLC®, College Station, Texas).

Results

Rivaroxaban plus aspirin eligible patients in COPART registry

Among 2 520 patients in COPART registry, 261 were excluded because missing data in COMPASS or VOYAGER-PAD exclusion or inclusion criteria (Figure 1).

Among the remaining 2 259 evaluable patients, 679 (30.1%) were eligible for COMPASS. Among the 1 580 evaluable patients not eligible for COMPASS, 1 374 had at least one COMPASS exclusion criteria; the main reasons for exclusion were indication for full-dose oral anticoagulant treatment, known malignancy and history of haemorrhagic or ischemic stroke (Figure 2).

Among the 1 073 evaluable patients undergoing limb revascularization, 101 (9.4%) were eligible for VOYAGER-PAD. Among the 972 evaluable patients undergoing limb revascularization not eligible for VOYAGER-PAD, 868 had at least one VOYAGER-PAD exclusion criteria; the main reasons for exclusion were indication for full-dose oral anticoagulant treatment, major tissue loss and known malignancy (Figure 3).

All the VOYAGER-PAD-eligible patients were also eligible for COMPASS, therefore, only 30.1% of the entire COPART evaluable cohort was eligible for rivaroxaban 2.5 mg bid plus aspirin 100 mg once daily regimen, based on either COMPASS or VOYAGER-PAD criteria. The proportion of patients eligible for this regimen varied according to the year of inclusion in the COPART registry, between 21.2% and 38.1% based on COMPASS criteria and between 3.6% and 16.0% based on VOYAGER-PAD criteria, with a trend toward a decrease over the most recent years.

Comparison of baseline characteristics and outcomes of COMPASS-eligible and COMPASS-non eligible subsets

As compared with other COPART patients, COMPASS-eligible patients are younger, with lower prevalence of diabetes, history of CAD, stroke, heart failure and renal failure (Table 1). The clinical presentation of LEAD was less severe and use of statin more frequent. Regarding outcome, the cumulative incidences of MACE and all-cause mortality were lower than those non eligible for COMPASS (Figure 4, Supplemental Table 5).

Comparison of baseline characteristics and outcomes of COMPASS-eligible subset and COMPASS participants with symptomatic LEAD in the aspirin alone arm

As compared with participants with symptomatic LEAD randomized in the aspirin alone arm of COMPASS, proportions of diabetes, history of CAD and heart failure are higher in the COMPASS-eligible subset, but LEAD is less severe (Table 2). Cumulative incidence of MACE is higher in COMPASS-eligible subset (Figure 4, Supplemental Table 6). One-year cumulative incidence of MALE in COMPASS participants with symptomatic LEAD of the aspirin alone arm is not available, but the risk of MALE in COMPASS-eligible subset is higher, given the 30-month Kaplan-Meier incidence risk in the former group [135].

Comparison of baseline characteristics and outcomes of VOYAGER-PAD-eligible and VOYAGER-PAD-non eligible subsets

As compared with other COPART revascularized patients, proportions of hypertension, history of CAD, heart failure and renal failure are lower in the VOYAGER-PAD-eligible subset (Table 3). There was no statistically significant difference between the two groups regarding cumulative incidences of outcomes (Figure 5, Supplemental Table 7).

Comparison of baseline characteristics and outcomes differences between VOYAGER-PAD-eligible subset and VOYAGER-PAD participants of the placebo arm

As compared with VOYAGER-PAD participants randomized in the placebo arm, LEAD presentation is more severe in the VOYAGER-PAD-eligible subset and qualifying revascularization required more often surgery (Table 4). Compared with VOYAGER-PAD placebo arm, VOYAGER-PAD-eligible patients from COPART have a higher risk of major amputation (Figure 5, Supplemental Table 8). Cumulative incidence of overall mortality is higher in VOYAGER-PAD-eligible subset than in the VOYAGER-PAD placebo arm, but the difference is driven by non-CV death (Figure 5, Supplemental Table 8).

Discussion

Randomized-controlled trials constitute the highest level of evidence for new drug applications, line extensions, and recommendations for treatment. In 2018, the Committee for Medicinal Products for Human Use (CHMP) of the European Medicines Agency (EMA) issued a favourable opinion about the use of low-dose rivaroxaban on top of aspirin in patients with CAD or symptomatic LEAD at high risk of ischemic events. Multiple international agencies also granted low-dose rivaroxaban approval. Since then, this combination is being incorporated into international guidance statements: this association is suggested in patients with CAD or LEAD, including CLTI [62,108,156]. However, patients samples in RCTs, including those focused on LEAD, are highly selected with a lower risk profile than real-world populations, and are not broadly representative of patients encountered in daily practice [5,63,122]. Moreover, new therapies may be used in a broader patient population in clinical practice than that enrolled in the RCTs, and the regulatory approval may also extend beyond the inclusion and exclusion criteria of the trials [157]. Application of real-world data (RWD), mostly drawn from registries, to support drug approvals is now accepted by regulatory authorities as the EMA and the United States Food and Drug Administration (FDA) [6]. RWD can complement the results of phase 3 RCTs by providing insight into their applicability to clinical practice and informing health technology assessment bodies and payers [157].

In the present study, only 30.1% of the COPART evaluable cohort was eligible for rivaroxaban 2.5 mg bid plus aspirin regimen. VOYAGER-PAD criteria did not allow more patients to be eligible to this regimen than COMPASS criteria alone.

To date, no other registry assessed the external applicability of the VOYAGER-PAD trial. Concerning the external applicability of the COMPASS trial, in the REACH registry (Reduction of Atherothrombosis for Continued Health), 59.3% of CAD or peripheral artery disease (PAD) patients were eligible to secondary prevention with low-dose rivaroxaban plus aspirin: 68.4% of PAD alone patients, 60.4% of PAD + CAD patients, and 48.7% of CAD alone patients [158]. Only CAD patients are included in other cohorts, with 14.9% to 44.2% of them eligible for combined low-dose rivaroxaban and aspirin treatment [159–162].

As found in the REACH registry, indication for full-dose anticoagulant or DAPT, and high bleeding risk, are among the main reasons for being ineligible for low-dose rivaroxaban plus aspirin regimen [158]. Severe renal failure and major tissue loss are exclusion criteria respectively in about 10% of ineligible patients in COMPASS and VOYAGER subsets, and in

a third in the VOYAGER-PAD subset. These patients are at high-risk of ischemic events. Indeed, in the COMPASS trial, patients with symptomatic LEAD with high-risk limb presentations (prior amputation, Fontaine III or IV symptoms, previous peripheral artery revascularization) or high-risk comorbidities (kidney dysfunction, heart failure, diabetes, polyvascular disease) had a higher incidence of MACE and MALE than those with neither high-risk limb presentation nor high-risk comorbidities [135]. Moreover, after MALE, one-year cumulative incidences of MACE, all-cause mortality, and amputation were high [121]. Since this non-eligible population is likely to be among those who will benefit the most from combination therapy, further studies are needed to explore the applicability of this strategy. ALI is found in 13% of ineligible patients in the VOYAGER-PAD subset; even if low-dose rivaroxaban plus aspirin has not primarily been investigated after ALI, a small subgroup of patients within the COMPASS trial who had ALI also had a marked reduction in amputation and mortality rate, as highlighted in the European Society for Vascular Surgery (ESVS) guidelines [163]. Finally, malignancy is another major reason for ineligibility, found in about 20% of ineligible patients.

In the COPART registry, VOYAGER-PAD eligible patients had the same risk of MACE than VOYAGER-PAD participants in the reference arm and even higher risks of non-CV death, MALE, and major amputation. COMPASS eligible patients had higher risks of MACE and MALE than COMPASS LEAD participants in the reference arm. These results support the fact that the COMPASS and VOYAGER-PAD inclusion criteria define high-risk patients, which may derive particular benefit from low-dose rivaroxaban plus aspirin regimen.

However, COMPASS ineligible patients had an even higher risk of MACE than COMPASS eligible patients, which advocates for further studies enrolling these high-risk patients to clarify the benefit of this intervention in more representative populations. This is an aim of the prospective post-approval XATO registry study (Xarelto plus Acetylsalicylic acid: Treatment patterns and Outcomes in patients with Atherosclerosis) [157].

There are limitations to this analysis. Because the COPART registry recruited patients from 2006 to 2015 (recent data being entered), the management and prognosis of some patients could be different from what is currently observed. The differences found in event rates between the registry and the RCT may, at least partially, reflect the developments of revascularization procedures and secondary prevention during this period. The population from the COPART registry represents a nationwide sample of patients hospitalized for LEAD in university hospitals in France but cannot necessarily be extrapolated to patients hospitalized in non-

university hospitals, outpatients or patients from other countries. Of note, results from the main COMPASS trial demonstrated different results by geographical location, with high efficacy seen in the Asia-Pacific and Eastern European regions [154]. A few COMPASS and VOYAGER-PAD eligibility criteria were not collected as part of the COPART registry, and others required adjustment. For example, the exclusion criteria of a high bleeding risk was not captured in the COPART registry, so it was approached by the composite of any medical history of major bleeding and/or platelet count less than 100 G/L. Moreover, the date of major bleeding during follow-up was not captured in the registry so the Cox analysis was not available for this outcome. Finally, the cost of rivaroxaban is not considered when it could be a factor limiting the use of the treatment.

Conclusion

Based on the COMPASS or VOYAGER-PAD criteria, patients eligible to low-dose rivaroxaban plus aspirin represent only a third of patients with LEAD hospitalized in vascular departments of 4 tertiary-care hospitals in France. The higher rates of MALE and major amputation in these patients compared with those in the VOYAGER-PAD trial, and the higher rates of MACE compared with those in the COMPASS trial, may result in greater benefits from this regimen in the real-world population. However, ineligible patients represent a higher risk population, who could have benefited more from this regimen, requiring further research. There is a need for further clinical trials assessing the LEAD population that may benefit from this treatment in routine clinical practice.

Conflicts of interest

Dr. Lapébie reports personal fees from Bayer healthcare SAS, outside the submitted work. Dr. Aboyans reports personal fees from Bayer Healthcare, BMS/Pfizer, Novartis, and Sanofi, outside the submitted work. Dr. Lacroix, Dr. Constans, Dr. Boulon, Dr. Messas and Dr. Ferrières have nothing to disclose. Dr. Bongard reports personal fees from Sanofi outside the submitted work. Dr. Bura-Rivière reports grants, personal fees, non-financial support and other from Bayer Healthcare and BMS/Pfizer, grants and non-financial support from Astra-Zeneca, outside the submitted work.

Author contributions

Conception and design: FXL, VA, ABR

Analysis and interpretation: FXL, VA, PL, JC, CB, EM, JF, VB, ABR

Data collection: FXL, VA, PL, JC, CB, EM, ABR

Writing the article: FXL, VA, ABR

Critical revision of the article: PL, JC, CB, EM, JF, VB

Final approval of the article: FXL, VA, PL, JC, CB, EM, JF, VB, ABR

Statistical analysis: FXL

Obtained funding: ABR

Overall responsibility: ABR

Acknowledgements

We thank the research-assistants team: Delphine Baudoin, Clémence Labetoulle, Emmanuelle Lorian, Marion Mangin, Béatrice Montalègre, Monique Pailhol, Brigitte Pelvet, and Irina Smoliy for their technical help and data collection.

Funding source

The COPART registry is supported by the French Vascular Medicine Society (SFMV) and received grants from Toulouse university hospital, AstraZeneca France, Bayer France, Bristol-Myers Squibb (BMS) France and Sanofi-Aventis France companies.

Tables

Table I: Baseline characteristics of COMPASS-eligible and COMPASS-non eligible patients from COPART registry (time origin: hospital discharge)

Characteristics	COMPASS-eligible n = 679	COMPASS non-eligible n = 1 580	<i>p</i> *
<i>Demographic data</i>			
Age – years	66.7 ±12.9	72.6 ±12.3	< .001
Men	523 /678 (77.1)	1168 /1572 (74.3)	.153
<i>Cardiovascular risk factors</i>			
Diabetes	282 /679 (41.5)	842 /1580 (53.3)	< .001
Dyslipidemia	540 /679 (79.5)	1171 /1578 (74.2)	.007
Hypertension	564 /679 (83.1)	1385 /1579 (87.7)	.003
Current smokers	250 /676 (37.0)	303 /1553 (19.5)	< .001
Obesity	119 /636 (18.7)	305 /1438 (21.2)	.193
<i>Medical history</i>			
CAD	211 /677 (31.2)	632 /1571 (40.2)	< .001
Ischemic stroke	23 /679 (3.4)	337 /1569 (21.5)	< .001
Heart failure	35 /674 (5.2)	252 /1549 (16.3)	< .001
Atrial fibrillation	0 /668 (.0)	502 /1556 (32.3)	< .001
COPD or asthma	124 /665 (18.7)	300 /1532 (19.6)	.610
Cancer	0 /679 (.0)	359 /1580 (22.7)	< .001
Renal failure	186 /679 (27.4)	774 /1580 (49.0)	< .001
Hepatic failure	0 /679 (.0)	21 /1580 (1.3)	.003
<i>Laboratory exams</i>			
eGFR – mL/min/1.73 m ²	77.6 ±32.4	59.2 ±25.2	< .001
<i>LEAD at hospital admittance</i>			
Prior revascularization	287 /675 (42.5)	464 /1564 (29.7)	< .001
Prior amputation	90 /673 (13.4)	220 /1564 (14.1)	.663
LEAD classification	-	-	< .001
Rutherford 0	4 /677 (.6)	13 /1575 (.8)	-
Rutherford 1, 2 or 3	261 /677 (38.6)	319 /1575 (20.3)	-
Rutherford 4	80 /677 (11.8)	150 /1575 (9.5)	-
Rutherford 5 or 6	285 /677 (42.1)	965 /1575 (61.3)	-
Acute limb ischemia	47 /677 (6.9)	128 /1575 (8.1)	-
Gangrene	59 /664 (8.9)	216 /1536 (14.1)	.001
<i>Medical treatment at hospital discharge</i>			
Aspirin or clopidogrel	637 /679 (93.8)	1033 /1390 (74.3)	< .001
Oral anticoagulant	0 /679 (.0)	432 /1575 (27.4)	< .001
Statin	542 /679 (79.8)	994 /1392 (71.4)	< .001
ACEI or ARB	417 /673 (62.0)	812 /1359 (59.8)	.337

Beta-blockers	170 /677 (25.1)	487 /1365 (35.7)	< .001
---------------	-----------------	------------------	--------

Data are presented as n (%) or mean \pm standard deviation (SD). CAD = coronary artery disease; COPD = chronic obstructive pulmonary disease; eGFR = estimated glomerular filtration rate; LEAD = lower extremity artery disease; ACEI = angiotensin-converting enzyme inhibitors; ARB = angiotensin receptor blockers.

* p-value of difference between COMPASS-eligible and COMPASS-non eligible subsets.

Table II: Baseline characteristics of COMPASS-eligible patients in COPART and actual subgroup of COMPASS participants with symptomatic LEAD from aspirin alone reference arm

Characteristics	COMPASS-eligible in COPART n = 679	Actual COMPASS participants with symptomatic LEAD (aspirin alone arm) n = 1 359	p *
Demographic data			
Age – years	66.7 ±12.9	66.7 ±8.8	1.000
Men	523 /678 (77.1)	960 /1359 (70.6)	.002
Risk factors and coexisting conditions			
Diabetes	282 /679 (41.5)	644 /1359 (47.4)	.012
Current smokers	250 /676 (37.0)	442 /1359 (32.5)	.046
Coronary artery disease	211 /677 (31.2)	725 /1359 (53.3)	< .001
Ischemic stroke	23 /679 (3.4)	65 /1359 (4.8)	.144
Heart failure	35 /674 (5.2)	226 /1359 (16.6)	< .001
Renal failure	186 /679 (27.4)	371 /1359 (27.3)	.964
LEAD history			
Previous revascularization	287 /675 (42.5)	563 /1359 (41.4)	.639
Previous amputation	90 /673 (13.4)	103 /1359 (7.6)	< .001
LEAD classification			
Rutherford 1, 2 or 3	261 /677 (38.6)	943 /1359 (69.4)	< .001
Rutherford 4	80 /677 (11.8)	44 /1359 (3.2)	< .001
Rutherford 5 or 6	285 /677 (42.1)	12 /1359 (.9)	< .001
Medication			
Aspirin or clopidogrel	637 /679 (93.8)	1359 /1359 (100.0)	< .001
Aspirin	497 /679 (73.2)	1359 /1359 (100.0)	< .001
Clopidogrel	140 /679 (20.6)	0 /1359 (.0)	< .001
Statin	542 /679 (79.8)	1051 /1359 (77.3)	.200
ACEI and/or ARB	417 /673 (62.0)	920 /1359 (67.7)	.010

Data are presented as n (%) or mean ± standard deviation (SD). LEAD = lower extremity artery disease; ACEI = angiotensin-converting enzyme inhibitors; ARB = angiotensin receptor blockers.

* p-value of difference between COMPASS-eligible subset and COMPASS participants with symptomatic LEAD in the aspirin alone arm.

Table III: Baseline characteristics of VOYAGER-PAD-eligible and VOYAGER-PAD-non eligible patients revascularized from COPART (time origin: index revascularization)

Characteristics	VOYAGER-PAD-eligible n = 101	VOYAGER-PAD non-eligible n = 972	p *
<i>Demographic data</i>			
Age – years	69.2 ±11.7	70.2 ±12.7	.411
Men	78 /101 (77.2)	725 /965 (75.1)	.642
<i>Cardiovascular risk factors</i>			
Diabetes	41 /101 (40.6)	491 /972 (50.5)	.058
Dyslipidemia	80 /101 (79.2)	750 /971 (77.2)	.653
Hypertension	78 /101 (77.2)	850 /971 (87.5)	.004
Current smokers	34 /101 (33.7)	232 /967 (24.0)	.032
Obesity	14 /95 (14.7)	189 /937 (20.2)	.204
<i>Medical history</i>			
CAD	30 /101 (29.7)	389 /967 (40.2)	.039
Ischemic stroke	0 /101 (.0)	163 /972 (16.8)	< .001
Heart failure	4 /101 (4.0)	128 /960 (13.3)	.007
Atrial fibrillation	0 /101 (.0)	238 /964 (24.7)	< .001
COPD or asthma	22 /100 (22.0)	185 /952 (19.4)	.539
Cancer	0 /101 (.0)	182 /972 (18.7)	< .001
Renal failure	30 /101 (29.7)	412 /972 (42.4)	.014
Hepatic failure	0 /101 (.0)	9 /972 (.9)	1.000
<i>Laboratory exams</i>			
eGFR – mL/min/1.73 m ²	75.3 ±23.3	63.8 ±31.9	.001
<i>LEAD at hospital admittance</i>			
Prior revascularization	41 /101 (40.6)	317 /968 (32.8)	.112
Prior amputation	7 /101 (6.9)	99 /963 (10.3)	.285
LEAD classification	-	-	< .001
Rutherford 0	0 /101 (.0)	3 /970 (.3)	-
Rutherford 1, 2 or 3	30 /101 (29.7)	276 /970 (28.5)	-
Rutherford 4	27 /101 (26.7)	79 /970 (8.1)	-
Rutherford 5 or 6	44 /101 (43.6)	496 /970 (51.1)	-
Acute limb ischemia	0 /101 (.0)	116 /970 (12.0)	-
Gangrene	7 /99 (7.1)	93 /957 (9.7)	.392
<i>Medical treatment at hospital discharge</i>			
Aspirin or clopidogrel	98 /101 (97.0)	723 /894 (80.9)	< .001
Oral anticoagulant	0 /101 (.0)	237 /972 (24.4)	< .001
Statin	83 /101 (82.2)	713 /896 (79.6)	.537
ACEI or ARB	64 /100 (64.0)	533 /875 (60.9)	.549
Beta-blockers	20 /101 (19.8)	316 /882 (35.8)	.001

Data are presented as n (%) or mean \pm standard deviation (SD). CAD = coronary artery disease; COPD = chronic obstructive pulmonary disease; eGFR = estimated glomerular filtration rate; LEAD = lower extremity artery disease; ACEI = angiotensin-converting enzyme inhibitors; ARB = angiotensin receptor blockers.

* p-value of difference between VOYAGER-PAD-eligible and VOYAGER-PAD-non eligible subsets.

Table IV: Baseline characteristics of VOYAGER-PAD-eligible patients in COPART and actual VOYAGER-PAD participants from aspirin reference arm

Characteristics	VOYAGER-PAD-eligible in COPART n = 101	Actual VOYAGER-PAD participants (aspirin reference arm) n = 3 278	p *
Age – years	69.2 ±11.7	67.0 ±12.0	NA
Men	78 /101 (77.2)	2421 /3278 (73.9)	.447
Body mass index – kg/m ²	24.6 ±4.3	26.0 ±5.9	NA
<i>Risk factors and coexisting conditions</i>			
Diabetes	41 /101 (40.6)	1316 /3278 (40.1)	.928
Dyslipidemia	80 /101 (79.2)	1968 /3278 (60.0)	< .001
Hypertension	78 /101 (77.2)	2658 /3278 (81.1)	.331
Current smokers	34 /101 (33.7)	1132 /3278 (34.5)	.856
Coronary artery disease	30 /101 (29.7)	1015 /3278 (31.0)	.787
Renal failure	30 /101 (29.7)	666 /3278 (20.3)	.022
<i>LEAD history</i>			
Prior revascularization	41 /101 (40.6)	1155 /3278 (35.2)	.267
Prior amputation	7 /101 (6.9)	196 /3278 (6.0)	.692
<i>Qualifying revascularization</i>			
Performed for claudication	30 /101 (29.7)	2504 /3278 (76.4)	< .001
Performed for critical limb ischemia	71 /101 (70.3)	771 /3278 (23.5)	< .001
Endovascular	35 /101 (34.7)	2140 /3278 (65.3)	< .001
Surgical	66 /101 (65.3)	1138 /3278 (34.7)	< .001
<i>Medication</i>			
Aspirin at randomization	76 /101 (75.3)	3248 /3278 (99.1)	< .001
Clopidogrel at randomization	22 /101 (21.8)	1655 /3278 (50.5)	< .001
Statin	83 /101 (82.2)	2641 /3278 (80.6)	.687
ACEI or ARB	64 /100 (64.0)	2063 /3278 (62.9)	.828

Data are presented as n (%) or median ± interquartile range (IQR). LEAD = lower extremity artery disease; ACEI = angiotensin-converting enzyme inhibitors; ARB = angiotensin receptor blockers.

* p-value of difference between VOYAGER-PAD-eligible subset and VOYAGER-PAD participants in the aspirin reference arm.

Supplemental Table I: COMPASS Randomized Controlled Trial inclusion and exclusion criterion for patients with peripheral arterial disease of the lower limb

Inclusion criteria	Willing and able to provide written informed consent;	
	Peripheral arterial disease of the lower limb is defined as:	<ul style="list-style-type: none"> a. previous aorto-femoral bypass surgery, limb bypass surgery, or percutaneous transluminal angioplasty revascularization of the iliac, or infra-inguinal arteries, or b. previous limb or foot amputation for arterial vascular disease, or c. history of intermittent claudication and one or more of the following: <ul style="list-style-type: none"> i. an ankle/arm blood pressure ratio <0.90, or ii. significant peripheral artery stenosis ($\geq 50\%$) documented by angiography, or by duplex ultrasound.
Exclusion criteria	High risk of bleeding;	
	Stroke within 1 month or any history of hemorrhagic or lacunar stroke;	
	Severe heart failure with known ejection fraction $<30\%$ or New York Heart Association (NYHA) class III or IV symptoms;	
	Estimated glomerular filtration rate <15 mL/min;	
	Need for dual antiplatelet therapy, other non-aspirin antiplatelet therapy, or oral anticoagulant therapy;	
	Known non-cardiovascular disease that is associated with poor prognosis (e.g., metastatic cancer) or that increases the risk of an adverse reaction to study interventions;	
	History of hypersensitivity or known contraindication for rivaroxaban, aspirin, pantoprazole, or excipients, if applicable;	
	Systemic treatment with strong inhibitors of both Cytochrome P450 isoenzyme 3A4 (CYP3A4) and p-glycoprotein (P-gp) (e.g., systemic azole antimycotics, such as ketoconazole, and human immunodeficiency virus [HIV]-protease inhibitors, such as ritonavir), or strong inducers of CYP3A4 (e.g., rifampicin, rifabutin, phenobarbital, phenytoin, and carbamazepine);	
	Any known hepatic disease associated with coagulopathy;	
	Subjects who are pregnant, breastfeeding, or are of childbearing potential, and sexually active and not practicing an effective method of birth control (e.g., surgically sterile, prescription oral contraceptives, contraceptive injections, intrauterine device, double-barrier method, contraceptive patch, male partner sterilization);	

	Previous assignment to treatment during the study;
	Concomitant participation in another study with investigational drug;
	Known contraindication for any study related procedures;
	An additional exclusion for the pantoprazole randomization is: need for continuous treatment with a proton pump inhibitor.

Supplemental Table II: Adaptation from COMPASS exclusion and inclusion criteria, to COPART registry

COMPASS inclusion criteria, as labeled in Trial Design	Adapted inclusion criteria on COPART registry
Willing and able to provide written informed consent.	Not applicable to COPART cohort.
COMPASS exclusion criteria, as labeled in Trial Design	Adapted exclusion criteria on COPART registry
High risk of bleeding;	No such information was available in COPART case report form. We then assumed that patients with any medical history of major bleeding, or platelet count less than 100 G/L at the time of the index hospitalization are at significant risk for major bleeding.
Stroke within 1 month or any history of hemorrhagic or lacunar stroke;	The date of stroke before the index hospitalization was not available in COPART case report form, and stroke and lacunar stroke are not differentiated. We then assumed that patients with a history of stroke potentially have a history of recent stroke.
Severe heart failure with known ejection fraction <30% or New York Heart Association (NYHA) class III or IV symptoms;	There are no mentions of neither ejection fraction nor NYHA status in COPART case report form. We have chosen to keep patients labeled with “heart failure” at baseline or during the index hospitalization, because the number of such patients without other cause of exclusion is low.
Known non-cardiovascular disease that is associated with poor prognosis (e.g., metastatic cancer) or that increases the risk of an adverse reaction to study interventions;	No such information was available in COPART case report form. We have chosen to exclude patients labeled with “cancer” in COPART case report form.
Subjects who are pregnant, breastfeeding, or are of childbearing potential, and sexually active and not practicing an effective method of birth control (e.g., surgically sterile, prescription oral contraceptives, contraceptive injections, intrauterine device, double-barrier method, contraceptive patch, male partner sterilization);	No such information was available in COPART case report form. However, women of childbearing potential were uncommon in COPART registry. We then assumed that no pregnant or breastfeeding women were included in the COPART registry and that women of reproductive potential had adequate contraception.
Known contraindication for any study related procedures;	Not applicable to COPART cohort.
Previous assignment to treatment during this study;	

<p>Concomitant participation in another study with investigational drug;</p> <p>An additional exclusion for the pantoprazole randomization is: need for continuous treatment with a proton pump inhibitor.</p>	
--	--

COMPASS Exclusion criteria that could not be applied to COPART Registry:

- History of hypersensitivity for rivaroxaban, aspirin, or excipients;
- Systemic treatment with strong inhibitors of both Cytochrome P450 isoenzyme 3A4 (CYP3A4) and p-glycoprotein (P-gp) (e.g., systemic azole antimycotics, such as ketoconazole, and human immunodeficiency virus [HIV]-protease inhibitors, such as ritonavir), or strong inducers of CYP3A4 (e.g., rifampicin, rifabutin, phenobarbital, phenytoin, and carbamazepine).

Supplemental Table III: VOYAGER-PAD Randomized Controlled Trial inclusion and exclusion criterion

Inclusion criteria	Willing and able to provide written informed consent;
	Age ≥ 50 years;
	Documented moderate to severe symptomatic lower extremity atherosclerotic peripheral artery disease as evidenced by all of the following: <ul style="list-style-type: none"> a. clinically, by functional limitations in walking activity, ischemic rest pain, or ischemic ulceration, b. anatomically, by imaging evidence of peripheral artery disease distal to the external iliac artery in the index leg within 12 months prior to or at the time of the qualifying revascularization, c. hemodynamically in either leg (within 12 months prior to, or at the time of, the qualifying revascularization) by: <ul style="list-style-type: none"> i. an $ABI \leq 0.80$ or $TBI \leq 0.60$ for patients without a prior history of limb revascularization, OR ii. an $ABI \leq 0.85$ or $TBI \leq 0.65$ for patients with a prior history of limb revascularization;
	Technically successful peripheral revascularization distal to the external iliac artery (surgical and/or endovascular) for symptomatic PAD within the last 10 days prior to randomization;
	Negative serum pregnancy test (in women of childbearing potential only);
	Women of reproductive potential must agree to use adequate contraception when sexually active. This applies for the time period between signing of the informed consent form to the last administration of study drug.
Exclusion criteria	Patients undergoing revascularization for asymptomatic PAD or mild claudication without functional limitation of the index leg;
	Patients undergoing revascularization of the index leg to treat an asymptomatic or minimally symptomatic restenosis of a bypass graft or target lesion restenosis;
	Prior revascularization on the index leg within 10 days of the qualifying revascularization;
	Acute limb ischemia within 2 weeks prior to the qualifying revascularization;
	Patients with major tissue loss (defined as significant ulceration/gangrene proximal to the metatarsal heads, i.e. heel or midfoot) in either leg;
	Patients requiring treatment with ASA at doses >100 mg;

	<p>Planned dual antiplatelet therapy (DAPT) use for the qualifying revascularization procedure of clopidogrel in addition to ASA for >6 months after the qualifying revascularization procedure; it is strongly recommended that any course of clopidogrel is kept to the minimum necessary in accordance with local standard of care and international practice guidelines (typically 30 days, or up to 60 days for some drug-coated products or devices), and is only allowed for up to 6 months for complex procedures or devices that in the investigator’s opinion require longer use;</p>
	<p>Planned* use of any additional antiplatelet agent other than clopidogrel and ASA after the qualifying revascularization procedure (*This exclusion criterion refers to the clinical condition at the time of randomization. The use of DAPT with ASA plus clopidogrel, for new indication(s) occurring after randomization is permitted);</p>
	<p>Any active clinical condition requiring systemic anticoagulation after the qualifying revascularization;</p>
	<p>Hypersensitivity or any other contraindication listed in the local labeling for ASA or rivaroxaban;</p>
	<p>Systemic treatment with strong inhibitors of both Cytochrome P450 isoenzyme 3A4 (CYP3A4) and p-glycoprotein (P-gp) (e.g., systemic azole antimycotics, such as ketoconazole, and human immunodeficiency virus [HIV]-protease inhibitors, such as ritonavir), or strong inducers of CYP3A4 (e.g., rifampicin, rifabutin, phenobarbital, phenytoin, and carbamazepine) anticipated after randomization or during the study;</p>
	<p>Medical history or active clinically significant bleeding, lesions, or conditions within the last 6 months prior to randomization, considered to be a significant risk for major bleeding (this may include current medically confirmed gastrointestinal ulceration, presence of malignant neoplasms at high risk of bleeding, current or recent brain or spinal injury, known esophageal varices, vascular aneurysms of the large arteries or major intraspinal or intracerebral vascular abnormalities);</p>
	<p>Any known hepatic disease associated with coagulopathy or bleeding risk;</p>
	<p>Any condition requiring dialysis or renal replacement therapy, or a renal impairment at screening assessed with an estimated glomerular filtration rate (eGFR) <15 mL/min/1.73 m² (if a patient’s eGFR is <30 mL/min/1.73 m² prior to the procedure, it must remain to be >15 mL/min/1.73 m² 72 hours after the procedure to enroll and randomize the patient);</p>
	<p>Confirmed acute coronary syndrome within 30 days prior to randomization;</p>
	<p>Major trauma or accidents within 30 days prior to randomization;</p>
	<p>Any medically documented history of intracranial hemorrhage, stroke, or transient ischemic attack;</p>
	<p>Known active malignancy (as determined through review of medical history), excluding local skin cancer (basal or squamous cell carcinoma);</p>

	Poorly controlled diabetes (at the discretion of investigator);
	Severe uncontrolled hypertension (at the discretion of investigator);
	Overall life expectancy <1 year;
	Previous assignment to treatment during this study;
	Previous (within 30 days) or concomitant participation in another clinical study with investigational product(s);
	Close affiliation with the investigational site; e.g., a close relative of the investigator, dependent person (e.g., employee or student of the investigational site);
	Breast feeding.

Supplemental Table IV: Adaptation from VOYAGER-PAD exclusion and inclusion criteria, to COPART registry

VOYAGER-PAD inclusion criteria, as labeled in Trial Design	Adapted inclusion criteria on COPART registry
Technically successful peripheral revascularization distal to the external iliac artery (surgical and/or endovascular) for symptomatic PAD within the last 10 days prior to randomization;	No information was available in COPART case report form regarding arterial patency after the qualifying revascularization. We then assumed that the qualifying revascularization of the index leg was technically successful if a new revascularization and/or major amputation on the index leg was not necessary within 10 days after.
Negative serum pregnancy test (in women of childbearing potential only); Women of reproductive potential must agree to use adequate contraception when sexually active. This applies for the time period between signing of the informed consent form to the last administration of study drug.	No such information was available in COPART case report form. However, women of childbearing potential were uncommon in COPART registry. We then assumed that no pregnant women were included in the COPART registry and that women of reproductive potential had adequate contraception.
Willing and able to provide written informed consent.	Not applicable to COPART cohort.
VOYAGER-PAD exclusion criteria, as labeled in Trial Design	Adapted exclusion criteria on COPART registry
Planned dual antiplatelet therapy (DAPT) use for the qualifying revascularization procedure of clopidogrel in addition to ASA for >6 months after the qualifying revascularization procedure;	No such information was available in COPART case report form. We excluded patients with DAPT at hospital discharge.
Medical history or active clinically significant bleeding, lesions, or conditions within the last 6 months prior to randomization, considered to be a significant risk for major bleeding (this may include current medically confirmed gastrointestinal ulceration, presence of malignant neoplasms at high risk of bleeding, current or recent brain or spinal injury, known esophageal varices, vascular aneurysms of the large arteries or major intraspinal or intracerebral vascular abnormalities);	No such information was available in COPART case report form. We then assumed that patients with any medical history of major bleeding, or platelet count less than 100 G/L at the time of the index hospitalization are at significant risk for major bleeding.
Confirmed acute coronary syndrome within 30 days prior to randomization;	No such information was available in COPART case report form. We then assumed that patients with a medical history

	of acute coronary syndrome and dual antiplatelet therapy (DAPT) at the start of the index hospitalization, or patients with a diagnosis of acute coronary syndrome during the index hospitalization, potentially have a history of recent acute coronary syndrome.
Major trauma or accidents within 30 days prior to randomization;	No such information was available in COPART case report form. However, patients with recent major trauma or accidents should have suffer from major bleeding, which is reported in COPART casa report form.
Poorly controlled diabetes (at the discretion of investigator);	Poorly controlled diabetes was defined by HbA _{1c} level >9%.
Severe uncontrolled hypertension (at the discretion of investigator);	Severe uncontrolled hypertension was defined by systolic blood pressure \geq 180 mmHg and/or diastolic blood pressure \geq 110 mmHg.
Breast feeding;	No such information was available in COPART case report form. We then assumed that no breastfeeding women were included in the COPART registry.
Previous assignment to treatment during this study; Previous (within 30 days) or concomitant participation in another clinical study with investigational product(s); Close affiliation with the investigational site; e.g., a close relative of the investigator, dependent person (e.g., employee or student of the investigational site).	Not applicable to COPART cohort.

VOYAGER-PAD Exclusion criteria that could not be applied to COPART Registry:

- Hypersensitivity to ASA or rivaroxaban;
- Systemic treatment with strong inhibitors of both Cytochrome P450 isoenzyme 3A4 (CYP3A4) and p-glycoprotein (P-gp) (e.g., systemic azole antimycotics, such as ketoconazole, and human immunodeficiency virus [HIV]-protease inhibitors, such as ritonavir), or strong inducers of CYP3A4 (e.g., rifampicin, rifabutin, phenobarbital, phenytoin, and carbamazepine) anticipated after randomization or during the study;
- Overall life expectancy <1 year.

Supplemental Table V: Outcomes cumulative incidences at one year, for COMPASS-eligible patients in COPART as compared with non-eligible patients (time origin: hospital discharge)

Outcome	COMPASS-eligible	COMPASS non-eligible	Hazard ratio	<i>p</i>
Cardiovascular death, myocardial infarction or ischemic stroke	6.0 (4.3-8.1)	12.1 (10.4-14.0)	0.47 (0.33-0.68)	< .001
Overall mortality	8.9 (6.9-11.4)	20.4 (18.3-22.6)	0.41 (0.31-0.54)	< .001
Cardiovascular death	3.8 (2.6-5.7)	7.5 (6.2-9.1)	0.51 (0.32-0.79)	.003
Non-cardiovascular death	5.3 (3.8-7.3)	13.9 (12.1-15.9)	0.36 (0.25-0.52)	< .001
Myocardial infarction	1.5 (0.8-2.9)	3.8 (2.9-5.0)	0.38 (0.19-0.77)*	.007
Stroke	1.0 (0.5-2.2)	2.5 (1.7-3.5)	0.39 (0.16-0.94)	.037
Major amputation or limb revascularization	28.1 (24.8-31.8)	28.4 (26.0-31.0)	1.01 (0.85-1.21)	.879
Major amputation	6.9 (5.2-9.2)	9.5 (8.0-11.3)	0.73 (0.51-1.03)	.071
Limb revascularization	24.0 (20.8-27.5)	21.9 (19.7-24.3)	1.14 (0.93-1.39)	.199

Data are presented as % with 95% confidence interval (95%CI) or hazard ratio with 95%CI.

* Proportional hazard assumption not respected.

Supplemental Table VI: Outcomes cumulative incidences at one year, for COMPASS-eligible patients in COPART compared with actual COMPASS participants with symptomatic LEAD (aspirin alone arm)

Outcome	COMPASS-eligible	Actual COMPASS participants with symptomatic LEAD (aspirin alone arm)	<i>p</i> *
Cardiovascular death, myocardial infarction or ischemic stroke	6.0 (4.3-8.1)	3.5 (2.9-4.3)	< .010
Cardiovascular death	3.8 (2.6-5.7)	1.6 (1.2-2.2)	< .001
Myocardial infarction	1.5 (0.8-2.9)	1.5 (1.1-2.0)	NS
Stroke	1.0 (0.5-2.2)	0.8 (0.5-1.3)	NS

Data are presented as % with 95% confidence interval. LEAD = lower extremity artery disease; NS = non-significant.

* p-value of difference between COMPASS-eligible subset and COMPASS participants with symptomatic LEAD in the aspirin alone arm. p-values are estimated using Krzywinski and Altman graphs.

Supplemental Table VII: Outcomes cumulative incidences at one year, for VOYAGER-PAD-eligible patients revascularized in COPART as compared with non-eligible patients revascularized (time origin: index revascularization)

Outcome	VOYAGER-PAD-eligible	VOYAGER-PAD non-eligible	Hazard ratio	<i>p</i>
Cardiovascular death, myocardial infarction or ischemic stroke	8.6 (4.4-16.5)	12.8 (10.7-15.3)	0.66 (0.32-1.36)	.260
Overall mortality	12.2 (7.1-20.5)	18.5 (16.1-21.1)	0.64 (0.35-1.14)	.129
Cardiovascular death	5.3 (2.2-12.3)	7.2 (5.6-9.1)	0.73 (0.29-1.80)	.489
Non-cardiovascular death	7.3 (3.6-14.7)	12.2 (10.2-14.5)	0.58 (0.27-1.25)	.168
Myocardial infarction	2.3 (0.6-9.0)	4.3 (3.1-6.1)	0.53 (0.13-2.23)	.389
Stroke	1.1 (0.2-7.7)	3.2 (2.1-4.7)	0.37 (0.05-2.73)	.330
Major amputation or limb revascularization	27.9 (19.9-38.3)	30.2 (27.2-33.5)	0.90 (0.60-1.35)	.616
Major amputation	10.7 (5.9-18.9)	13.2 (11.1-15.7)	0.78 (0.41-1.50)	.461
Limb revascularization	20.8 (13.8-30.7)	22.7 (19.9-25.7)	0.92 (0.57-1.47)	.714

Data are presented as % with 95% confidence interval (95%CI) or hazard ratio with 95%CI.

Supplemental Table VIII: Outcomes cumulative incidences at one year, for VOYAGER-PAD-eligible patients in COPART compared with actual VOYAGER-PAD participants from aspirin reference arm

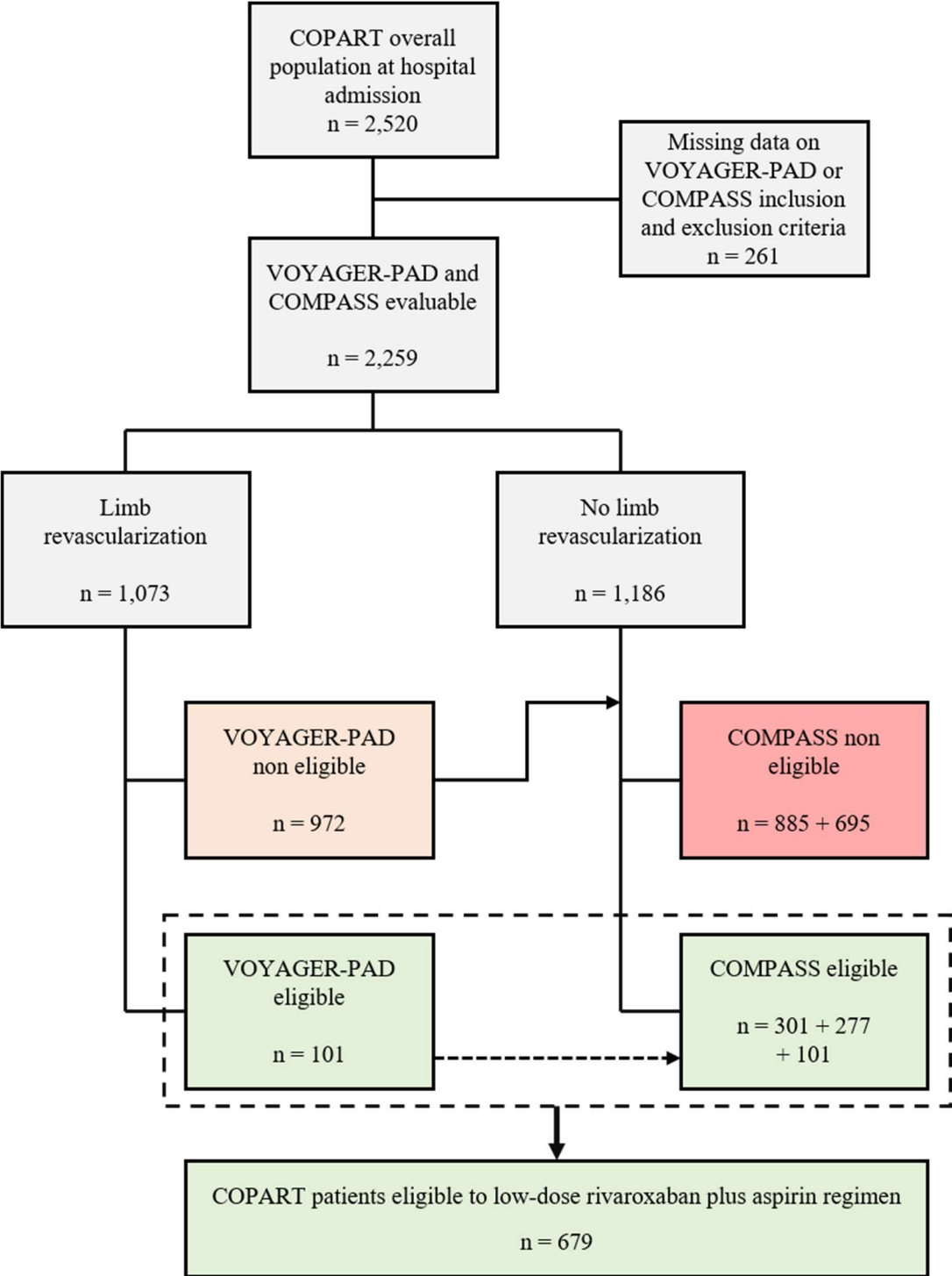
Outcome	VOYAGER-PAD-eligible in COPART	Actual VOYAGER-PAD participants (aspirin reference arm)	<i>p</i> *
Cardiovascular death, myocardial infarction or ischemic stroke	8.6 (4.4-16.5)	5.0 (4.3-5.8)	NS
Overall mortality	12.2 (7.1-20.5)	3.4 (2.9-4.1)	< .001
Cardiovascular death	5.3 (2.2-12.3)	2.2 (1.8-2.8)	NS
Non-cardiovascular death	7.3 (3.6-14.7)	1.2 (0.9-1.7)	< .001
Myocardial infarction	2.3 (0.6-9.0)	2.3 (1.8-2.9)	NS
Stroke	1.1 (0.2-7.7)	1.1 (0.8-1.5)	NS
Major amputation or limb revascularization	27.9 (19.9-38.3)	6.0 (5.3-6.9)	< .001
Major amputation	10.7 (5.9-18.9)	2.1 (1.7-2.7)	< .001

Data are presented as % with 95% confidence interval. NS = non-significant.

* p-value of difference between VOYAGER-PAD-eligible subset and VOYAGER-PAD participants in the aspirin reference arm. p-values are estimated using Krzywinski and Altman graphs.

Figures

Figure 1: Flow chart for identification of COMPASS and VOYAGER-PAD eligible patients in COPART registry



Footnote: Among the VOYAGER-PAD non eligible patients, 277 were eligible for COMPASS and 695 were not. All the VOYAGER-PAD-eligible patients were also eligible for COMPASS.

Figure 2: Main reasons for exclusion of patients in COPART registry, according to the exclusion criteria in COMPASS trial

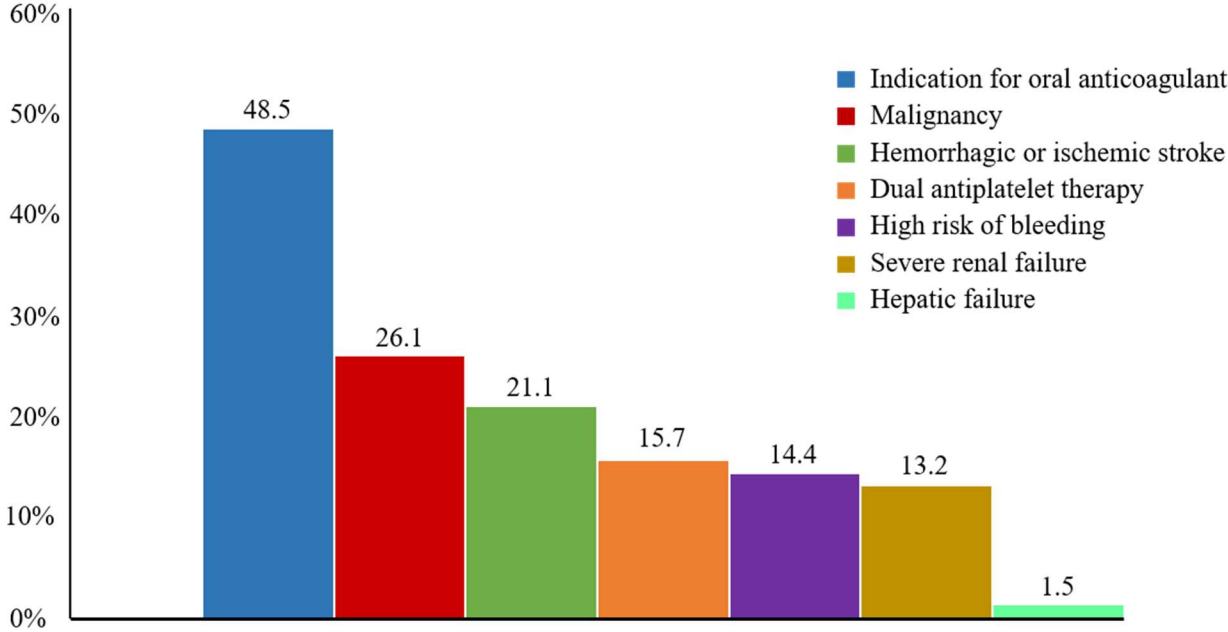


Figure 3: Main reasons for exclusion of patients revascularized in COPART registry, according to the exclusion criteria in VOYAGER-PAD trial

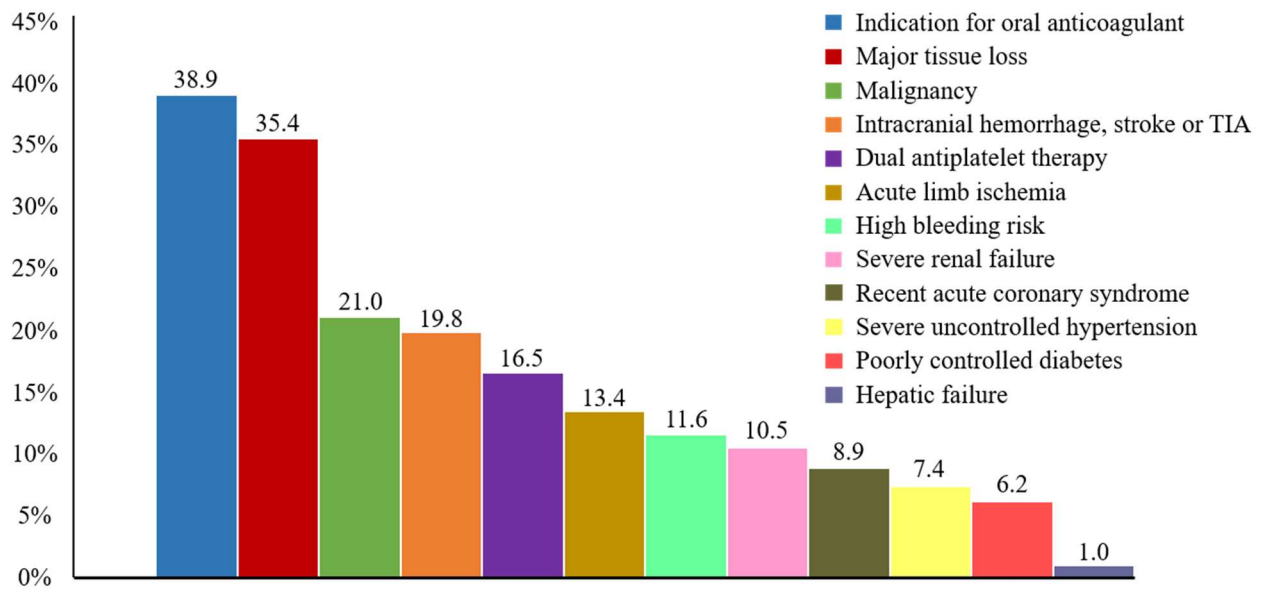
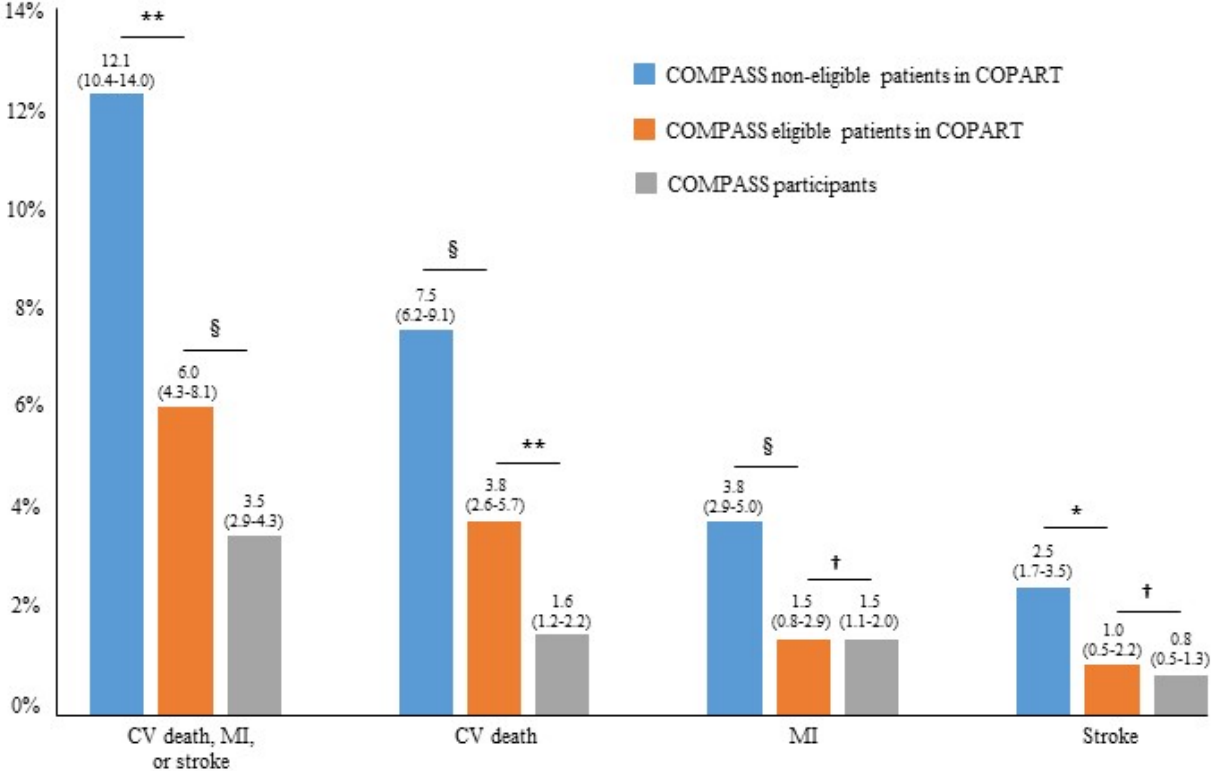
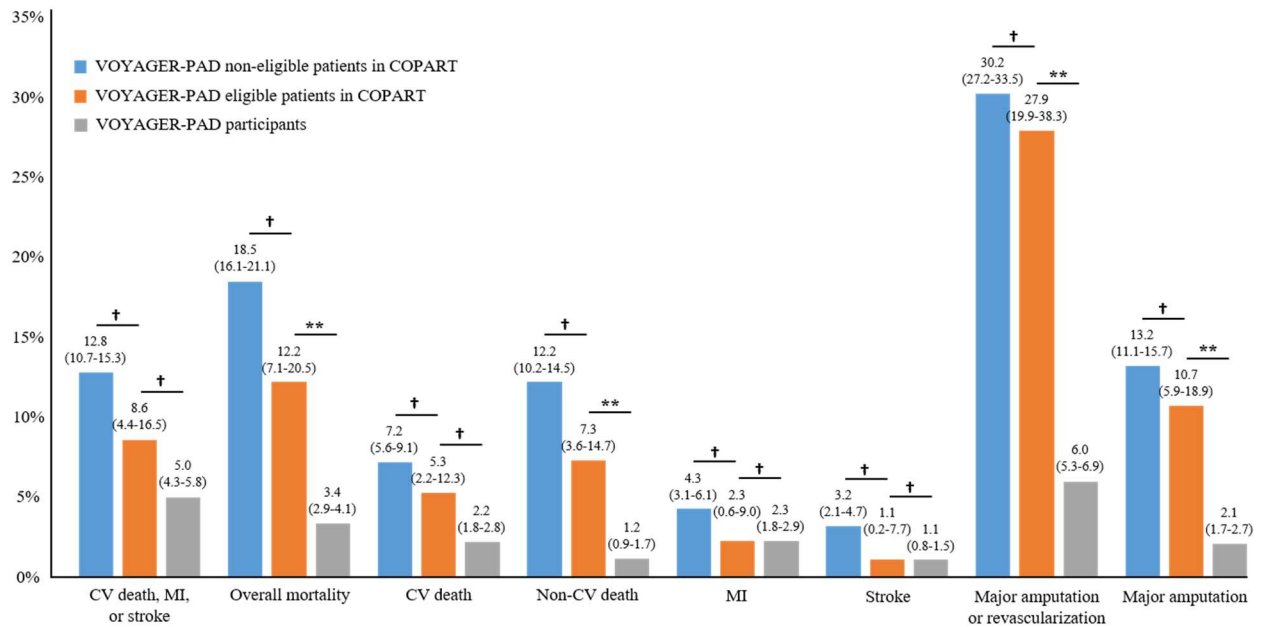


Figure 4: Comparison of one-year cumulative incidences of outcomes for the COMPASS eligible and non-eligible subsets from COPART, and COMPASS participants with symptomatic LEAD of the aspirin alone arm



Footnote: CV, cardiovascular; MI, myocardial infarction; †, p-value > .05 for comparison between groups; *, p-value < .05 for comparison between groups; §, p-value < .010 for comparison between groups; **, p-value < .001 for comparison between groups. Values are in percentage with 95%CI.

Figure 5: Comparison of one-year cumulative incidences of outcomes for the VOYAGER-PAD eligible and non-eligible subsets from COPART, and VOYAGER-PAD participants of the placebo arm



Footnote: CV, cardiovascular; MI, myocardial infarction; †, p-value > .05 for comparison between groups; *, p-value < .05 for comparison between groups; **, p-value < .001 for comparison between groups. Values are in percentage with 95%CI.

3.4.2. Résumé et remarques

Récemment, un anticoagulant oral direct à faible dose, le rivaroxaban, a été proposé en association à l'aspirine en prévention secondaire de la maladie athéromateuse. L'essai contrôlé randomisé COMPASS s'est intéressé aux patients présentant une maladie athéromateuse stable, alors que l'essai contrôlé randomisé VOYAGER-PAD s'adressait aux patients présentant une AAMI venant d'être revascularisée. Cette dernière étude évaluait également l'effet sur le devenir du membre opéré.

Les patients inclus dans ces essais, notamment VOYAGER-PAD, étaient très sélectionnés, poussant à s'interroger sur leur validité externe, et principalement leur transposabilité.

Les critères d'inclusion et de non-inclusion ont donc été appliqués à la cohorte COPART, et le pronostic des patients du registre qui auraient pu être inclus dans ces études a été comparé à celui des patients qui en auraient été exclus, ainsi qu'à celui des bras contrôle des essais. Le but de ces comparaisons était d'évaluer la gravité de la maladie, l'effet de l'intervention (rivaroxaban) pouvant ne pas être le même chez des patients à des stades différents de la maladie athéromateuse.

Ainsi, sur les 2259 patients retenus pour cette nouvelle analyse de COPART, 30,1% étaient éligibles à l'association rivaroxaban à faible dose + aspirine. Globalement, ces patients présentaient un risque cardiovasculaire plus élevé que ceux inclus dans COMPASS et VOYAGER-PAD. Il faut noter que ce risque était encore supérieur chez les patients de COPART qui auraient été exclus de ces deux essais.

Ces résultats remettent donc en cause la transposabilité des résultats de ces essais, et notamment la magnitude de l'effet de l'intervention chez les patients nécessitant une hospitalisation pour la prise en charge d'une AAMI.

Cet article a été publié dans *European Journal of Vascular and Endovascular Surgery* (en 2020, IF 7,069, rang A pour « Surgery » et « Peripheral vascular disease ») [164].

**3.5. Imputation multiple des données manquantes par équations chaînées :
article “Identification of factors associated with recurrence of cancer-
associated venous thrombo-embolism after discontinuation of
anticoagulant therapy: the RIETE registry”**

3.5.1. Manuscrit

Identification of factors associated with recurrence of cancer-associated venous thrombo-embolism after discontinuation of anticoagulant therapy: the RIETE registry

François-Xavier Lapébie, MD^{1,2}, Alessandra Bura-Rivière, MD, PhD^{1,3}, Olivier Espitia, MD, PhD⁴, Vanina Bongard, MD, PhD^{2,5,6}, Jean Ferrières, MD, PhD^{2,5,6}, Laurent Bertolletti, MD, PhD^{7,8,9}, Manuel Monreal, MD, PhD^{10,11}, and the RIETE investigators

¹Department of Vascular Medicine, Toulouse University Hospital, Toulouse, France

²UMR 1295 INSERM, CERPOP, Toulouse III – Paul Sabatier University, Toulouse, France

³UMR 1031 INSERM, StromaLab, Toulouse III – Paul Sabatier University, Toulouse, France

⁴Department of Internal Medicine and Vascular Medicine, Nantes University Hospital, Nantes, France

⁵Department of Epidemiology, Toulouse University Hospital, Toulouse, France

⁶Federation of Cardiology, Toulouse University Hospital, Toulouse, France

⁷Department of Vascular and Therapeutic Medicine, Saint-Etienne University Hospital, Saint-Etienne, France

⁸CIC 1408 INSERM, Saint-Etienne University Hospital, Saint-Etienne, France

⁹UMR 1059 INSERM, SAINBIOSE, Jean Monnet University, Saint-Etienne, France

¹⁰Department of Internal Medicine, Germans Trias i Pujol University Hospital, Badalona, Barcelona, Spain

¹¹Universidad Católica de Murcia, Murcia, Spain

Corresponding author

François-Xavier Lapébie

Department of Vascular Medicine, Toulouse University Hospital, Toulouse, France

Hôpital Rangueil, 1 avenue du Pr Jean Poulhès, TSA50032, 31059 Toulouse Cedex 9, France

Phone number: +33 5 61 32 24 38

Fax number: +33 5 61 32 26 34

E-mail: lapebie.fx@chu-toulouse.fr; fx.lapebie@gmail.com

Introduction

Cancer has long been recognized as a major persistent risk factor for venous thromboembolism (VTE), with at least a 4 to 6 fold increase of the risk of a first event [78,82]. However, incidence rates of VTE varies widely across specific cancer subpopulations, from 13 per 1,000 person-years (95% confidence interval [95%CI], 7-23) for average-risk patients to 68 per 1,000 person-years (95%CI, 48-96) for high-risk patients in a meta-analysis [83]. The incidence of VTE is highest among patients with pancreatic, brain, lung, stomach, esophageal or ovary cancer, metastatic disease, and adenocarcinoma [165,166]. The risk of VTE is also higher in hospitalized than in ambulatory cancer patients, within the first months after cancer diagnosis, and with cancer-directed therapy, especially chemotherapy [165,167].

Patients with malignancy that suffer an incident VTE are also at higher risk of VTE recurrence compared with those without cancer, both during and after anticoagulant therapy [165,168]. Thus, cumulative incidence of recurrent VTE after anticoagulant discontinuation for cancer-associated thrombosis (CAT) is about 9% at 6 months and 15% at one year, while the risk of recurrence is 10% in the first year after treatment for a first unprovoked VTE [76,81,84,85]. Recurrent VTE complicates treatment of cancer, reduces quality of life and is potentially fatal. Data on recurrence after anticoagulant discontinuation are scarce, but the case fatality rate of recurrent VTE in patients undergoing anticoagulation for CAT is 14.8% (95%CI, 6.6-30.1) [169]. During the first 12 months following a diagnosis of CAT, fatal VTE is the second leading cause of death [170].

The current guidelines recommend anticoagulation for at least 3 to 6 months for the treatment of CAT [171–173]. Treatment beyond should be considered for patients with active cancer, but an individual evaluation of benefits and risks is mandatory because the risk of bleeding during treatment is also higher in these patients, with an incidence of major bleeding of 13.1 per 100 patient-years (95%CI, 10.3-16.7), which is twice the risk of non-cancer patients, and a case fatality rate of 8.9% (95%CI, 3.5-21.1) [169,174].

The identification of factors associated with VTE recurrence is important for making a decision regarding management of anticoagulation beyond the first 3 to 6 months. However, published data on the risk of recurrence after the end of anticoagulant therapy are scarce, with most of data concerning recurrence during anticoagulant therapy [175]. The aim of our study is to determine risk factors for recurrence after a time-limited anticoagulant therapy of at least 3 months for CAT in a large cohort of patients diagnosed with VTE.

Material and Methods

The RIETE registry

The RIETE registry (*Registro Informatizado de la Enfermedad TromboEmbolica venosa*) is an ongoing, international, multicenter, prospective cohort study of consecutive patients diagnosed with symptomatic acute deep vein thrombosis (DVT), pulmonary embolism (PE), and/or superficial venous thrombosis (SVT) confirmed by objective tests (ClinicalTrials.gov identifier: NCT02832245). The rationale and methodology of RIETE have been published elsewhere [92]. Patients are excluded if they are currently participating in a therapeutic clinical trial or if they will not be available for a 3-month follow-up. All patients provide informed consent to their participation in the study, according to the requirements of the ethics committee within each hospital.

To be eligible in the present analysis, patients had to have an active cancer at the time of VTE and had to have received a time-limited full-dose anticoagulation (unfractionated heparin [UFH], low molecular weight heparin [LMWH], vitamin K antagonist [VKA], fondaparinux, and/or direct oral anticoagulant [DOA]) of at least 3 months after VTE diagnosis without recurrence during anticoagulant therapy.

Definition of active cancer was: a cancer diagnosed within the 6 months preceding or following the VTE diagnosis, and/or a metastatic cancer, and/or any cancer treatment (surgery, chemotherapy, radiotherapy, immunotherapy, hormonotherapy) at the time of VTE diagnosis. Basal or squamous-cell skin cancer were excluded. Index VTE has to be a PE (subsegmental or more proximal artery), a lower-extremity DVT (proximal or isolated distal), and/or an upper-extremity DVT (including catheter-related DVT). Isolated SVT was excluded. Proximal DVT was defined as the thrombus involving at least the popliteal vein and above, and distal DVT was defined as the thrombus involving veins below the popliteal level.

Primary and secondary outcomes

For this study, the primary outcome was first symptomatic recurrence of VTE during the one-year follow-up after discontinuation of an anticoagulant therapy of at least 3 months for a CAT. VTE recurrence was defined as a PE (subsegmental or more proximal artery), and/or a lower or upper-extremity DVT. SVT was excluded.

The secondary outcome was all-cause death during the one-year following the end of an anticoagulation of at least 3 months for a CAT.

Collected variables

Baseline variables collected were: age, male sex, body mass index (BMI), obesity (defined as BMI ≥ 30 kg/m²), estimated glomerular filtration rate (eGFR) according to the Chronic Kidney Disease-Epidemiology collaboration (CKD-EPI) equation, chronic kidney disease (defined as eGFR < 60 mL/min/1.73 m²), chronic heart failure, chronic lung disease, previous VTE, time between cancer and VTE, site of the cancer, metastases, chemotherapy, radiotherapy, immunotherapy, hormonal therapy, surgery during the 2 months before, confined to bed during the 2 months before, type of VTE (PE without lower extremity [LE]-DVT, LE-DVT without PE, upper extremity [UE]-DVT without PE, PE plus DVT), type of LE-DVT (proximal or isolated distal), bilateral DVT, type of PE (pulmonary artery trunk, main pulmonary artery, lobar artery, segmental artery, subsegmental artery), haemoglobin, platelets count, anticoagulant therapy after the first 7 days (UFH, LMWH, fondaparinux, VKA, DOA), length of anticoagulant therapy, major bleeding with anticoagulant therapy (International Society of Thrombosis and Haemostasis [ISTH]-defined), minor bleeding with anticoagulant therapy, inferior vena cava filter, post-thrombotic syndrome (PTS) according to Ginsberg's criteria, residual venous obstruction after 3 months (venous ultrasonography, angiography computerized tomography), residual pulmonary artery obstruction after 3 months (angiography computerized tomography), and D-dimers at discontinuation of anticoagulant therapy [113,114,176,177]. A four-class categorical variable was constructed for describing site of cancer, based on the risk of VTE by cancer site in the published data [178]. Of note, only 369 patients were included with an isolated UE-DVT, 69.1% being catheter-related thrombosis. In the bivariate Cox models, results were closed: hazard ratio (HR) 0.49 (95% confidence interval [95%CI], 0.30-0.79) for UE-DVT as compared with no UE-DVT, HR 0.49 (95%CI, 0.27-0.87) for catheter-associated UE-DVT as compared with no catheter-associated UE-DVT, and HR 0.57 (95%CI, 0.25-1.28) for UE-DVT not catheter-related as compared with no UE-DVT not catheter-related. Therefore, all isolated UE-DVT, catheter-related or not, were pooled in a single class in the categorical variable "type of VTE".

Statistical analysis

Baseline characteristics are described using mean \pm standard deviation or median \pm interquartile range for continuous variables and frequencies and percentages for categorical variables. Comparisons were made using Chi² test for categorical variables or Fisher exact test when appropriate, while Student t test or Mann-Whitney test were used for continuous variables.

The Kaplan-Meier method was used to estimate event cumulative incidences. The index date is the date of the end of anticoagulant for the index CAT. Patients are censored at the date of the first event. If they do not develop an event during the year of follow-up, they are censored at the date of the one-year final record or at the date of loss of follow-up, whichever comes first.

Unadjusted and multivariate analyses were performed using Cox proportional-hazard models, after checking for the proportional-hazard assumption and the log-linearity assumption for continuous variables.

As a sensitivity analysis, unadjusted and multivariate analyses were also performed using Fine and Gray proportional-hazard models for the subdistribution of a competing risk, with all-cause death as a competing risk, i.e. an event whose occurrence precludes the occurrence of the primary event of interest.

For the multivariate models, a two-tier predictor selection process was performed. The first tier included variables chosen based on knowledge regarding their relationship with VTE recurrence from the literature review, regardless the p-values in the unadjusted models [1]. The variables of the second tiers were identified by multivariate Cox model with backward stepwise elimination with a p-value greater than 0.157 for removal of variables [1]. Finally, the following variables were included in the Cox and Fine and Gray proportional-hazard multivariate models: male sex, eGFR, previous VTE, surgery during the 2 months before, anticoagulant therapy after the first 7 days, inferior vena cava filter, PTS, residual venous obstruction after 3 months, residual pulmonary artery obstruction after 3 months, time between cancer and VTE, site of the cancer, metastases, chemotherapy, radiotherapy, immunotherapy, and hormonal therapy.

Several potential predictors of VTE recurrence have missing data. Complete case analysis could be biased if data are not missing completely at random (MCAR) and would have prevented multivariate analyses by excluding many observations. Therefore, a multiple imputation by chained equations (MICE) was performed on missing values. MICE creates imputed datasets based on a set of imputation models, one model for each variable with missing values. The distribution of observed data is used to estimate a set of plausible values for missing data, instead of filling in a single value; that reflect the uncertainty around the true value. The missing values are replaced by the estimated plausible values to create a “complete” dataset. First, missing-value patterns were examined to assume that the data are not missing at random (MNAR). Then, missing values were predicted based on all other potential predictors, with the outcomes and auxiliary variables also included in the imputation models [179]. Auxiliary

variables were chosen a priori, based on literature, because of their association with potential predictors with missing data. Variables included in the imputation models were all the baseline variables listed above, the outcomes (first recurrence of VTE and all-cause death), cardiac frequency at VTE diagnosis, blood pressure at VTE diagnosis, and body weight. The predictive mean matching imputation method using 10 nearest neighbors was used for continuous variables. The logistic regression imputation method was used for binary variables. The ordered logistic regression imputation method was used for ordinal variables. Due to a high proportion of missing data in some variables, 100 imputed datasets were used [9]. Collinearity between variables and the convergence of MICE were verified; as a result of convergence issues, some variables were excluded from the imputation models of some missing-value variables. Finally, analyses ran on each imputed dataset were combined using the Rubin rule to produce an overall estimate of each regression coefficient or model performance measure, while taking into account uncertainty in the imputed values [180].

HR and sub-hazard ratio (sHR) were estimated with a 95%CI. A p-value <0.05 was considered to be statistically significant. Analyses were performed on STATA Statistical software (release 17.1, StataCorp LLC®, College Station, Texas).

Results

From March 2001 to August 2018, 14,318 patients were included in the RIETE registry with a CAT (Figure 1). Eventually, 3,414 patients were eligible for this study. Comparison between these 3,414 patients and the 6,532 patients with extended anticoagulant therapy was presented in Supplemental Table 1. The median follow-up of the studied group after the end of anticoagulant was 113 days, interquartile range (IQR) 370 days. During a one-year follow-up, 253 patients were diagnosed with a recurrent VTE with a cumulative incidence of 13.5% (95%CI, 11.9-15.2) (Figure 2), and 993 patients died with a cumulative incidence of 35.9% (95%CI, 34.0-37.8).

A history of VTE, cancer of the pancreas, biliary tract or unknown origin, the association of PE plus DVT, proximal DVT, the presence of a vena cava filter after the end of anticoagulant, PTS, and residual venous or pulmonary artery obstruction, were more frequent in the recurrence group than in the non-recurrence group (Table 1).

The percentage of missing values ranged from 0 for some demographic variables to as high as 82,5% for data about D-dimers at discontinuation of anticoagulant therapy (Table 2), and only 135 of the 3,414 patients (4,0%) in the sample would have been available for analysis under the traditional complete case analysis. Imputed values compare reasonably to observed values, and results using complete case are similar to MICE, so imputed results are presented (Table 2). Factors associated with VTE recurrence in the unadjusted Cox analyses were a history of VTE, chronic kidney disease, cancer of lung, brain, stomach, esophagus, liver, ovary, pancreas, biliary tract or unknown origin compared with cancer of oropharynx, larynx or melanoma, metastasis, the association of PE plus DVT compared with isolated PE, proximal DVT compared with no DVT, the presence of a vena cava filter after the end of anticoagulant, PTS, and residual pulmonary artery obstruction. Conversely, a surgery during the two months before VTE was associated with half the recurrence. This was confirmed in the unadjusted Fine and Gray analyses, except for the presence of metastasis, which was no longer statistically significant (Supplemental Table 2).

After adjustment, surgery remained negatively associated with VTE recurrence (Table 3). A positive relationship was confirmed for a history of VTE, chronic kidney disease, cancer of lung, brain, stomach, esophagus, liver, ovary, pancreas, biliary tract or unknown origin, inferior vena cava filter, PTS, and residual pulmonary obstruction. This was confirmed in the multivariate Fine and Gray model (Supplemental Table 3).

Discussion

VTE is one of the major complications of cancer. The risk of VTE recurrence after stopping anticoagulant therapy was particularly high in the present study. Data on the safety of stopping anticoagulant after a CAT are scarce, but this risk is closed to that found in another cohort study, with a incidence rate of 19 per 100 person-years (95%CI, 9.3-33) for patients who stopped anticoagulant treatment despite active cancer [181]. The active nature of cancer, which is recognized as a major risk factor of recurrence, was not recorded at the time of the end of anticoagulant therapy in RIETE.

A metastatic disease is also recognized as a risk factor of recurrence [182]. In our study, this association was no longer statistically significant with the adjustment on cancer site in the Cox analysis.

The association of cancer of lung, brain, stomach, esophagus, liver, ovary, pancreas or biliary tract with a first CAT is constantly reported, so their association with VTE recurrence after stopping anticoagulant therapy was expected [178]. It was also confirmed in the multivariate Fine and Gray analysis, while the important risk of death linked to these cancers could have modified the results, as was the case with metastasis.

Residual pulmonary artery or venous obstruction after anticoagulant treatment for PE or DVT, and PTS have already been described as predictors of recurrent VTE [183–185]. But to our knowledge, it is the first time that this is shown for CAT.

It has already been shown that inferior vena cava filter in cancer patients with DVT or PE was associated with a higher risk of recurrent DVT, but our study confirmed this result after the end of anticoagulant therapy [186].

In the multivariate Fine and Gray model, but not in the multivariate Cox model, DOA were associated with a higher risk of VTE recurrence than LMWH or fondaparinux after the end of anticoagulant therapy. Besides inhibiting coagulation, LMWH could reduce metastatic spread, which is associated with the risk of VTE, through inhibiting cell adhesion molecules [187]. However, a recent meta-analysis of randomized controlled trials (RCT) showed a non-significant reduction in recurrent VTE during the 6 months of treatment with DOA as compared to LMWH and comparable risks of mortality, and we cannot exclude the possibility of variables non controlled by the multivariate model, affecting the evaluation of risk [188].

Surgery is a well known major transient risk factor of VTE, usually associated with a low risk of recurrence after a time-limited anticoagulant therapy [78]. In our study, the risk of recurrence was halved for surgery associated VTE, despite the presence of active cancer. This association and its magnitude were consistent in all the analyses. To our knowledge, no published data had noted this.

Our analysis has some limitations. Because of the observational nature of the study, and despite multivariate analyses, we cannot exclude the possibility of non-controlled variables differing in the two groups which could affect the outcome. MICE has been shown to be a robust method to deal with missing data, but even if a high number of 100 imputed datasets was used, relative efficiency for effect estimates, especially standard errors, may be penalized for the covariates with missing values greater than 50%. Cancer-directed therapy and the active nature of cancer, which are known to be associated with VTE recurrence, were not recorded at the time of the end of anticoagulant therapy.

Conclusion

Cancer of lung, brain, stomach, esophagus, liver, ovary, pancreas or biliary tract, history of VTE, chronic kidney disease, PTS, residual pulmonary artery obstruction, and persistent inferior vena cava filter were associated with VTE recurrence after anticoagulant discontinuation for a first episode of CAT. Among the transient risk factors of VTE, surgery during the two months before VTE was associated with a halving of recurrence. These findings could help to develop a multivariable prognostic model, to predict the risk of VTE recurrence after anticoagulant discontinuation. This model would help clinicians to decide which patients can safely discontinue anticoagulant therapy after the first 3 to 6 months, and which would receive extended treatment.

Tables

Table 1: Characteristics of the study population at baseline

Characteristics	All (n = 3,414)	Recurrent VTE (n = 253)	No recurrent VTE (n = 3,161)	p
Age – year	66.5 ±13.2	67.4 ±12.3	66.5 ±13.3	0.292
Male sex	1803 /3414 (52.8)	138 /253 (54.6)	1665 /3161 (52.7)	0.566
Obesity	788 /3414 (23.1)	75 /253 (29.6)	713 /3161 (22.6)	0.010
BMI – kg/m ²	27.2 ±5.1	27.9 ±5.1	27.2 ±5.1	0.042
Chronic kidney disease	701 /3317 (21.1)	71 /244 (29.1)	630 /3073 (20.5)	0.002
eGFR – mL/min/1.73 m ²	74.5 ±26.6	68.9 ±25.4	74.9 ±26.6	0.004
Chronic heart failure	146 /3414 (4.3)	12 /253 (4.7)	134 /3161 (4.2)	0.703
Chronic lung disease	362 /3414 (10.6)	24 /253 (9.5)	338 /3161 (10.7)	0.549
Previous VTE	323 /3414 (9.5)	34 /253 (13.4)	289 /3161 (9.1)	0.025
Time between cancer and DVT – day	616 ±1185	639 ±1097	614 ±1192	0.704
Site of the cancer	/	/	/	0.006
Oropharynx, larynx, melanoma	118 /3414 (3.5)	3 /253 (1.2)	115 /3161 (3.6)	/
Others, haematological, colorectal, uterus, bladder, kidney, prostate, breast, vulva	2311 /3414 (67.7)	163 /253 (64.4)	2148 /3161 (68.0)	/
Lung, cerebral, stomach, oesophagus, liver, ovary	843 /3414 (24.7)	68 /253 (26.9)	775 /3161 (24.5)	/
Pancreas, biliary system, carcinoma of unknown origin	142 /3414 (4.2)	19 /253 (7.5)	123 /3161 (3.9)	/
Metastases	1352 /3113 (43.4)	103 /232 (44.4)	1249 /2881 (43.4)	0.758

Chemotherapy	1759 /3277 (53.7)	130 /246 (52.9)	1629 /3031 (53.7)	0.786
Radiotherapy	476 /3224 (14.8)	42 /245 (17.1)	434 /2979 (14.6)	0.275
Immunotherapy	11 /2121 (0.5)	1 /177 (0.6)	10 /1944 (0.5)	1.000
Hormonal therapy	404 /3383 (11.9)	26 /251 (10.4)	378 /3132 (12.1)	0.421
Surgery during the 2 months before	627 /3414 (18.4)	28 /253 (11.1)	599 /3161 (19.0)	0.002
Confined to bed during the 2 months before	516 /3414 (15.1)	35 /253 (13.8)	481 /3161 (15.2)	0.555
Type of venous thromboembolism event	/	/	/	0.002
PE without LE-DVT	931 /3414 (27.3)	58 /253 (22.9)	873 /3161 (27.6)	/
LE-DVT without PE	1344 /3414 (39.4)	100 /253 (39.5)	1244 /3161 (39.4)	/
UE-DVT without PE	369 /3414 (10.8)	17 /253 (6.7)	352 /3161 (11.1)	/
PE and DVT	770 /3414 (22.3)	78 /253 (30.8)	692 /3161 (21.9)	/
Type of LE-DVT	/	/	/	0.001
No LE-DVT	1370 /3383 (40.5)	78 /252 (31.0)	1292 /3131 (41.3)	/
Isolated distal LE-DVT	274 /3383 (8.1)	17 /252 (6.8)	257 /3131 (8.2)	/
Proximal LE-DVT	1739 /3383 (51.4)	157 /252 (62.3)	1582 /3131 (50.5)	/
Bilateral DVT	151 /3332 (4.5)	15 /248 (6.1)	136 /3084 (4.4)	0.233
Type of PE	/	/	/	0.410
No PE	1713 /2421 (70.8)	117 /169 (69.2)	1596 /2252 (70.9)	/
Subsegmental artery	35 /2421 (1.5)	0 /169 (0.0)	35 /2252 (1.6)	/
Segmental artery	185 /2421 (7.6)	13 /169 (7.7)	172 /2252 (7.6)	/
Lobar artery	229 /2421 (9.5)	15 /169 (8.9)	214 /2252 (9.5)	/
Main pulmonary artery	183 /2421 (7.6)	17 /169 (10.1)	166 /2252 (7.4)	/
Pulmonary artery trunk	76 /2421 (3.1)	7 /169 (4.1)	69 /2252 (3.1)	/
Hemoglobin – g/dL	11.9 ±2.0	12.1 ±1.9	11.9 ±2.0	0.149
Platelets count – G/L	243.7 ±119.4	230.7 ±112.2	244.7 ±120.0	0.071

Anticoagulant therapy after the first 7 days	/	/	/	0.019
LMWH or fondaparinux	2408 /3414 (70.5)	175 /253 (69.2)	2233 /3161 (70.6)	/
VKA or UFH	921 /3414 (27.0)	65 /253 (25.7)	856 /3161 (27.1)	/
Direct oral anticoagulant	85 /3414 (2.5)	13 /253 (5.1)	72 /3161 (2.3)	/
Length of anticoagulant therapy – day	262 ±275	275 ±303	261 ±273	0.679
Major bleeding with anticoagulant therapy	163 /3414 (4.8)	15 /253 (5.9)	148 /3161 (4.7)	0.371
Minor bleeding with anticoagulant therapy	214 /3414 (6.3)	20 /253 (7.9)	194 /3161 (6.1)	0.264
Inferior vena cava filter	/	/	/	<0.001
No vena cava filter	3298 /3396 (97.1)	236 /252 (93.7)	3062 /3144 (97.4)	/
Retrieval before stopping anticoagulant therapy	30 /3396 (0.9)	1 /252 (0.4)	29 /3144 (0.9)	/
Maintained after stopping anticoagulant therapy	68 /3396 (2.0)	15 /252 (6.0)	53 /3144 (1.7)	/
Post-thrombotic syndrome	52 /1418 (3.7)	12 /110 (10.9)	40 /1308 (3.1)	<0.001
Residual venous obstruction after 3 months	163 /1278 (12.8)	19 /89 (21.4)	144 /1189 (12.1)	0.012
Residual pulmonary artery obstruction after 3 months	36 /1968 (1.8)	8 /136 (5.9)	28 /1832 (1.5)	0.002
D-dimers at discontinuation of anticoagulant therapy – µg/L	2573 ±34356	919 ±1361	2775 ±36390	0.005

Data are presented as n (%) or mean \pm standard deviation (SD). BMI: body mass index. DVT: deep vein thrombosis. HR: hazard ratio. eGFR: estimated glomerular filtration rate. LE-DVT: lower extremity deep vein thrombosis. PE: pulmonary embolism. UE-DVT: upper extremity deep vein thrombosis. LMWH: low molecular weight heparin. UFH: unfractionated heparin. VKA: vitamin K antagonist. VTE: venous thromboembolism.

Table 2: Characteristics of the study population at baseline associated with VTE recurrence during the one-year follow-up after the end of anticoagulant therapy (unadjusted Cox analysis)

	Observed case			Multiple imputation by chained equation	
	NA	HR (95%CI)	p	HR (95%CI)	p
Age (for an increase of one year)	0	1.01 (1.00-1.02)	0.072	/	/
Male sex	0	1.15 (0.90-1.47)†	0.275	/	/
Obesity	0	1.26 (0.96-1.65)	0.092	/	/
BMI (for an increase of one kg/m ²)	769	1.01 (0.99-1.04)†	0.282	1.02 (0.99-1.04)	0.136
Chronic kidney disease	97	1.57 (1.19-2.07)†	0.001	1.58 (1.20-2.08)†	0.001
eGFR (for an increase of one mL/min/1.73 m ²)	1065	0.99 (0.99-1.00)	0.002	0.99 (0.99-1.00)	0.002
Chronic heart failure	0	1.24 (0.69-2.21)	0.471	/	/
Chronic lung disease	0	0.89 (0.59-1.36)†	0.595	/	/
Previous VTE	0	1.78 (1.24-2.55)	0.002	/	/
Time between cancer and DVT (for one day increase)	0	1.00 (1.00-1.00)†	0.582	/	/
Site of the cancer	0	/	/	/	/
Oropharynx, larynx, melanoma	/	Ref.	Ref.	/	/
Others, haematological, colorectal, uterus, bladder, kidney, prostate, breast, vulva	/	3.00 (0.96-9.41)	0.059	/	/
Lung, cerebral, stomach, oesophagus, liver, ovary	/	4.31 (1.36-13.71)	0.013	/	/
Pancreas, biliary system, carcinoma of unknown origin	/	9.76 (2.88-32.99)	<0.001	/	/
Metastases	301	1.52 (1.17-1.97)	0.002	1.52 (1.17-1.97)	0.002
Chemotherapy	137	0.95 (0.74-1.22)	0.693	0.95 (0.74-1.21)	0.657
Radiotherapy	190	1.17 (0.84-1.63)	0.358	1.17 (0.84-1.63)	0.364
Immunotherapy	1293	1.29 (0.18-9.24)	0.797	1.01 (0.15-6.74)	0.991
Hormonal therapy	31	0.77 (0.51-1.15)	0.199	0.78 (0.52-1.16)	0.219

Surgery during the 2 months before	0	0.45 (0.30-0.67)	<0.001	/	/
Confined to bed during the 2 months before	0	0.94 (0.66-1.34)	0.731	/	/
Type of venous thromboembolism event	0	/	/	/	/
PE without LE-DVT	/	Ref.	Ref.	/	/
LE-DVT without PE	/	1.27 (0.92-1.76)	0.146	/	/
UE-DVT without PE	/	0.65 (0.38-1.12)	0.123	/	/
PE and DVT	/	1.56 (1.11-2.19)	0.010	/	/
Type of LE-DVT	31	/	/	/	/
No LE-DVT	/	Ref.	Ref.	Ref.	Ref.
Isolated distal LE-DVT	/	1.22 (0.72-2.06)	0.465	1.20 (0.71-2.03)	0.491
Proximal LE-DVT	/	1.66 (1.26-2.18)	<0.001	1.64 (1.25-2.16)	<0.001
Bilateral DVT	82	1.50 (0.89-2.52)	0.131	1.48 (0.88-2.49)	0.140
Type of PE	993	/	/	/	/
No PE	/	Ref.	Ref.	Ref.	Ref.
Subsegmental artery	/	3.38e-14 (0-.)	1.000	0.07 (.)	.
Segmental artery	/	0.92 (0.52-1.63)	0.778	0.97 (0.60-1.57)	0.909
Lobar artery	/	0.84 (0.49-1.43)	0.513	1.02 (0.69-1.52)	0.923
Main pulmonary artery	/	1.35 (0.81-2.24)	0.249	1.34 (0.89-2.03)	0.161
Pulmonary artery trunk	/	1.73 (0.81-3.70)	0.161	1.61 (0.89-2.93)	0.115
Hemoglobin (for an increase of 1 g/dL)	0	1.03 (0.97-1.10)	0.293	/	/
Platelets count (for an increase of 1 G/L)	0	1.00 (1.00-1.00)	0.052	/	/
Anticoagulant therapy after the first 7 days	0	/	/	/	/
LMWH or fondaparinux	/	Ref.	Ref.	/	/
VKA or UFH	/	0.93 (0.70-1.24)	0.638	/	/
Direct oral anticoagulant	/	2.34 (1.33-4.11)	0.003	/	/
Length of anticoagulant therapy (for one day increase)	0	1.00 (1.00-1.00)	0.500	/	/

Major bleeding with anticoagulant therapy	0	1.56 (0.92-2.63)	0.096	/	/
Minor bleeding with anticoagulant therapy	0	1.47 (0.93-2.32)	0.097	/	/
Inferior vena cava filter	18	/	/	/	/
No vena cava filter	/	Ref.	Ref.	Ref.	Ref.
Retrieval before stopping anticoagulant therapy	/	0.36 (0.05-2.56)	0.306	0.36 (0.05-2.55)	0.305
Maintained after stopping anticoagulant therapy	/	3.02 (1.79-5.08)	<0.001	2.98 (1.77-5.02)	<0.001
Post-thrombotic syndrome	1996	2.46 (1.35-4.48)	0.003	2.20 (1.23-3.96)	0.009
Residual venous obstruction after 3 months	2136	1.79 (1.08-2.98)	0.024	1.57 (0.94-2.62)	0.087
Residual pulmonary artery obstruction after 3 months	1446	3.17 (1.55-6.48)	0.002	2.58 (1.45-4.59)	0.001
D-dimers at discontinuation of anticoagulant therapy (for an increase of 1 µg/L)	2817	1.00 (1.00-1.00)	0.837	1.00 (1.00-1.00)	0.804

BMI: body mass index. DVT: deep vein thrombosis. HR: hazard ratio. eGFR: estimated glomerular filtration rate. LE-DVT: lower extremity deep vein thrombosis. NA: not applicable. PE: pulmonary embolism. UE-DVT: upper extremity deep vein thrombosis. LMWH: low molecular weight heparin. UFH: unfractionated heparin. VKA: vitamin K antagonist. VTE: venous thromboembolism. †: Proportional hazard assumption not respected. *: Log-linearity assumption not respected.

Table 3: Characteristics of the study population at baseline associated with VTE recurrence during the one-year follow-up after the end of anticoagulant therapy (multivariate Cox analysis; missing values are imputed by MICE)

	Hazard ratio	95%CI	p
Male sex	1.31†	1.00-1.71	0.050
eGFR	0.99	0.99-1.00	0.008
Previous VTE	1.52	1.03-2.24	0.037
Surgery during the 2 months before	0.50	0.33-0.76	0.001
Anticoagulant therapy after the first 7 days	/	/	/
LMWH or fondaparinux	Ref.	Ref.	Ref.
VKA or UFH	1.02	0.74-1.40	0.889
Direct oral anticoagulant	1.90	0.99-3.66	0.055
Inferior vena cava filter	/	/	/
No vena cava filter	Reference	Reference	Ref.
Retrieval before stopping anticoagulant therapy	0.36	0.05-2.63	0.314
Maintained after stopping anticoagulant therapy	3.06	1.72-5.45	<0.001
Post-thrombotic syndrome (Ginsberg criteria)	2.04	1.05-3.95	0.035
Residual venous obstruction after 3 months	1.66	0.94-2.95	0.082
Residual pulmonary artery obstruction after 3 months	2.48	1.34-4.56	0.004
Time between cancer and VTE (for one day increase)	1.00	1.00-1.00	0.719
Site of the cancer	/	/	/
Oropharynx, larynx, melanoma	Reference	Reference	Ref.
Others, haematological, colorectal, uterus, bladder, kidney, prostate, breast, vulva	3.11	0.97-10.01	0.057
Lung, cerebral, stomach, oesophagus, liver, ovary	4.36	1.33-14.28	0.015
Pancreas, biliary system, carcinoma of unknown origin	10.85	3.02-38.92	<0.001
Metastases	1.34	0.98-1.82	0.064
Chemotherapy	0.82	0.61-1.08	0.160
Radiotherapy	1.19	0.84-1.69	0.335
Immunotherapy	0.88	0.13-5.79	0.890
Hormonal therapy	0.77	0.49-1.19	0.241

95%CI: 95% confidence interval. eGFR: estimated glomerular filtration rate. LMWH: low molecular weight heparin. MICE: multiple imputation by chained equations. UFH: unfractionated heparin. VKA: vitamin K antagonist. VTE: venous thromboembolism. †: Proportional hazard assumption not respected.

Supplemental Table 1: Characteristics of the study population at baseline, compared with patients not included because of extended anticoagulant therapy

Characteristics	Time-limited anticoagulant ≥ 3 months (n = 3,414)	Extended anticoagulant (n = 6,532)	p
Age – year	66.5 \pm 13.2	68.1 \pm 13.0	<0.001
Male sex	1803 /3414 (52.8)	3457 /6532 (52.9)	0.915
Obesity	788 /3414 (23.1)	1382 /6532 (21.2)	0.027
BMI – kg/m ²	27.2 \pm 5.1	27.1 (\pm 5.0)	0.188
Chronic kidney disease	701 /3317 (21.1)	1385 /6331 (21.9)	0.400
eGFR – mL/min/1.73 m ²	74.5 \pm 26.6	71.6 (26.2)	<0.001
Chronic heart failure	146 /3414 (4.3)	368 /6532 (5.6)	0.004
Chronic lung disease	362 /3414 (10.6)	779 /6532 (11.9)	0.049
Previous VTE	323 /3414 (9.5)	926 /6532 (14.2)	<0.001
Time between cancer and DVT – day	616 \pm 1185	717 \pm 1361	0.055
Site of the cancer	/	/	0.003
Oropharynx, larynx, melanoma	118 /3414 (3.5)	148 /6532 (2.3)	/
Others, haematological, colorectal, uterus, bladder, kidney, prostate, breast, vulva	2311 /3414 (67.7)	4430 /6532 (67.8)	/
Lung, cerebral, stomach, oesophagus, liver, ovary	843 /3414 (24.7)	1643 /6532 (25.2)	/
Pancreas, biliary system, carcinoma of unknown origin	142 /3414 (4.2)	311 /6532 (4.8)	/
Metastases	1352 /3113 (43.4)	2630 /5893 (44.6)	0.276
Chemotherapy	1759 /3277 (53.7)	3022 /6109 (49.5)	<0.001
Radiotherapy	476 /3224 (14.8)	740 /5988 (12.4)	0.001
Immunotherapy	11 /2121 (0.5)	19 /3748 (0.5)	0.952
Hormonal therapy	404 /3383 (11.9)	856 /6435 (13.3)	0.055
Surgery during the 2 months before	627 /3414 (18.4)	1067 /6532 (16.3)	0.011
Confined to bed during the 2 months before	516 /3414 (15.1)	936 /6532 (14.3)	0.293
Type of venous thromboembolism event	/	/	<0.001
PE without LE-DVT	931 /3414 (27.3)	1887 /6472 (29.2)	/
LE-DVT without PE	1344 /3414 (39.4)	2428 /6472 (37.5)	/
UE-DVT without PE	369 /3414 (10.8)	425 /6472 (6.6)	/
PE and DVT	770 /3414 (22.3)	1732 /6472 (26.8)	/
Bilateral DVT	151 /3332 (4.5)	416 /6283 (6.6)	<0.001
Hemoglobin – g/dL	11.9 \pm 2.0	12.0 \pm 2.0	<0.001
Platelets count – G/L	243.7 \pm 119.4	242.0 \pm 115.6	0.484

Anticoagulant therapy after the first 7 days	/	/	<0.001
LMWH or fondaparinux	2408 /3414 (70.5)	3922 /6530 (60.1)	/
VKA or UFH	921 /3414 (27.0)	2338 /6530 (35.8)	/
Direct oral anticoagulant	85 /3414 (2.5)	270 /6530 (4.1)	/
Major bleeding with anticoagulant therapy	163 /3414 (4.8)	152 /6532 (2.3)	<0.001
Minor bleeding with anticoagulant therapy	214 /3414 (6.3)	283 /6532 (4.3)	<0.001
Inferior vena cava filter	/	/	0.339
No vena cava filter	3298 /3396 (97.1)	6253 /6465 (96.7)	/
Retrieval before stopping anticoagulant therapy	30 /3396 (0.9)	53 /6465 (0.8)	/
Maintained after stopping anticoagulant therapy	68 /3396 (2.0)	159 /6465 (2.5)	/
Post-thrombotic syndrome	52 /1418 (3.7)	23 /2176 (1.1)	<0.001
Residual venous obstruction after 3 months	163 /1278 (12.8)	242 /2353 (10.3)	0.024
Residual pulmonary artery obstruction after 3 months	36 /1968 (1.8)	50 /3168 (1.6)	0.496
D-dimers at discontinuation of anticoagulant therapy – µg/L	2573 ±34356	92658 ±1288585	0.105

Data are presented as n (%) or mean ± standard deviation (SD). BMI: body mass index. DVT: deep vein thrombosis. HR: hazard ratio. eGFR: estimated glomerular filtration rate. LE-DVT: lower extremity deep vein thrombosis. PE: pulmonary embolism. UE-DVT: upper extremity deep vein thrombosis. LMWH: low molecular weight heparin. UFH: unfractionated heparin. VKA: vitamin K antagonist. VTE: venous thromboembolism.

Supplemental Table 2: Characteristics of the study population at baseline associated with VTE recurrence during the one-year follow-up after the end of anticoagulant therapy (unadjusted Fine and Gray analysis)

	Observed case			Multiple imputation by chained equation	
	NA	sHR (95%CI)	p	sHR (95%CI)	p
Age (for an increase of one year)	0	1.01 (1.00-1.02)	0.162	/	/
Male sex	0	1.09 (0.85-1.39)†	0.494	/	/
Obesity	0	1.39 (1.06-1.82)	0.016	/	/
Body mass index (for an increase of one kg/m ²)	769	1.02 (1.00-1.05)†	0.064	1.03 (1.01-1.05)	0.009
Renal failure	97	1.49 (1.13-1.96)†	0.004	1.50 (1.14-1.97)	0.004
eGFR	1065	0.99 (0.99-1.00)	0.002	0.99 (0.99-1.00)	0.002
Chronic heart failure	0	1.21 (0.67-2.17)	0.532	/	/
Chronic lung disease	0	0.85 (0.55-1.29)†	0.436	/	/
Previous VTE	0	1.68 (1.17-2.42)	0.005	/	/
Time between cancer and DVT (for one day increase)	0	1.00 (1.00-1.00)†	0.537	/	/
Site of the cancer	0	/	/	/	/
Oropharynx, larynx, melanoma	/	Reference	Ref.	/	/
Others, haematological, colorectal, uterus, bladder, kidney, prostate, breast, vulva	/	2.93 (0.93-9.25)	0.066	/	/
Lung, cerebral, stomach, oesophagus, liver, ovary	/	3.34 (1.04-10.69)	0.042	/	/
Pancreas, biliary system, carcinoma of unknown origin	/	5.98 (1.75-20.38)	0.004	/	/
Metastases	301	1.04 (0.80-1.35)	0.766	1.05 (0.81-1.36)	0.701
Chemotherapy	137	0.95 (0.74-1.21)	0.660	0.94 (0.73-1.20)	0.616
Radiotherapy	190	1.19 (0.85-1.67)	0.303	1.19 (0.85-1.66)	0.318
Immunotherapy	1293	1.66 (0.21-13.32)†	0.634	1.10 (0.16-7.56)	0.926
Hormonal therapy	31	0.86 (0.57-1.29)	0.464	0.87 (0.58-1.30)	0.501

Surgery during the 2 months before	0	0.53 (0.36-0.79)	0.002	/	/
Confined to bed during the 2 months before	0	0.85 (0.60-1.21)	0.374	/	/
Type of venous thromboembolism event	0	/	/	/	/
PE without LE-DVT	/	Ref.	Ref.	/	/
LE-DVT without PE	/	1.30 (0.94-1.79)	0.116	/	/
UE-DVT without PE	/	0.74 (0.43-1.27)	0.270	/	/
PE and DVT	/	1.61 (1.15-2.26)	0.006	/	/
Type of LE-DVT	31	/	/	/	/
No LE-DVT	/	Ref.	Ref.	Ref.	Ref.
Isolated distal LE-DVT	/	1.22 (0.72-2.06)	0.462	1.21 (0.71-2.04)	0.483
Proximal LE-DVT	/	1.64 (1.25-2.15)	<0.001	1.63 (1.24-2.14)	<0.001
Bilateral DVT	82	1.31 (0.78-2.20)	0.315	1.29 (0.77-2.18)	0.330
Type of PE	993	/	/	/	/
No PE	/	Ref.	Ref.	Ref.	Ref.
Subsegmental artery	/	3.96e-08 (2.69e-08-5.83e-08)	<0.001	0.19 (0.00-181.36)	0.636
Segmental artery	/	0.94 (0.53-1.66)	0.825	0.99 (0.62-1.61)	0.982
Lobar artery	/	0.83 (0.49-1.41)	0.490	1.01 (0.68-1.50)	0.947
Main pulmonary artery	/	1.23 (0.74-2.04)	0.424	1.27 (0.84-1.91)	0.263
Pulmonary artery trunk	/	1.27 (0.59-2.73)	0.535	1.38 (0.77-2.49)	0.281
Hemoglobin (for an increase of 1 g/dL)	0	1.05 (0.99-1.12)	0.101	/	/
Platelets count (for an increase of 1 G/L)	0	1.00 (1.00-1.00)	0.074	/	/
Anticoagulant therapy after the first 7 days	0	/	/	/	/
LMWH or fondaparinux	/	Ref.	Ref.	/	/
UFH or VKA	/	1.07 (0.80-1.42)	0.654	/	/
Direct oral anticoagulant	/	2.65 (1.53-4.57)	<0.001	/	/
Length of anticoagulant therapy (for one day increase)	0	1.00 (1.00-1.00)	0.541	/	/

Major bleeding with anticoagulant therapy	0	1.32 (0.78-2.23)	0.302	/	/
Minor bleeding with anticoagulant therapy	0	1.29 (0.82-2.03)	0.266	/	/
Inferior vena cava filter	18	/	/	/	/
No vena cava filter	/	Ref.	Ref.	Ref.	Ref.
Retrieval before stopping anticoagulant therapy	/	0.42 (0.06-3.00)†	0.387	0.42 (0.06-2.99)†	0.385
Maintained after stopping anticoagulant therapy	/	2.99 (1.75-5.09)	<0.001	2.96 (1.73-5.04)	<0.001
Post-thrombotic syndrome (Ginsberg criteria)	1996	3.02 (1.67-5.48)	<0.001	2.30 (1.26-4.21)	0.007
Residual venous obstruction after 3 months	2136	2.05 (1.24-3.40)	0.005	1.64 (0.97-2.76)	0.063
Residual pulmonary artery obstruction after 3 months	1446	3.26 (1.58-6.74)	0.001	2.60 (1.45-4.65)	0.001
D-dimers at discontinuation of anticoagulant therapy (for an increase of 1 µg/L)	2817	1.00 (1.00-1.00)	0.366	1.00 (1.00-1.00)	0.782

BMI: body mass index. DVT: deep vein thrombosis. HR: hazard ratio. eGFR: estimated glomerular filtration rate. LE-DVT: lower extremity deep vein thrombosis. NA: not applicable. PE: pulmonary embolism. UE-DVT: upper extremity deep vein thrombosis. LMWH: low molecular weight heparin. UFH: unfractionated heparin. VKA: vitamin K antagonist. VTE: venous thromboembolism. †: Proportional hazard assumption not respected.

Supplemental Table 3: Characteristics of the study population at baseline associated with VTE recurrence during the one-year follow-up after the end of anticoagulant therapy (multivariate Fine and Gray analysis; missing values are imputed by MICE)

	SubHazard ratio	95%CI	p
Male sex	1.18	0.90-1.54	0.228
eGFR	0.99	0.99-1.00	0.011
Previous VTE	1.49	1.01-2.19	0.044
Surgery during the 2 months before	0.58	0.38-0.87	0.009
Anticoagulant therapy after the first 7 days	/	/	/
LMWH or fondaparinux	Ref.	Ref.	Ref.
VKA or UFH	1.06	0.78-1.45	0.698
Direct oral anticoagulant	1.93	1.02-3.66	0.045
Inferior vena cava filter	/	/	/
No vena cava filter	Ref.	Ref.	Ref.
Retrieval before stopping anticoagulant therapy	0.42†	0.06-2.96	0.381
Maintained after stopping anticoagulant therapy	3.03	1.68-5.45	<0.001
Post-thrombotic syndrome (Ginsberg criteria)	2.11	1.06-4.19	0.034
Residual venous obstruction after 3 months	1.68	0.92-3.05	0.088
Residual pulmonary artery obstruction after 3 months	2.68	1.43-5.03	0.002
Time between cancer and DVT (for one day increase)	1.00	1.00-1.00	0.785
Site of the cancer	/	/	/
Oropharynx, larynx, melanoma	Reference	Reference	Ref.
Others, haematological, colorectal, uterus, bladder, kidney, prostate, breast, vulva	2.97	0.92-9.57	0.068
Lung, cerebral, stomach, oesophagus, liver, ovary	3.59	1.09-11.80	0.036
Pancreas, biliary system, carcinoma of unknown origin	6.90	1.92-24.79	0.003
Metastases	0.94	0.70-1.26	0.670
Chemotherapy	0.87	0.66-1.15	0.331
Radiotherapy	1.26	0.89-1.79	0.191
Immunotherapy	0.95	0.14-6.51	0.960
Hormonal therapy	0.80	0.52-1.24	0.318

95%CI: 95% confidence interval. eGFR: estimated glomerular filtration rate. LMWH: low molecular weight heparin. MICE: multiple imputation by chained equations. UFH: unfractionated heparin. VKA: vitamin K antagonist. VTE: venous thromboembolism. †: Proportional hazard assumption not respected.

Figures

Figure 1: Flow chart for identification of eligible patients in RIETE registry

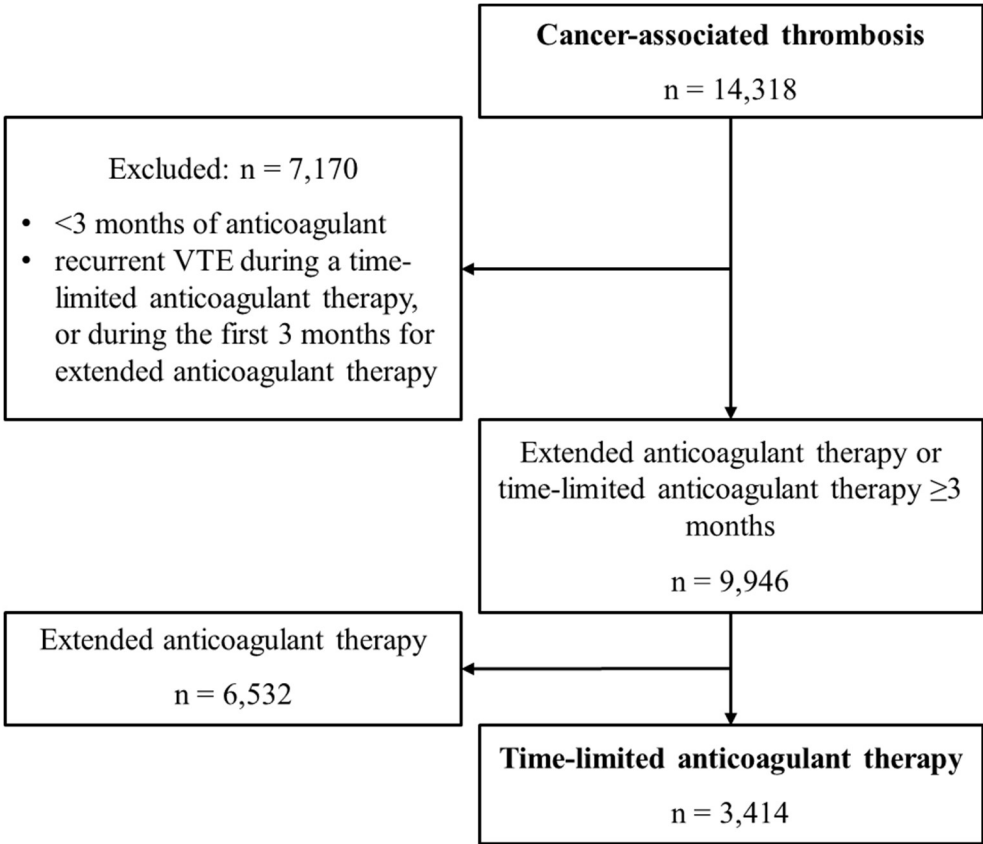
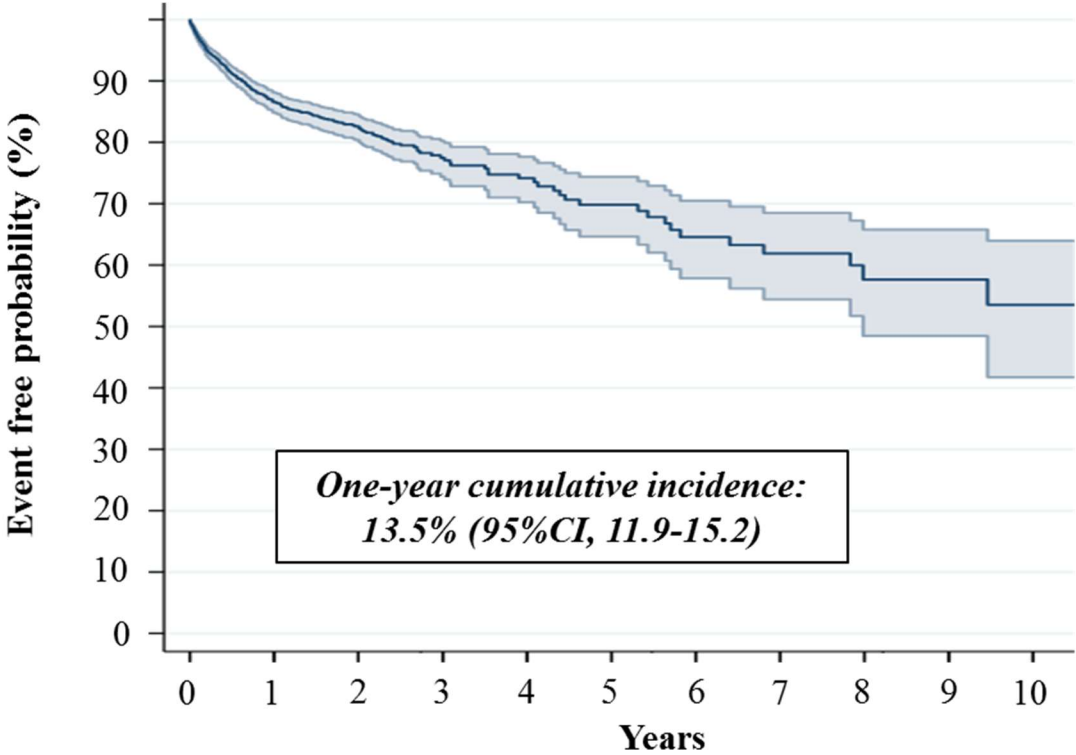


Figure 2: Kaplan-Meier analysis of venous thromboembolism recurrence after the end of a time-limited anticoagulant therapy of at least 3 months for cancer-associated thrombosis



No. at risk	3414	485	122	52	25	8
--------------------	------	-----	-----	----	----	---

3.5.2. Résumé et remarques

Ce travail réalisé sur la cohorte RIETE a été un prérequis nécessaire à la dérivation d'un modèle multivarié à but pronostique estimant le risque de récurrence thromboembolique veineuse à l'arrêt d'un traitement anticoagulant d'une durée d'au moins 3 mois pour une EP ou une TVP associée à un cancer. En effet, la proportion importante de données manquantes rendait impossible l'estimation des coefficients β de régression du modèle, le nombre d'observations complètes étant trop réduit.

Le mécanisme de génération des données manquantes a été évalué, afin de raisonnablement exclure un mécanisme de type MNAR.

Une imputation multiple des données manquantes par équations chaînées a ensuite été réalisée. Dans cette méthode, développée dans les parties précédentes de cette thèse, chaque variable est imputée en utilisant sa propre équation. Ici, un appariement d'après la moyenne prédictive (PMM : *predictive mean matching*) a été préféré pour les variables quantitatives du fait d'un doute sur la normalité de certaines distributions, et un modèle de régression logistique multinomiale a été utilisé pour les variables qualitatives (non ordonnées ou ordonnées). En pratique, un modèle différent pour chaque variable a dû être spécifié, du fait de problèmes de convergence dans les différentes itérations. En effet, les tirages itératifs des différentes distributions conditionnelles ne convergeaient pas vers les tirages de la distribution jointe, malgré l'augmentation du nombre d'itérations, avec des modèles comprenant tous les mêmes variables. Les modèles posant problèmes ont dû être identifiés, puis corrigés un à un. Le choix des covariables à conserver pour chaque modèle d'imputation a été fait sur les données de la littérature. L'évènement d'intérêt (la récurrence thromboembolique veineuse à l'arrêt du traitement) a été conservé dans le modèle d'imputation. Le code utilisé sous STATA pour l'imputation multiple est reproduit ci-dessous.

```
////////// VARIABLES ////////////  
  
codebook popétudiée //population traitée au moins 3 mois mais pour une durée  
//limitée, et sans récurrence lors du traitement  
//anticoagulant  
codebook cancerassoceptvp //EP ou TVP associées au cancer  
codebook ageini //âge à l'inclusion  
codebook sexemasc //sexe: 1=homme, 0=femme
```

```

codebook obesite3 //obésité: 1=oui, 0=non
codebook imc //indice de masse corporelle
codebook peso //poids
codebook frequenc //fréquence cardiaque
codebook tension //pression artérielle systolique
codebook insuffren60v2 //insuffisance rénale: DFG <60 mL/min/1,73 m2 CKD-EPI=1,
//DFG>=60 =2
codebook dfgckd //débit de filtration glomérulaire selon CKD-EPI
codebook insufcard //insuffisance cardiaque
codebook insufresp //insuffisance respiratoire
codebook atcdeptvp //antécédent d'EP et/ou de TVP
codebook delaicancereptvp //délai entre le diagnostic de cancer et le diagnostic
//d'EP/TVP
codebook typecancer4 //localisation du cancer, en 4 classes
codebook metastase3 //métastase
codebook radiotherapie //radiothérapie
codebook chimiotherapie //chimiothérapie
codebook hormonotherapie //hormonothérapie
codebook immunotherapie //immunothérapie
codebook tvpbilat //TVP bilatérale
codebook tvpmbreinftype //localisation de la TVP au membre inférieur: pas TVP=0,
//distale=1, proximale=2
codebook limiteeptype //localisation de l'EP: pas EP=0, tronc de l'artère
//pulmonaire=1, artère pulmonaire=2, lobaire=3,
//segmentaire=4, soussegmentaire=5
codebook localisationVTE3 //type d'évènement thromboembolique veineux: EP sans
//TVP=1, TVP membre inf=2, TVP membre sup sans EP ou TVP
//membre inf=3, EP+TVP=4
codebook chirurgie //chirurgie dans les deux mois précédant
codebook alitement //alitement strict dans les deux mois précédant
codebook ttt_phaseentretient2 //traitement à la phase d'entretien : HBPM ou
//fondaparinux=1, AVK ou HNF=2, AOD=3
codebook etatfiltrefinttt //filtre cave à l'arrêt du traitement anticoagulant:
//1=pas de filtre, 2=filtre présent avant mais retiré à
//l'arrêt, 3=filtre toujours en place
codebook hemorsousttt_maj //hémorragie majeure
codebook hemorsousttt_min //hémorragie mineure
codebook anticoag_duree //durée de l'anticoagulation
codebook syndrpostthromb //présence d'un syndrome postthrombotique à l'arrêt du
//traitement anticoagulant
codebook seqtvpapr3m //séquelle de TVP à l'imagerie à plus de 3 mois de
//l'épisode
codebook seqepapr3m //séquelle d'EP à l'imagerie à plus de 3 mois de
//l'épisode
codebook e_petro //thrombus persistant
codebook ddfinttt //D-dimères en fin de traitement
codebook ana_hemo //hémoglobine à l'inclusion
codebook ana_plaq //plaquettes à l'inclusion
codebook recidive_1an //récidive EP ou TVP dans l'année suivant l'arrêt du
//traitement anticoagulant
codebook outcomeFG_1an //pour les modèles de Fine et Gray: pas récidive=0,
//récidive=1, décès=2
codebook an_c_mol //créatinine en µmol/L
codebook ep_ele_f //fréquence cardiaque
codebook ep_gas_s //saturation en oxygène
codebook varl41 //troponine augmentée
codebook ana_trop //troponine
codebook ana_bnp //BNP
codebook ana_pbnp //pro-BNP
codebook ep_eco_p //pression artérielle pulmonaire
codebook ep_e_dia //diamètre du ventricule droit

```

```

codebook ep_e_ddv //rapport entre le diamètre du VD et celui du VG
codebook sptdolor //douleur à la compression du mollet à 1 an
codebook sptcalam //crampes à 1 an
codebook sptpesad //lourdeur à 1 an
codebook sptpruri //prurit à 1 an
codebook sptpares //paresthésies à 1 an
codebook sptedema //œdème à 1 an
codebook sptincut //induration cutanée à 1 an
codebook spthippi //hyperpigmentation à 1 an
codebook sptcirvc //circulation veineuse collatérale à 1 an
codebook spterit //érythème à 1 an
codebook sptdolpa //douleur à la compression du mollet à 1 an
codebook sptulcve //ulcère veineux à 1 an
codebook sptdol24 //douleur à 2 ans
codebook sptcal24 //crampes à 2 ans
codebook sptpes24 //lourdeur à 2 ans
codebook sptpru24 //prurit à 2 ans
codebook sptpar24 //paresthésies à 2 ans
codebook sptede24 //œdème à 2 an
codebook spticu24 //induration cutanée à 2 ans
codebook spthip24 //hyperpigmentation à 2 ans
codebook sptcvc24 //circulation veineuse collatérale à 2 ans
codebook spteri24 //érythème à 2 ans
codebook sptdpa24 //douleur à la compression du mollet à 2 ans
codebook sptuve24 //ulcère veineux à 2 ans
codebook sptdol36 //douleur à 3 ans
codebook sptcal36 //crampes à 3 ans
codebook sptpes36 //lourdeur à 3 ans
codebook sptpru36 //prurit à 3 ans
codebook sptpar36 //paresthésies à 3 ans
codebook sptede36 //œdème à 3 ans
codebook spticu36 //induration cutanée à 3 ans
codebook spthip36 //hyperpigmentation à 3 ans
codebook sptcvc36 //circulation veineuse collatérale à 3 ans
codebook spteri36 //érythème à 3 ans
codebook sptdpa36 //douleur à la compression du mollet à 3 ans
codebook sptuve36 //ulcère veineux à 3 ans
codebook e_neoact //cancer actif

```

```

////////// DECLARATION DES DONNEES MANQUANTES ET RECHERCHE D'UN MECANISME MNAR
//////////

```

```

mi set mlong //déclaration des données manquantes

```

```

mi misstable summarize recidive_1an outcomeFG_1an ageini sexemasc peso imc obeseite3
an_c_mol dfgckd insuffren60v2 insufcard insufresp atcdeptvp delaicancereptvp
typecancer4 metastase3 radiotherapie chimiotherapie hormonotherapie immunotherapie
tvpbilat tvpmbreinfytype frequenc ep_ele_f tension ep_gas_s varl41 ana_trop ana_bnp
ana_pbnp ep_eco_p ep_e_dia ep_e_ddv limiteeptype localisationVTE3 chirurgie alitement
ttt_phaseentretient2 etatfiltrefinttt hemorsousttt_maj hemorsousttt_min anticoag_duree
sptdolor sptcalam sptpesad sptpruri sptpares sptedema sptincut spthippi sptcirvc
spterit sptdolpa sptulcve sptdol24 sptcal24 sptpes24 sptpru24 sptpar24 sptede24
spticu24 spthip24 sptcvc24 spteri24 sptdpa24 sptuve24 sptdol36 sptcal36 sptpes36
sptpru36 sptpar36 sptede36 spticu36 spthip36 sptcvc36 spteri36 sptdpa36 sptuve36
syndrpostthromb seqtvpapr3m seqepapr3m e_petro ana_hemo ana_plaq ddinttt e_neoact if
popétudiée==1 & cancerassoceptvp==1

```

```

misstable summarize recidive_1an outcomeFG_1an ageini sexemasc peso imc obeseite3
an_c_mol dfgckd insuffren60v2 insufcard insufresp atcdeptvp delaicancereptvp
typecancer4 metastase3 radiotherapie chimiotherapie hormonotherapie immunotherapie

```

```

tvpbilat tvpmbreinfytype frequenc ep_ele_f tension ep_gas_s var141 ana_trop ana_bnp
ana_pbnp ep_eco_p ep_e_dia ep_e_ddv limiteeptype localisationVTE3 chirurgie alitement
ttt_phaseentretient2 etatfiltrefinttt hemorsousttt_maj hemorsousttt_min anticoag_duree
sptdolor sptcalam sptpesad sptpruri sptpares sptedema sptincut spthippi sptcirvc
spterit sptdolpa sptulcve sptdol24 sptcal24 sptpes24 sptpru24 sptpar24 sptede24
spticu24 spthip24 sptcvc24 spteri24 sptdpa24 sptuve24 sptdol36 sptcal36 sptpes36
sptpru36 sptpar36 sptede36 spticu36 spthip36 sptcvc36 spteri36 sptdpa36 sptuve36
syndrpostthromb seqtvpapr3m seqepapr3m e_petro ana_hemo ana_plaq ddfinttt e_neoact if
popétudiée==1 & cancerassoceptvp==1, gen(na_) //pour générer de nouvelles variables
avec 1 si données manquantes et 0 si pas de données manquantes

```

```

mi misstable patterns ageini sexemasc imc dfgckd atcdeptvp delaicancereptvp typecancer4
metastase3 radiotherapie chimiotherapie hormonotherapie immunotherapie tvpbilat
tvpmbreinfytype limiteeptype localisationVTE3 chirurgie alitement ttt_phaseentretient2
etatfiltrefinttt anticoag_duree syndrpostthromb seqtvpapr3m seqepapr3m ana_hemo
ana_plaq ddfinttt e_neoact if popétudiée==1 & cancerassoceptvp==1

```

```

mi misstable nested ageini sexemasc imc dfgckd atcdeptvp delaicancereptvp typecancer4
metastase3 radiotherapie chimiotherapie hormonotherapie immunotherapie tvpbilat
tvpmbreinfytype limiteeptype localisationVTE3 chirurgie alitement ttt_phaseentretient2
etatfiltrefinttt anticoag_duree syndrpostthromb seqtvpapr3m seqepapr3m ana_hemo
ana_plaq ddfinttt e_neoact if popétudiée==1 & cancerassoceptvp==1 //les patterns de
données NA ne peuvent pas être résumés en une ligne, donc elles ne sont pas monotones
(définition de données NA monotones: if Variable J is missing then Variable K is also
missing for all K>J)

```

```

mi misstable tree ageini sexemasc imc dfgckd atcdeptvp delaicancereptvp typecancer4
metastase3 radiotherapie chimiotherapie hormonotherapie immunotherapie tvpbilat
tvpmbreinfytype limiteeptype localisationVTE3 chirurgie alitement ttt_phaseentretient2
etatfiltrefinttt anticoag_duree syndrpostthromb seqtvpapr3m seqepapr3m ana_hemo
ana_plaq ddfinttt e_neoact if popétudiée==1 & cancerassoceptvp==1, frequency //pour voir
les liens entre données NA et non NA selon un arbre

```

```

mi register imputed imc an_c_mol dfgckd insuffren60v2 metastase3 radiotherapie
chimiotherapie hormonotherapie immunotherapie tvpbilat tvpmbreinfytype frequenc
ep_ele_f tension ep_gas_s var141 ana_trop ana_bnp ana_pbnp ep_eco_p ep_e_dia ep_e_ddv
limiteeptype etatfiltrefinttt sptdolor sptcalam sptpesad sptpruri sptpares sptedema
sptincut spthippi sptcirvc spterit sptdolpa sptulcve sptdol24 sptcal24 sptpes24
sptpru24 sptpar24 sptede24 spticu24 spthip24 sptcvc24 spteri24 sptdpa24 sptuve24
sptdol36 sptcal36 sptpes36 sptpru36 sptpar36 sptede36 spticu36 spthip36 sptcvc36
spteri36 sptdpa36 sptuve36 syndrpostthromb seqtvpapr3m seqepapr3m e_petro ddfinttt
e_neoact //pour déclarer les données qui devront être imputées; certaines de ces
variables sont des variables auxiliaires (mais avec des données manquantes donc
enregistrées en imputed), c'est à dire soit associées à des données manquantes, soit
associées au fait d'être NA, et on les gardera dans l'imputation car permet d'améliorer
l'imputation des données manquantes, et on peut les identifier selon les données de la
littérature (fait comme ça ici) ou à l'aide des coefficients de corrélation (r>0.4)

```

```

mi register regular recidive_lan outcomeFG_lan dateorigine recidive dureesuivi_recidive
dureeavantrecidive recidive_EPsansTVP recidive_EPavecTVP recidive_TVPsansEP deces
datedeces dureesuivi_deces datepoint dureesuivi_lan deces_lan dureesuivi_deceslan
outcomeFG ageini sexemasc peso obesite3 insufcard insufresp atcdeptvp delaicancereptvp
typecancer4 localisationVTE3 chirurgie alitement ttt_phaseentretient2 hemorsousttt_maj
hemorsousttt_min anticoag_duree ana_hemo ana_plaq //pour déclarer les
variables qui ne devront pas être imputées car sans données manquantes

```

```

mi describe //pour vérifier comment les variables sont enregistrées

```

```

mi varying //pour vérifier s'il y a des "varying and super-varying
variables" (variables dont les valeurs dans les observations incomplètes diffèrent
selon les itérations: par exemple les variables imputées) et si elles sont toutes

```

enregistrées correctement (soit "imputed" soit "passive"); toutes les variables imputées doivent être "varying"

```
codebook recidive_1an outcomeFG_1an ageini sexemasc peso imc obesite3 an_c_mol dfgckd
insuffren60v2 insufcard insufresp atcdeptvp delaicancereptvp typecancer4 metastase3
radiotherapie chimiotherapie hormonotherapie immunotherapie tvpbilat tvpmbreinfytype
frecuenc ep_ele_f tension ep_gas_s varl41 ana_trop ana_bnp ana_pbnp ep_eco_p ep_e_dia
ep_e_ddv limiteeptype localisationVTE3 chirurgie alitement ttt_phaseentretient2
etatfiltrefinttt hemorsousttt_maj hemorsousttt_min anticoag_duree sptdolor sptcalam
sptpesad sptpruri sptpare sptedema sptincut spthippi sptcirvc spteri sptdolpa
sptulcve sptdol24 sptcal24 sptpes24 sptpru24 sptpar24 sptede24 spticu24 spthip24
sptcvc24 spteri24 sptdpa24 sptuve24 sptdol36 sptcal36 sptpes36 sptpru36 sptpar36
sptede36 spticu36 spthip36 sptcvc36 spteri36 sptdpa36 sptuve36 syndrpostthromb
seqtvpapr3m seqepapr3m e_petro ana_hemo ana_plaq ddfinttt e_neoact if popétudiée==1 &
cancerassoceptvp==1 //pour voir les données avec NA et sans NA, et leur type
(qualitative, quantitative etc) pour aider à la constitution du modèle d'imputation
```

```
mi impute chained (pmm, knn(10) omit(i.tvpbilat i.tvpmbreinfytype frecuenc
i.limiteeptype i.etatfiltrefinttt i.seqepapr3m delaicancereptvp i.localisationVTE3
ana_plaq)) imc (pmm, knn(10) omit(i.tvpbilat i.tvpmbreinfytype i.limiteeptype
i.syndrpostthromb i.seqtvpapr3m i.seqepapr3m i.e_petro delaicancereptvp
i.localisationVTE3 ana_plaq)) dfgckd (logit, omit(dfgckd i.tvpbilat i.tvpmbreinfytype
i.limiteeptype i.etatfiltrefinttt i.syndrpostthromb i.seqtvpapr3m i.seqepapr3m
i.e_petro delaicancereptvp i.typecancer4 i.localisationVTE3 ana_plaq i.outcomeFG_1an
frecuenc i.metastase3 i.radiotherapie i.immunotherapie imc i.obesite3)) insuffren60v2
(logit, omit(frecuenc tension i.tvpmbreinfytype i.limiteeptype i.syndrpostthromb
i.seqepapr3m i.seqtvpapr3m i.e_petro i.localisationVTE3)) metastase3 (logit,
omit(i.tvpbilat i.tvpmbreinfytype frecuenc tension i.limiteeptype i.syndrpostthromb
i.seqepapr3m i.seqtvpapr3m i.e_petro i.localisationVTE3)) radiotherapie (logit,
omit(i.tvpbilat i.tvpmbreinfytype frecuenc tension i.limiteeptype i.syndrpostthromb
i.seqepapr3m i.seqtvpapr3m i.e_petro i.localisationVTE3)) chimiotherapie (logit,
omit(i.tvpbilat i.tvpmbreinfytype frecuenc tension i.limiteeptype i.syndrpostthromb
i.seqepapr3m i.seqtvpapr3m i.e_petro i.localisationVTE3)) hormonotherapie (logit,
omit(i.tvpbilat i.tvpmbreinfytype frecuenc tension i.limiteeptype i.syndrpostthromb
i.seqepapr3m i.seqtvpapr3m i.e_petro i.localisationVTE3 i.etatfiltrefinttt dfgckd
i.outcomeFG_1an i.atcdeptvp ddfinttt peso)) immunotherapie (logit, omit(dfgckd
i.insuffren60v2 i.radiotherapie i.chimiotherapie i.immunotherapie i.hormonotherapie
frecuenc tension ana_hemo ana_plaq i.outcomeFG_1an i.etatfiltrefinttt i.tvpmbreinfytype
i.e_petro imc ddfinttt peso)) tvpbilat (ologit, omit(dfgckd i.insuffren60v2
i.radiotherapie i.chimiotherapie i.immunotherapie i.hormonotherapie frecuenc tension
ana_hemo ana_plaq)) tvpmbreinfytype (pmm, knn(10) omit(i.metastase3 i.radiotherapie
i.hormonotherapie i.immunotherapie i.tvpbilat i.tvpmbreinfytype i.etatfiltrefinttt
i.syndrpostthromb i.seqepapr3m i.seqtvpapr3m i.e_petro i.atcdeptvp delaicancereptvp
i.typecancer4)) frecuenc (pmm, knn(10) omit(i.metastase3 i.radiotherapie
i.hormonotherapie i.immunotherapie i.tvpbilat i.tvpmbreinfytype i.etatfiltrefinttt
i.syndrpostthromb i.seqepapr3m i.seqtvpapr3m i.e_petro i.atcdeptvp delaicancereptvp
i.typecancer4)) tension (ologit, omit(i.metastase3 i.radiotherapie i.chimiotherapie
i.hormonotherapie i.immunotherapie)) limiteeptype (ologit, omit(i.outcomeFG_1an
frecuenc tension_ dfgckd ddfinttt i.immunotherapie i.hormonotherapie i.radiotherapie
i.tvpbilat i.e_petro i.obesite3 peso i.tvpmbreinfytype i.syndrpostthromb
i.typecancer4)) etatfiltrefinttt (logit, omit(frecuenc tension i.etatfiltrefinttt
i.limiteeptype i.immunotherapie i.outcomeFG_1an i.typecancer4 i.hemorsousttt_min
i.e_petro)) syndrpostthromb (logit, omit(frecuenc tension i.radiotherapie
i.immunotherapie i.hormonotherapie i.tvpbilat dfgckd ddfinttt i.outcomeFG_1an
i.e_petro)) seqtvpapr3m (logit, omit(i.etatfiltrefinttt i.hormonotherapie
i.chimiotherapie i.radiotherapie i.immunotherapie i.tvpmbreinfytype i.tvpbilat
i.insuffren60v2 i.typecancer4 i.limiteeptype)) seqepapr3m (logit, omit(i.seqepapr3m
i.seqtvpapr3m i.outcomeFG_1an i.hormonotherapie i.immunotherapie ddfinttt tension_
frecuenc dfgckd)) e_petro (pmm, knn(10)) ddfinttt = i.recidive_1an i.outcomeFG_1an
ageini i.sexemasc peso i.obesite3 i.insufcard i.insufresp i.atcdeptvp delaicancereptvp
i.typecancer4 i.localisationVTE3 i.chirurgie i.alitement i.ttt_phaseentretient2
```

```
i.hemorsousttt_maj i.hemorsousttt_min anticoag_duree ana_hemo ana_plaq if popétudiée==1
& cancerassoceptvp==1, add(100) rseed(53421) savetrace(trace2, replace detail) augment
noisily
```

Deux modèles multivariés ont ensuite été constitués, l'un à partir d'un modèle de Cox, l'autre à partir d'un modèle de Fine et Gray. Les variables ont été sélectionnées en deux temps, sur les données de la littérature puis à l'aide d'une sélection descendante utilisant les estimations des coefficients β dans chaque base de données imputées et combinées selon les règles de Rubin ainsi qu'un seuil de *p-value* $<0,157$. Au final, les mêmes variables ont été incluses dans ces deux modèles. Leurs résultats sont globalement superposables. On confirme ainsi que le cancer du pancréas et des voies biliaires, les antécédents d'EP ou de TVP, l'insuffisance rénale, le syndrome post-thrombotique, les séquelles d'embolie pulmonaire à l'imagerie et le maintien d'un filtre cave à l'arrêt du traitement anticoagulant sont associés à la récurrence thromboembolique veineuse, alors que le fait que l'évènement initial ait été causé par une chirurgie diminue par deux la récurrence.

Ce travail a été présenté à la 10^{ème} *International Conference on Thrombosis and Hemostasis Issues in Cancer (ICTHIC)* en 2021 ; l'écriture de l'article est en cours de finalisation, en vue d'une soumission en décembre 2021.

3.6. Validation interne d'un modèle prédictif par *bootstrap*

3.6.1. Présentation du modèle multivarié

Un modèle multivarié de type prédictif pour le risque de récurrence des ETEV dans l'année suivant l'arrêt d'un traitement anticoagulant d'au moins 3 mois pour un évènement thromboembolique veineux associé au cancer a été développé, à partir d'un modèle de Cox, et avec une sélection des variables guidées sur les données de la littérature et les résultats de l'analyse précédente. Ce travail a été réalisé sur les 100 jeux de données imputées, en plus du jeu de données original. Le tableau 1 présente les résultats de l'estimation des coefficients β de régression du modèle.

Tableau 1 : Coefficients de régression β (avec IC95%) des variables du modèle multivarié de prédiction du risque de récurrence des événements thromboemboliques veineux.

Prédicteurs	Coefficient β	Erreur-type	p	IC95%
Localisation du cancer				
Oropharynx, pharynx, mélanome (référence)	0	/	/	/
Colorectal, utérus, vessie, rein, prostate, sein, vulve, hématologique, autre	0,973	0,587	0,098	- 0,179 à 2,124
Poumon, cerveau, estomac, œsophage, foie, ovaire	1,318	0,595	0,027	0,151 à 2,485
Pancréas, voies biliaires, origine inconnue	2,232	0,634	<0,001	0,989 à 3,475
Métastase	0,249	0,148	0,092	- 0,041 à 0,540
Cancer sous traitement au moment de la thrombose (chimiothérapie, immunothérapie, hormonothérapie ou radiothérapie)	- 0,115	0,145	0,430	- 0,399 à 0,170
Chirurgie dans les 2 mois précédant la thrombose	- 0,706	0,212	0,001	- 1,121 à - 0,291
Sexe masculin	0,287*	0,136	0,035	0,021 à 0,554
Antécédent d'EP ou de TVP	0,420	0,196	0,032	0,037 à 0,804
DFG, en mL/min/1,73 m ²	- 0,007	0,003	0,009	- 0,013 à - 0,002
Filtre cave en place à l'arrêt du traitement anticoagulant	1,007	0,281	<0,001	0,457 à 1,557
Syndrome post-thrombotique à l'arrêt du traitement anticoagulant	0,796	0,316	0,013	0,172 à 1,419
Séquelle d'embolie pulmonaire à l'imagerie à l'arrêt du traitement anticoagulant	0,941	0,297	0,002	0,355 à 1,527

* Hypothèse des risques proportionnels non respectée.

Un modèle de Cox a été préféré à un modèle de Fine et Gray, bien que le décès apparaisse comme un risque compétitif vis-à-vis de la récurrence thromboembolique veineuse, du fait de résultats globalement superposables dans le travail réalisé sur l'identification des facteurs

associés à la récurrence, et surtout du fait d'un temps de calcul fortement augmenté, en particulier dans ce contexte d'analyse de données manquantes par imputation multiple et de *bootstrap*.

Les coefficients β du modèle multivarié présentés dans le tableau 1 ont donc été estimés sur la base de données initiales (avec imputation des données manquantes).

La probabilité \hat{P} de survenue de l'évènement d'intérêt après t années de suivi pour un patient donné est définie par $\hat{P} = 1 - S_0(t)^{\exp(\sum_{i=1}^p \beta_i X_i - \sum_{i=1}^p \beta_i \bar{X}_i)}$, avec :

- $S_0(t)$ la survie de base à t ans, donc pour notre modèle, la survie globale sans évènement à 1 an dans la cohorte, estimée par la méthode de Kaplan-Meier,
- X_i la valeur du prédicteur pour le patient étudié,
- \bar{X}_i la valeur moyenne dans la cohorte pour une variable quantitative, ou la proportion dans la cohorte pour une variable qualitative,
- β_i le coefficient de régression estimé,
- p le nombre de prédicteurs dans le modèle.

Dans notre base, on a $S_0(t) = 0,8655$.

L'équation de l'indice pronostique, exprimée avec les coefficients β du tableau 1, est : $\sum_{i=1}^p \beta_i X_i = 0$ (si oropharynx, pharynx ou mélanome) + 0,973 (si colorectal, utérus, vessie, rein, prostate, sein, vulve, hématologique ou autre) + 1,318 (si poumon, cerveau, estomac, œsophage, foie ou ovaire) + 2,232 (si pancréas, voies biliaires ou origine inconnue) + 0,249 (si métastase) – 0,115 (si cancer sous traitement lors de la thrombose) – 0,706 (si chirurgie dans les 2 mois précédant la thrombose) + 0,287 (si sexe masculin) + 0,420 (si antécédent d'EP ou de TVP) – 0,007xDFG + 1,007 (si filtre cave en place à l'arrêt de l'anticoagulant) + 0,796 (si syndrome post-thrombotique) + 0,941 (si séquelle d'EP à l'imagerie).

Pour un hypothétique patient « moyen », chaque coefficient β est multiplié par la proportion de sa variable associée si celle-ci est qualitative ou par la moyenne si la variable associée est quantitative, et l'ensemble est additionné. On a donc : $\sum_{i=1}^p \beta_i \bar{X}_i = (0 \times 0,0346) + (0,973 \times 0,6769) + (1,318 \times 0,2469) + (2,232 \times 0,0416) + (0,249 \times 0,4321) - (0,115 \times 0,6471) - (0,706 \times 0,1837) + (0,287 \times 0,5281) + (0,420 \times 0,0946) - (0,007 \times 77,9141) + (1,007 \times 0,0203) + (0,796 \times 0,0828) + (0,941 \times 0,0710) = 0,7591$.

Ainsi, pour un patient donné, sa probabilité \hat{P} de récurrence dans l'année suivant l'arrêt du traitement anticoagulant est : $\hat{P} = 1 - 0,8655^{\exp(\sum_{i=1}^p \beta_i X_i - 0,7591)}$.

3.6.2. Estimation de la performance apparente du modèle multivarié (performance sur la base de données initiales)

La performance de ce modèle a été évaluée sur la cohorte de dérivation, à partir des données imputées : il s'agit de la performance apparente. Comme détaillé dans la partie 2.6.2.2, chaque estimateur de la performance a été estimé dans chacun des 100 jeux de données imputées, aboutissant ainsi à 100 variantes d'un même paramètre, et ces variantes ont ensuite été combinées selon les règles de Rubin pour fournir un estimateur unique.

La calibration du modèle a été évaluée grâce à la droite avec la probabilité prédite par le modèle de la survenue de l'évènement sur l'axe des X, et la fréquence effectivement observée de l'évènement sur l'axe des Y. Le calcul de la pente est donné, devant être égal à 1 ; un test d'hypothèse avec H_0 la valeur de la pente égale à 1 a aussi été réalisé.

Le code utilisé sous STATA est présenté ci-dessous :

```
mi xeq 0: stcoxcal xb_apparperf, times(365) nograph test
return list
scalar gamma1 = r(gamma1)
scalar gamma1_se = r(gamma1_se)
```

On remarque que la pente de la droite et son erreur-type ont été stockées sous forme de scalaires ; cela sera utilisé plus tard pour sauvegarder les résultats de l'estimation des paramètres dans les différents échantillons *bootstrap* pour la validation interne, mais n'est pas réellement utile à ce stade.

La pente de la droite de calibration était ici de 1,1335, avec une erreur-type de 0,1218, donc un IC95% de 0,8948 à 1,3722, et un p-value de 0,273 ne permettant pas de rejeter l'hypothèse H_0 définie comme l'absence de différence entre la valeur de la pente et 1. La calibration est ici correcte.

La discrimination du modèle a donc été évaluée par la statistique D de Royston et Sauerbrei (*prognostic separation D-statistic*). Enfin, le R_D^2 de Royston et Sauerbrei, basé sur leur statistique D, a été calculé.

Le code utilisé sous STATA pour le calcul de la statistique D et le R_D^2 est présenté ci-dessous :

```

scalar r2 = 0 //pour définir le contenu de départ de la variable scalaire r2
scalar r2se = 0 //pour définir le contenu de départ de la variable scalaire r2se
scalar D = 0 //pour définir le contenu de départ de la variable scalaire D
scalar sD = 0 //pour définir le contenu de départ de la variable scalaire sD
xi: mi xeq 1/100 : str2d stcox i.typecancer4 i.metastase3 i.tttcancer i.sexemasc dfgckd
i.atcdeptvp i.chirurgie i.filtreaDO i.syndrpostthromb i.seqepapr3m ; scalar r2 = r2 +
r(r2) ; scalar r2se = r2se + r(r2se) ; scalar D = D + r(D) ; scalar sD = sD + r(sD)
//lance le calcul de R^2 et de D dans chacun des jeux de données imputées (ici, de
1 à 100) et accumule les valeurs de R^2 dans la variable scalaire r2 et les valeurs de
l'erreur-type de R^2 dans la variable scalaire r2se, et idem pour D
scalar r2 = r2/100 //pour obtenir la moyenne, on divise r2 par le nombre
d'imputations (ici, 100)
scalar r2se = r2se/100 //idem pour la SE de r2
scalar D = D/100 //idem pour D
scalar sD = sD/100 //idem pour la SE de D
di as text "R^2 calculé dans les données imputés = " as res r2
di as text "Standard error de R^2 calculée dans les données imputées = " as res r2se
di as text "D calculé dans les données imputés = " as res D
di as text "Standard error de D calculée dans les données imputés = " as res sD
scalar list

postfile parametres gamma1 gamma1_se rcarre rcarre_se droyston droyston_se using
validation_interne, replace //on crée une base "validation_interne" où sauvegarder
les estimations des paramètres de la validation interne des différents modèles

post parametres (gamma1) (gamma1_se) (r2) (r2se) (D) (sD) //pour stocker dans
"validation_interne" les scalaires gamma1 et gamma1_se, r2 et r2se, D et sD, sur une
même ligne correspondant à la population initiale (échantillon 0)

```

Une difficulté était l'absence d'implémentation de la méthode d'estimation de ces paramètres sur des données imputées dans STATA. Ces paramètres et leur erreur-type ont donc été estimés sur chacun des jeux de données imputés (jeux 1 à 100), stockés ensuite sous forme de scalaires, puis combinés manuellement selon les règles de Rubin. Ces estimations sont alors sauvegardées dans une nouvelle base, pour pouvoir ensuite être manipulées.

Au final, les résultats étaient :

- pour la statistique D de Royston et Sauerbrei : 1,1431, avec une erreur-type de 0,1011,
- pour le R_D^2 de Royston et Sauerbrei : 0,2378, avec une erreur-type de 0,0319.

3.6.3. Tirage au sort avec remplacement des échantillons *bootstrap*

A partir de la base de données initiales, des échantillons *bootstrap* ont été constitués. Plusieurs difficultés sont apparues. Les données manquantes ayant été imputées avec constitution de 100 jeux de données complètes, chaque observation tirée au sort devait en fait

être prélevée dans chacun des 100 jeux de données imputées. Cette fonctionnalité de *bootstrap* sur des données imputées n'étant pas implémentée dans STATA, pour permettre ce tirage et ne pas considérer les 100 copies d'une même observation comme 100 observations différentes dans l'échantillon *bootstrap*, les 100 copies d'une même observation ont été considérées comme un *cluster*, et il a été réalisé un tirage au sort des *clusters* et non pas des observations individuelles. La taille de chaque échantillon *bootstrap* a également posé problème. Comme évoqué dans la partie 2.6.2.1, certains auteurs proposent une taille identique à celle de la base de données initiales, donc dans notre cas 3414 patients. Pour mémoire, un échantillon *bootstrap* étant constitué par tirage au sort avec remplacement, le même patient peut se retrouver plusieurs fois dans un échantillon *bootstrap* donné, donc même s'ils comprennent le même nombre d'observations (3414), l'échantillon *bootstrap* et la base de données initiales ne sont pas identiques. Un nombre d'échantillons *bootstrap* minimal de 100 étant proposé pour la validation interne, comme discuté dans la partie 2.6.2.1, l'ensemble des échantillons *bootstrap* contiendrait donc 341 400 observations, chacune en 100 exemplaires du fait de l'imputation multiple des données manquantes, donc 34 140 000 observations au total. Cela pose un problème de temps de calcul pour l'analyse ultérieure. Le choix a donc été fait de restreindre chaque échantillon *bootstrap* à 1000 observations (et chacune en 100 exemplaires, du fait de l'imputation multiple). La variable « *_mi_id* » permet d'identifier une observation donnée dans les 100 jeux de données imputées et dans le jeu de données initial ; le problème pour les calculs ultérieurs est que cette variable identifiait parfois plusieurs observations dans l'échantillon *bootstrap*, un même patient pouvant avoir été tiré plusieurs fois. Il a donc fallu renommer les patients tirés plusieurs fois dans l'échantillon.

Le code utilisé sous STATA pour la constitution des échantillons *bootstrap* et la correction de la variable « *_mi_id* » est présenté ci-dessous :

```

bsample 1000, cluster(_mi_id)

duplicates tag _mi_id _mi_m, gen(obsduplic)           //pour créer une variable
identifiant les observations dupliquées (grâce à _mi_id et _mi_m): 0 si pas dupliquée,
1 si dupliquées 1 fois, 2 si dupliquées 2 fois etc
sort _mi_id (_mi_m)                                 //pour classer par _mi_id, et dans chaque _mi_id par _mi_m
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 3414) if _n == 2 //la
première réplification, qui vient en 2ème position (_n==2) après l'observation originale,
est re-numérotée à partir de 3415
sort _mi_id (_mi_m)                                 //reclasse avec la nouvelle numérotation; la 2ème
réplification est maintenant en 2ème position après l'observation originale
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 6828) if _n == 2 //la 2ème
réplification, qui est maintenant en 2ème position après l'observation originale, est re-
numérotée à partir de 6829
sort _mi_id (_mi_m)

```

```

by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 10242) if _n == 2 //en
cas de 3ème réplification
sort _mi_id (_mi_m)
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 13656) if _n == 2 //en
cas de 4ème réplification
sort _mi_id (_mi_m)
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 17070) if _n == 2 //en
cas de 5ème réplification
sort _mi_id (_mi_m)

save basebootstrap, replace //pour sauvegarder l'échantillon bootstrap dans une
nouvelle base

```

3.6.4. Evaluation de la performance du modèle multivarié sur l'échantillon *bootstrap* (*bootstrap performance*)

Le modèle multivarié a ensuite été évalué sur l'échantillon *bootstrap*. Comme précisé dans la partie 2.6.2.1, les étapes de sélection des variables du modèle doivent avant cela être reproduite dans l'échantillon *bootstrap*. Cela n'a pas été nécessaire ici, la sélection des variables ayant été réalisée sur les données de la littérature et celles de l'analyse précédente.

Les coefficients β du modèle multivarié ont donc été estimés sur l'échantillon *bootstrap*, puis utilisés pour évaluer la calibration (pente de la droite), la discrimination (statistique D de Royston et Sauerbrei) et le R_D^2 de Royston et Sauerbrei.

Le code utilisé sous STATA pour l'estimation des coefficients β du modèle multivarié dans l'échantillon *bootstrap* et l'évaluation de la performance dans ce même échantillon est présenté ci-dessous :

```

use basebootstrap.dta, clear //pour charger l'échantillon bootstrap

xi: mi estimate, saving(coefficients_bsperf, replace): stcox i.typecancer4 i.metastase3
i.tttcancer i.sexemasc dfgckd i.atcdeptvp i.chirurgie i.filtreaDO i.syndrpostthromb
i.seqepapr3m

mi predict xb_bsperf using coefficients_bsperf, xb
mi xeq 0: stcoxcal xb_bsperf, times(365) nograph test
return list
scalar gamma1 = r(gamma1)
scalar gamma1_se = r(gamma1_se)

scalar r2 = 0

```

```

scalar r2se = 0
scalar D = 0
scalar sD = 0
xi: mi xeq 1/100 : str2d stcox i.typecancer4 i.metastase3 i.tttcancer i.sexemasc dfgckd
i.atcdeptvp i.chirurgie i.filtreaD0 i.syndrpostthromb i.seqepapr3m ; scalar r2 = r2 +
r(r2) ; scalar r2se = r2se + r(r2se) ; scalar D = D + r(D) ; scalar sD = sD + r(sD)
scalar r2 = r2/100
scalar r2se = r2se/100
scalar D = D/100
scalar sD = sD/100
di as text "R^2 calculé dans les données imputés = " as res r2
di as text "Standard error de R^2 calculée dans les données imputées = " as res r2se
di as text "D calculé dans les données imputés = " as res D
di as text "Standard error de D calculée dans les données imputés = " as res sD
scalar list

post parametres (gamma1) (gamma1_se) (r2) (r2se) (D) (sD) //pour stocker dans
"validation_interne" les paramètres estimés et leurs erreurs-type

```

3.6.5. Evaluation de la performance du modèle multivarié avec coefficients β estimés sur l'échantillon *bootstrap*, sur la base de données initiale (*test performance*)

Le modèle multivarié a ensuite été évalué dans la base de données initiale, mais en utilisant les estimations des coefficients β obtenues dans l'échantillon *bootstrap* et non pas dans la base de données initiale. La commande calculant la statistique D et le R_D^2 de Royston et Sauerbrei ne permettant pas d'utiliser les coefficients β évalués dans une autre base, une fusion de la base contenant les données initiales avec imputation des données manquantes, et de la base contenant l'échantillon *bootstrap* avec imputation des données manquantes a été nécessaire, avec création d'une variable identifiant les deux bases, puis utilisation d'un option de la commande « *str2d* » normalement utilisée pour la validation interne selon une méthode *split-sample*, qui permet de dériver un modèle dans un sous-ensemble de la base et de le tester dans un autre sous-ensemble de la même base. Dans notre cas, on considère, dans la base fusionnée, que la base *bootstrap* est celle où les coefficients β sont estimés, et la base initiale celle où la validation est réalisée.

Le code utilisé sous STATA pour l'évaluation, sur la base de données initiales, de la performance du modèle estimé sur l'échantillon *bootstrap* est présenté ci-dessous :

```

use RIETEimputations2b.dta, clear           //pour charger la base initiale

mi predict xb_testperf using coefficients_bsperf, xb //on utilise les coefficients
stockés dans coefficients_bsperf, qui sont issus du modèle appliqué sur l'échantillon
bootstrap
mi xeq 0: stcoxcal xb_testperf, times(365) nograph test
return list
scalar gamma1 = r(gamma1)
scalar gamma1_se = r(gamma1_se)

gen testdata =.
replace testdata = 1                       //pour identifier ensuite la base de données
où tester le modèle issu du bootstrap, quand on aura fusionné les bases avec la
commande ci-dessous

append using basebootstrap                 //on rattache la base de données de
l'échantillon bootstrap à la base de données initiale, ce qui est nécessaire ensuite
pour str2d
replace testdata = 0 if testdata=.. //pour identifier la base de données où estimer les
coefficients du modèle
codebook testdata

duplicates tag _mi_id _mi_m, generate(obsduplic2)
by _mi_id _mi_m obsduplic2, sort: replace _mi_id = (_mi_id + 20000) if _n == 2 //pour
renuméroter les observations identiques entre l'échantillon bootstrap et la base
initiale
sort _mi_id (_mi_m)

scalar r2 = 0
scalar r2se = 0
scalar D = 0
scalar sD = 0
xi: mi xeq 1/100 : str2d stcox i.typecancer4 i.metastase3 i.tttcancer i.sexemasc dfgckd
i.atcdeptvp i.chirurgie i.filtreaD0 i.syndrpostthromb i.seqepapr3m, validate(testdata)
; scalar r2 = (r2 + r(r2)) ; scalar r2se = (r2se + r(r2se)) ; scalar D = (D + r(D)) ;
scalar sD = (sD + r(sD))
scalar r2 = r2/100
scalar r2se = r2se/100
scalar D = D/100
scalar sD = sD/100
di as text "R^2 calculé dans la base initiale (test performance) à partir du modèle
évalué dans l'échantillon bootstrap = " as res r2
di as text "Standard error de R^2 calculée dans la base initiale (test performance) à
partir du modèle évalué dans l'échantillon bootstrap = " as res r2se
di as text "D calculé dans la base initiale (test performance) à partir du modèle
évalué dans l'échantillon bootstrap = " as res D
di as text "Standard error de D calculée dans la base initiale (test performance) à
partir du modèle évalué dans l'échantillon bootstrap = " as res sD
scalar list

post parametres (gamma1) (gamma1_se) (r2) (r2se) (D) (sD)
postclose parametres //pour fermer la base "validation_interne", maintenant
qu'on a récupéré tous les paramètres et leurs erreurs-types

```

3.6.6. Calcul de l'optimisme du modèle multivarié estimé sur l'échantillon *bootstrap*

Comme évoqué dans la partie 2.6.2.1, la performance du modèle multivarié estimée sur l'échantillon *bootstrap* peut être surévaluée (optimisme), les données de cet échantillon ayant servi à dériver le modèle. On peut calculer cet optimisme comme la différence entre la performance du modèle calculée dans l'échantillon *bootstrap* (*bootstrap performance*) et sa performance calculée dans la base de données initiale (*test performance*).

Le code utilisé sous STATA pour le calcul de l'optimisme des différents paramètres (pente de la droite de calibration, statistique D de Royston et Sauerbrei et R_D^2 de Royston et Sauerbrei) est présenté ci-dessous :

```
use validation_interne.dta, clear //pour charger la base où ont été enregistrées
les estimations des différents paramètres

gen optimisme_g1 =.
replace optimisme_g1 = gamma1[2] - gamma1[3] if _n==2

gen optimisme_g1se =.
replace optimisme_g1se = gamma1_se[2] - gamma1_se[3] if _n==2

gen optimisme_r2 =.
replace optimisme_r2 = rcarre[2] - rcarre[3] if _n==2

gen optimisme_r2se =.
replace optimisme_r2se = rcarre_se[2] - rcarre_se[3] if _n==2

gen optimisme_d =.
replace optimisme_d = droyston[2] - droyston[3] if _n==2

gen optimisme_dse =.
replace optimisme_dse = droyston_se[2] - droyston_se[3] if _n==2
```

3.6.7. Génération de 100 échantillons *bootstrap* et estimations des paramètres de la performance du modèle

Avec la méthodologie présentée ci-dessus, 100 échantillons *bootstrap* ont été générés, suivi à chaque fois de l'estimation des coefficients β de régression du modèle multivarié dans l'échantillon *bootstrap*, de l'évaluation de la performance du modèle spécifié avec ces coefficients β dans l'échantillon *bootstrap* (*bootstrap performance*) et dans la base de données initiale (*test performance*), et du calcul de l'optimisme, comme la différence entre *bootstrap performance* et *test performance*.

Le code utilisé sous STATA pour la réalisation automatique de l'ensemble de ces étapes dans chacun des 100 échantillons est présenté ci-dessous :

```
* 1) Déterminer la performance du modèle sur la cohorte initiale (apparent performance)
quietly cd "D:\francois xavier lapebie\ANALYSE RIETE\"
use RIETEimputations2b.dta, clear
postfile parametres gamma1 gamma1_se rcarre rcarre_se droyston droyston_se using
validation_interne, replace
* 1a) Calibration
mi stset dureesuivi_1an, failure(recidive_1an)
quietly xi: mi estimate, saving(coefficients_apparperf, replace): stcox i.typecancer4
i.metastase3 i.tttcancer i.sexemasc dfgckd i.atcdeptvp i.chirurgie i.filtreaD0
i.syndrpostthromb i.seqepapr3m
mi predict xb_apparperf using coefficients_apparperf, xb
mi xeq 0: stcoxcal xb_apparperf, times(365) nograph test
return list
scalar gamma1 = r(gamma1)
scalar gamma1_se = r(gamma1_se)
* 1b) Discrimination
scalar r2 = 0
scalar r2se = 0
scalar D = 0
scalar sD = 0
xi: mi xeq 1/100 : str2d stcox i.typecancer4 i.metastase3 i.tttcancer i.sexemasc dfgckd
i.atcdeptvp i.chirurgie i.filtreaD0 i.syndrpostthromb i.seqepapr3m ; scalar r2 = r2 +
r(r2) ; scalar r2se = r2se + r(r2se) ; scalar D = D + r(D) ; scalar sD = sD + r(sD)
scalar r2 = r2/100
scalar r2se = r2se/100
scalar D = D/100
scalar sD = sD/100
* 1c) Sauvegarde des estimateurs de la performance
post parametres (gamma1) (gamma1_se) (r2) (r2se) (D) (sD)
forvalues i = 1(1)100 {
    display "Boucle bootstrap `i'"
    * 2) Créer l'échantillon bootstrap `i' de n patients avec remise, en faisant en
sorte qu'un même patient soit tiré dans les 101 bases (1 base de données observées +
100 bases de données imputées), pour permettre ensuite les analyses avec l'imputation
multiple
    * 2a) Tirage au sort avec remise, à partir de la base initiale
    quietly cd "D:\francois xavier lapebie\ANALYSE RIETE\"
    use RIETEimputations2b.dta, clear
    bsample 1000, cluster(_mi_id)
    * 2b) Renumérotation, pour que chaque valeur de _mi_id n'identifie qu'un patient
(car comme il y a remise, un même patient peut être présent plusieurs fois dans
l'échantillon bootstrap)
    duplicates tag _mi_id _mi_m, gen(obsduplic)
    sort _mi_id ( _mi_m)
```

```

by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 3414) if _n == 2
sort _mi_id (_mi_m)
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 6828) if _n == 2
sort _mi_id (_mi_m)
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 10242) if _n == 2
sort _mi_id (_mi_m)
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 13656) if _n == 2
sort _mi_id (_mi_m)
by _mi_id _mi_m obsduplic, sort: replace _mi_id = (_mi_id + 17070) if _n == 2
sort _mi_id (_mi_m)
* 2c) Sauvegarder l'échantillon bootstrap
save basebootstrap, replace
* 3) Déterminer la performance du modèle sur l'échantillon bootstrap (bootstrap
performance)
* 3a) Calibration
mi stset dureesuivi_1an, failure(recidive_1an)
quietly xi: mi estimate, saving(coefficients_bsperf, replace): stcox i.typecancer4
i.metastase3 i.tttcancer i.sexemasc dfgckd i.atcdeptvp i.chirurgie i.filtreaDO
i.syndrpostthromb i.seqepapr3m
mi predict xb_bsperf using coefficients_bsperf, xb
mi xeq 0: stcoxcal xb_bsperf, times(365) nograph test
scalar gammad1 = r(gammad1)
scalar gammad1_se = r(gammad1_se)
* 3b) Discrimination
scalar r2 = 0
scalar r2se = 0
scalar D = 0
scalar sD = 0
xi: mi xeq 1/100 : str2d stcox i.typecancer4 i.metastase3 i.tttcancer i.sexemasc
dfgckd i.atcdeptvp i.chirurgie i.filtreaDO i.syndrpostthromb i.seqepapr3m ; scalar r2 =
r2 + r(r2) ; scalar r2se = r2se + r(r2se) ; scalar D = D + r(D) ; scalar sD = sD +
r(sD)
scalar r2 = r2/100
scalar r2se = r2se/100
scalar D = D/100
scalar sD = sD/100
* 3c) Sauvegarde des estimateurs de la performance
post parametres (gammad1) (gammad1_se) (r2) (r2se) (D) (sD)
* 4) Déterminer la performance du modèle sur la population initiale (test
performance)
quietly cd "D:\francois xavier lapebie\ANALYSE RIETE\"
use RIETEimputations2b.dta, clear
* 4a) Calibration
mi stset dureesuivi_1an, failure(recidive_1an)
mi predict xb_testperf using coefficients_bsperf, xb
mi xeq 0: stcoxcal xb_testperf, times(365) nograph test
scalar gammad1 = r(gammad1)
scalar gammad1_se = r(gammad1_se)
* 4b) Desccrimination
gen testdata =.
replace testdata = 1
append using basebootstrap
replace testdata = 0 if testdata==.
duplicates tag _mi_id _mi_m, generate(obsduplic2)
sort _mi_id (_mi_m)
by _mi_id _mi_m obsduplic2, sort: replace _mi_id = (_mi_id + 20000) if _n == 2
sort _mi_id (_mi_m)
scalar r2 = 0
scalar r2se = 0
scalar D = 0
scalar sD = 0

```

```

mi stset dureesuivi_1an, failure(recidive_1an)
quietly xi: mi xeq 1/100 : str2d stcox i.typecancer4 i.metastase3 i.tttcancer
i.sexemasc dfgckd i.atcdeptvp i.chirurgie i.filtreaD0 i.syndrpostthromb i.seqepapr3m,
validate(testdata) ; scalar r2 = (r2 + r(r2)) ; scalar r2se = (r2se + r(r2se)) ; scalar
D = (D + r(D)) ; scalar sD = (sD + r(sD))
scalar r2 = r2/100
scalar r2se = r2se/100
scalar D = D/100
scalar sD = sD/100
* 4c) Sauvegarde des estimateurs de la performance
post parametres (gamma1) (gamma1_se) (r2) (r2se) (D) (sD)
}
postclose parametres

```

Les estimations de la pente de la droite de calibration, de la statistique D de Royston et Sauerbrei et du R_D^2 de Royston et Sauerbrei, calculées à partir du modèle multivarié dans la base de données initiale avec les coefficients β estimés dans la base de données initiale, et à partir du modèle multivarié dans chaque échantillon *bootstrap* et dans la base de données initiale avec les coefficients β estimés dans chaque échantillon *bootstrap*, sont ainsi stockées dans une nouvelle base de donnée, qui permettra le calcul des optimismes.

3.6.8. Calcul de l'optimisme et correction de la performance du modèle multivarié évalué dans la base de données initiale

On a ainsi un modèle multivarié différent par échantillon *bootstrap* : tous ces modèles comprennent les mêmes variables, mais les estimations des coefficients β diffèrent selon chaque échantillon. La performance de chacun de ces 100 modèles a été évaluée sur chaque échantillon *bootstrap* (*bootstrap performance*), et sur la base de données initiale (*test performance*).

Pour chacun des 100 modèles, l'optimisme a été calculé comme la différence entre *bootstrap performance* et *test performance*. Cela a été réalisé pour les paramètres suivants :

- la pente de la droite de calibration,
- la statistique D de Royston et Sauerbrei,
- le R_D^2 de Royston et Sauerbrei.

Les 100 optimismes ont ensuite été moyennés, pour obtenir un indicateur unique.

Enfin, cet optimisme moyen a été soustrait de chacun des paramètres de la performance du modèle initial (dont les coefficients β sont présentés dans le tableau 1), estimés dans la base de données initiale.

Le code utilisé sous STATA pour le calcul de l'optimisme de chaque paramètre de performance, évalué dans chacun des 100 échantillons, le calcul de l'optimisme moyen et la correction de la performance apparente du modèle initial est présenté ci-dessous :

```
* 5) Calculer l'optimisme: différence entre bootstrap performance et test performance
quietly cd "D:\francois xavier lapebie\ANALYSE RIETE\"
use validation_interne.dta, clear
egen nummodelebs = seq() if _n!=1, from(1) to(100) block(2) //pour attribuer un
numéro aux observations (= lignes), correspondant à chaque modèle estimé sur un
échantillon bootstrap: donc pas de numéro pour la 1ère ligne (qui contient les
paramètres du modèle estimés dans la population initiale), et ensuite même numéro pour
2 lignes successives (car chaque modèle avec les coefficients estimés sur un
échantillon bootstrap donné est testé deux fois: sur l'échantillon bootstrap =
bootstrap performance, et sur la population initiale = test performance)
replace nummodelebs=0 if _n==1
egen typepopstest = seq() if _n!=1, from(1) to(2) block(1) //1 correspond à un
échantillon bootstrap, 2 correspond à la population initiale
replace typepopstest=2 if _n==1
reshape wide gammal gammal_se rcarre rcarre_se droyston droyston_se, i(nummodelebs)
j(typepopstest) //pour modifier la base: chaque échantillon bootstrap sur une seule
ligne, qui comprend tous les paramètres de performance estimés pour un modèle donné
(d'abord les paramètres estimés sur l'échantillon bootstrap, puis ceux estimés sur la
population initiale); la 1ère ligne correspond aux paramètres estimés par le modèle
développé sur la population initiale
gen optimisme_g1 = gammal1 - gammal2 if _n!=1 //optimisme de gammal
gen optimisme_r2 = rcarre1 - rcarre2 if _n!=1 //optimisme de R^2
gen optimisme_d = droyston1 - droyston2 if _n!=1 //optimisme de D
* 6) Calculer la moyenne des optimismes de chaque paramètre
egen moy_optg1 = mean(optimisme_g1) if _n!=1 //moyenne des optimismes de gammal
egen moy_optr2 = mean(optimisme_r2) if _n!=1 //moyenne des optimismes de R^2
egen moy_optd = mean(optimisme_d) if _n!=1 //moyenne des optimismes de D
* 7) Soustraire la moyenne de l'optimisme de chaque paramètre, à l'estimation du
paramètre obtenue dans la population initiale = optimism-corrected estimate of
performance
gen optcorr_g1 = gammal2[1] - moy_optg1[2] if _n==1 //gammal2[1] identifie la valeur
de gammal estimée dans la population de la ligne 1, donc dans la population initiale
codebook optcorr_g1 if optcorr_g1!=. //optimism-corrected estimate of performance
gen optcorr_r2 = rcarre2[1] - moy_optr2[2] if _n==1
codebook optcorr_r2 if optcorr_r2!=. //optimism-corrected estimate of performance
gen optcorr_d = droyston2[1] - moy_optd[2] if _n==1
codebook optcorr_d if optcorr_d!=. //optimism-corrected estimate of performance
```

Les différents optimismes moyens ainsi calculés sont :

- pour la pente de la droite de calibration : 0,3094,
- pour la statistique D de Royston et Sauerbrei : 0,3134,
- pour le R_D^2 de Royston et Sauerbrei : 0,0974.

Et les paramètres de la performance du modèle multivarié, corrigés à partir de la performance apparente par les optimismes moyens, sont :

- pour la pente de la droite de calibration : 0,8241,
- pour la statistique D de Royston et Sauerbrei : 0,8298,
- pour le R_D^2 de Royston et Sauerbrei : 0,1405.

Le calcul des IC95% de ces différents paramètres de la performance apparente corrigés par les optimismes a été réalisée selon la méthode « *location-shifted bootstrap confidence interval* », décrite dans la partie 2.6.2.1, avec 1000 réplifications *bootstrap*.

Le code utilisé sous STATA pour le calcul de ces IC95% est présenté ci-dessous :

```
* 8) Calculer l'intervalle de confiance des paramètres corrigés par l'optimisme :
location-shifted bootstrap confidence interval

quietly cd "D:\francois xavier lapebie\ANALYSE RIETE\"
use RIETEimputations2b.dta, clear
mi stset dureesuivi_1an, failure(recidive_1an)

xi: mi estimate, saving(coefficients_apparperf, replace): stcox i.typecancer4
i.metastase3 i.tttcancer i.sexemasc dfgckd i.atcdeptvp i.chirurgie i.filtreaD0
i.syndrpostthromb i.seqepapr3m
mi predict xb_apparperf using coefficients_apparperf, xb
mi xeq 0: bootstrap gamma1_bs = r(gamma1), reps(1000): stcoxal xb_apparperf,
times(365) nograph test //pour obtenir un 95%CI de la pente de la droite, calculé par
bootstrap sur la base de données initiale: pente 1.133495, standard error par bootstrap
.1056134, (95%CI, .9264962, 1.340493)
//Donc borne inférieure de 95%CI de la pente corrigée par l'optimisme = .9264962 -
optimisme = 0,61706158, et borne supérieure = 1.340493 - optimisme = 1,03105838

scalar r2 = 0
scalar r2se = 0
scalar D = 0
scalar sD = 0
xi: mi xeq 1/100 : str2d stcox i.typecancer4 i.metastase3 i.tttcancer i.sexemasc dfgckd
i.atcdeptvp i.chirurgie i.filtreaD0 i.syndrpostthromb i.seqepapr3m, bootreps(1000) ;
scalar r2 = r2 + r(r2) ; scalar r2se = r2se + r(r2se) ; scalar D = D + r(D) ; scalar sD
= sD + r(sD)
scalar r2 = r2/100
scalar r2se = r2se/100
scalar D = D/100
scalar sD = sD/100

di as text "R^2 dans la base de données initiale avec imputation des données manquantes
(R^2 de la performance apparente) = " as res r2 //R^2 = .23782741
di as text "Standard error de R^2 calculée par bootstrap dans la base de données
initiale avec imputation des données manquantes = " as res r2se //SE = .03312304; et
faire 1,96 x SE = 0,0649211584 pour obtenir la borne basse et la borne haute du 95%CI
de R^2 apparent: 95%CI, 0,1729062516 à 0,3027485684
//Donc borne inférieure de 95%CI de R^2 corrigée par l'optimisme = 0,1729062516 -
optimisme = 0,0755457016, et borne supérieure = 0,3027485684 - optimisme = 0,2053880184
di as text "D dans la base de données initiale avec imputation des données manquantes
(D de la performance apparente) = " as res D //D = 1.1431383
di as text "Standard error de D calculée par bootstrap dans la base de données initiale
avec imputation des données manquantes = " as res sD //SE = .10108506; et faire 1,96 x
```

SE = 0,1981267176 pour obtenir la borne basse et la borne haute du 95%CI du D apparent :
95%CI, 0,9450115824 à 1,24422336
// Donc borne inférieure de 95%CI de D corrigée par l'optimisme = 0,9450115824 -
optimisme = 0,6316465124, et borne supérieure = 1,24422336 - optimisme = 0,93085829

Ainsi, on obtient dans un premier temps les IC95%, calculés par *bootstrap*, des paramètres de la performance apparente évaluée sur la base de données initiales, en tenant compte de l'imputation multiple des données manquantes :

- pour la pente de la droite de calibration : IC95%, 0,9265 à 1,3405,
- pour la statistique D de Royston et Sauerbrei : IC95%, 0,9450 à 1,2442,
- pour le R_D^2 de Royston et Sauerbrei : IC95%, 0,1729 à 0,3027.

L'optimisme sera ensuite soustrait aux bornes basse et haute de ces IC95%.

Au final, les paramètres de la performance apparente du modèle multivarié et leurs IC95%, tous corrigés par l'optimisme, sont :

- pour la pente de la droite de calibration : 0,8241 (IC95%, 0,6171-1,0311),
- pour la statistique D de Royston et Sauerbrei : 0,8298 (IC95%, 0,6316-0,9309),
- pour le R_D^2 de Royston et Sauerbrei : 0,1405 (IC95%, 0,0755-0,2054).

La performance du modèle prédictif semble donc correcte.

La pente de la droite de calibration est donc une mesure de la calibration, elle est ici proche de 1 avec un IC95% comprenant la valeur 1, correspondant à une calibration correcte. En effet, une valeur très inférieure à 1 indiquerait un surajustement aux données, avec des risques prédits qui pourraient être trop faibles pour les patients à risque faible, ou trop élevés pour les patients à risque élevé.

La statistique D de Royston et Sauerbrei est une mesure de la discrimination, qui quantifie la séparation entre les sujets avec un risque prédit bas et ceux avec un risque prédit élevé. Elle s'interprète comme le logarithme du HR qui comparerait deux groupes à risque de mêmes effectifs, définis par une dichotomisation, selon la valeur médiane, de la distribution des indices pronostiques des patients. Ici, cet HR serait donc $e^{0,8298} = 2,29$. Son IC95% serait de 1,88 à 2,54. La discrimination est donc correcte.

Le R_D^2 de Royston et Sauerbrei quantifie la proportion de la variance dans l'évènement d'intérêt expliquée par les variables indépendantes du modèle prédictif. Ici, le résultat est médiocre, avec 14,05% de la variance expliquée. Cependant, l'intérêt des mesures de type R^2 semble inférieur aux mesures de la calibration et de la discrimination.

4. Conclusion

Le développement et la validation d'un modèle multivarié à but prédictif à partir d'une cohorte sont donc complexes, reposant sur plusieurs étapes essentielles. Pour les données longitudinales, le choix du type de modèle se porte souvent sur le modèle à risques proportionnels de Cox, mais le modèle de Fine et Gray est une alternative intéressante en présence de risques compétitifs. La sélection des variables à inclure est fondamentale, reposant souvent sur des algorithmes appliqués aux données étudiées, telle que la sélection descendante, mais qui posent plusieurs problèmes tels que l'exclusion de facteurs de confusion importants, un biais de sélection, un surajustement, et une instabilité notamment si le nombre d'évènements d'intérêt rapporté au nombre de variables sur lesquelles sont opérées la sélection est faible. Une sélection a priori des variables du modèle final sur une expertise et les données de la littérature, indépendamment des données étudiées, peut être préférée. Dans cette optique, les graphes acycliques dirigés sont utiles pour formaliser les liens entre variables, étudier une association causale, identifier les facteurs de médiation, les facteurs de confusion et d'autres types de biais. Des méthodes mixtes entre une sélection sur les connaissances préalables et une sélection automatique sont enfin possibles. Un choix devra aussi être fait concernant la gestion des données manquantes, l'analyse des seuls cas complets étant par nature biaisée. Une imputation multiple des données manquantes, notamment par la méthode des équations chaînées, est actuellement préférée, une condition d'application importante étant que le mécanisme rendant compte de l'absence de données soit aléatoire. Après dérivation, le modèle multivarié devra être validé. La validation interne consiste à évaluer ses capacités prédictives sur la base de données initiale ; elle sera préférentiellement réalisée par un rééchantillonnage de type *bootstrap*. Appliquée après une imputation multiple des données manquantes de la base, elle reste complexe. La validation externe étudie la reproductibilité et la transposabilité du modèle. La transposabilité notamment nécessitera de nouvelles bases de données, ou des études prospectives de validation. Dans ce travail, nous avons proposé des applications des différentes étapes du développement et de la validation des modèles multivariés à partir des cohortes COPART et RIETE. En particulier, un modèle multivarié prédictif du risque de récurrence d'évènement thromboembolique veineux au cours du cancer, à l'arrêt du traitement anticoagulant, a été dérivé de la cohorte RIETE avec des résultats satisfaisants lors de la validation interne. L'étape suivante serait une validation externe de ce score dans d'autres cohortes, suivie, selon les résultats, d'une étude interventionnelle prospective évaluant son

impact en pratique clinique. Si le score se révélait fiable, une version simplifiée pourrait être développée, afin de faciliter son utilisation.

5. Bibliographie

1. Heinze G, Wallisch C, Dunkler D. Variable selection - A review and recommendations for the practicing statistician. *Biom J.* 2018; 60 (3): 431-49.
2. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. *BMJ.* 2009; 338: b604.
3. Shmueli G. To Explain or to Predict? *Statistical Science.* 2010; 25 (3): 289-310.
4. Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015; 162 (1): W1-73.
5. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials.* 2015; 16: 495.
6. Bolisliis WR, Fay M, Kühler TC. Use of Real-world Data for New Drug Applications and Line Extensions. *Clin Ther.* 2020; 42 (5): 926-38.
7. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet.* 2002; 359 (9302): 248-52.
8. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ.* 2020; 368: m441.
9. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011; 30 (4): 377-99.
10. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol.* 1995; 142 (12): 1255-64.
11. Tsvetanova A, Sperrin M, Peek N, Buchan I, Hyland S, Martin GP. Missing data was handled inconsistently in UK prediction models: a review of method used. *J Clin Epidemiol.* 2021; S0895-4356(21)00288-2.
12. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. *BMC Med Res Methodol.* 2015; 15: 30.

13. Nguyen CD, Carlin JB, Lee KJ. Model checking in multiple imputation: an overview and case study. *Emerg Themes Epidemiol.* 2017; 14: 8.
14. Rao SR, Schoenfeld DA. Survival methods. *Circulation.* 2007; 115 (1): 109-13.
15. Timsit JF, Alberti C, Chevret S. [Cox proportional hazards regression analysis]. *Rev Mal Respir.* 2005; 22 (6 Pt 1): 1058-64.
16. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. *Stat Med.* 2017; 36 (27): 4391-400.
17. Austin PC, Lee DS, D'Agostino RB, Fine JP. Developing points-based risk-scoring systems in the presence of competing risks. *Stat Med.* 2016; 35 (22): 4056-72.
18. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. *Circulation.* 2016; 133 (6): 601-9.
19. Dunkler D, Plischke M, Leffondré K, Heinze G. Augmented backward elimination: a pragmatic and purposeful way to develop statistical models. *PLoS One.* 2014; 9 (11): e113677.
20. Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology.* 2009; 20 (4): 488-95.
21. Wallisch C, Dunkler D, Rauch G, de Bin R, Heinze G. Selection of variables for multivariable models: Opportunities and limitations in quantifying model stability by resampling. *Stat Med.* 2021; 40 (2): 369-81.
22. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol.* 1999; 28 (5): 964-74.
23. Heinze G, Dunkler D. Five myths about variable selection. *Transpl Int.* 2017; 30 (1): 6-10.
24. Yiu ZZN, Sorbe C, Lunt M, Rustenbach SJ, Kühl L, Augustin M, et al. Development and validation of a multivariable risk prediction model for serious infection in patients with psoriasis receiving systemic therapy. *Br J Dermatol.* 2019; 180 (4): 894-901.
25. Etminan M, Collins GS, Mansournia MA. Using Causal Diagrams to Improve the Design and Interpretation of Medical Research. *Chest.* 2020; 158 (1S): S21-8.

26. Ferguson KD, McCann M, Katikireddi SV, Thomson H, Green MJ, Smith DJ, et al. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *Int J Epidemiol.* 2020; 49 (1): 322-9.
27. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. *Int J Epidemiol.* 2021; 50 (2): 620-32.
28. Suzuki E, Shinozaki T, Yamamoto E. Causal Diagrams: Pitfalls and Tips. *J Epidemiol.* 2020; 30 (4): 153-62.
29. Greenland S, Mansournia MA. Limitations of individual causal models, causal graphs, and ignorability assumptions, as illustrated by random confounding and design unfaithfulness. *Eur J Epidemiol.* 2015; 30 (10): 1101-10.
30. Textor J, van der Zander B, Gilthorpe MS, Liskiewicz M, Ellison GT. Robust causal inference using directed acyclic graphs: the R package « dagitty ». *Int J Epidemiol.* 2016; 45 (6): 1887-94.
31. Wood AM, White IR, Royston P. How should variable selection be performed with multiply imputed data? *Stat Med.* 2008; 27 (17): 3227-46.
32. Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol.* 2013; 13: 33.
33. Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998; 97 (18): 1837-47.
34. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart.* 2012; 98 (9): 683-90.
35. Crowson CS, Atkinson EJ, Therneau TM. Assessing calibration of prognostic risk scores. *Stat Methods Med Res.* 2016; 25 (4): 1692-706.
36. Rahman MS, Ambler G, Choodari-Oskoei B, Omar RZ. Review and evaluation of performance measures for survival prediction models in external validation settings. *BMC Med Res Methodol.* 2017; 17 (1): 60.

37. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. *Crit Care Med.* 2007; 35 (9): 2052-6.
38. Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biom J.* 2015; 57 (4): 614-32.
39. Pencina MJ, D'Agostino RB, Song L. Quantifying discrimination of Framingham risk functions with different survival C statistics. *Stat Med.* 2012; 31 (15): 1543-53.
40. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med.* 2004; 23 (5): 723-48.
41. Choodari-Oskooei B, Royston P, Parmar MKB. A simulation study of predictive ability measures in a survival model I: explained variation measures. *Stat Med.* 2012; 31 (23): 2627-43.
42. Royle K-L, Cairns DA. The development and validation of prognostic models for overall survival in the presence of missing data in the training dataset: a strategy with a detailed example. *Diagn Progn Res.* 2021; 5 (1): 14.
43. Royston P, Parmar MKB, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Stat Med.* 2004; 23 (6): 907-26.
44. Choodari-Oskooei B, Royston P, Parmar MKB. A simulation study of predictive ability measures in a survival model II: explained randomness and predictive accuracy. *Stat Med.* 2012; 31 (23): 2644-59.
45. Rufibach K. Use of Brier score to assess binary predictions. *J Clin Epidemiol.* 2010; 63 (8): 938-9; author reply 939.
46. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006; 26 (6): 565-74.
47. Baker SG, Cook NR, Vickers A, Kramer BS. Using relative utility curves to evaluate risk prediction. *J R Stat Soc Ser A Stat Soc.* 2009; 172 (4): 729-48.
48. Wahl S, Boulesteix A-L, Zierer A, Thorand B, van de Wiel MA. Assessment of predictive performance in incomplete data by combining internal validation and multiple imputation. *BMC Med Res Methodol.* 2016; 16 (1): 144.

49. Steyerberg EW, Harrell FE, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* 2001; 54 (8): 774-81.
50. Steyerberg EW. Validation in prediction research: the waste by data splitting. *J Clin Epidemiol.* 2018; 103: 131-3.
51. Noma H, Shinozaki T, Iba K, Teramukai S, Furukawa TA. Confidence intervals of prediction accuracy measures for multivariable prediction models based on the bootstrap-based optimism correction methods. *Stat Med.* 2021 Jul 24. doi: 10.1002/sim.9148.
52. Andrews DWK, Buchinsky M. A Three-step Method for Choosing the Number of Bootstrap Repetitions. *Econometrica.* 2000; 68 (1): 23-51.
53. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Stat Med.* 2018; 37 (14): 2252-66.
54. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med.* 1999; 130 (6): 515-24.
55. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J.* 2021; 14 (1): 49-58.
56. Snell KIE, Archer L, Ensor J, Bonnett LJ, Debray TPA, Phillips B, et al. External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb. *J Clin Epidemiol.* 2021; 135: 79-89.
57. Riley RD, Debray TPA, Collins GS, Archer L, Ensor J, van Smeden M, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med.* 2021; 40 (19): 4230-51.
58. Archer L, Snell KIE, Ensor J, Hudda MT, Collins GS, Riley RD. Minimum sample size for external validation of a clinical prediction model with a continuous outcome. *Stat Med.* 2021; 40 (1): 133-46.
59. Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KGM. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol.* 2015; 68 (3): 279-89.

60. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012; 98 (9): 691-8.
61. Aboyans V, Criqui MH, Abraham P, Allison MA, Creager MA, Diehm C, et al. Measurement and interpretation of the ankle-brachial index: a scientific statement from the American Heart Association. *Circulation*. 2012; 126 (24): 2890-909.
62. Conte MS, Bradbury AW, Kolh P, White JV, Dick F, Fitridge R, et al. Global vascular guidelines on the management of chronic limb-threatening ischemia. *J Vasc Surg*. 2019; 69 (6S): 3S-125S.e40.
63. Sigvant B, Hasvold P, Kragsterman B, Falkenberg M, Johansson S, Thuresson M, et al. Cardiovascular outcomes in patients with peripheral arterial disease as an initial or subsequent manifestation of atherosclerotic disease: Results from a Swedish nationwide study. *J Vasc Surg*. 2017; 66 (2): 507-514.e1.
64. Sigvant B, Kragsterman B, Falkenberg M, Hasvold P, Johansson S, Thuresson M, et al. Contemporary cardiovascular risk and secondary preventive drug treatment patterns in peripheral artery disease patients undergoing revascularization. *J Vasc Surg*. 2016; 64 (4): 1009-1017.e3.
65. Sigvant B, Lundin F, Wahlberg E. The Risk of Disease Progression in Peripheral Arterial Disease is Higher than Expected: A Meta-Analysis of Mortality and Disease Progression in Peripheral Arterial Disease. *Eur J Vasc Endovasc Surg*. 2016; 51 (3): 395-403.
66. Aboyans V, Ricco J-B, Bartelink M-LEL, Björck M, Brodmann M, Cohnert T, et al. 2017 ESC Guidelines on the Diagnosis and Treatment of Peripheral Arterial Diseases, in collaboration with the European Society for Vascular Surgery (ESVS): Document covering atherosclerotic disease of extracranial carotid and vertebral, mesenteric, renal, upper and lower extremity arteries. Endorsed by: the European Stroke Organization (ESO) The Task Force for the Diagnosis and Treatment of Peripheral Arterial Diseases of the European Society of Cardiology (ESC) and of the European Society for Vascular Surgery (ESVS). *Eur Heart J*. 2018; 39 (9): 763-816.
67. Gerhard-Herman MD, Gornik HL, Barrett C, Barshes NR, Corriere MA, Drachman DE, et al. 2016 AHA/ACC Guideline on the Management of Patients With Lower Extremity Peripheral Artery Disease: Executive Summary: A Report of the American College of

Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation*. 2017; 135 (12): e686-725.

68. Cambou JP, Aboyans V, Constans J, Lacroix P, Dentans C, Bura A. Characteristics and outcome of patients hospitalised for lower extremity peripheral artery disease in France: the COPART Registry. *Eur J Vasc Endovasc Surg*. 2010; 39 (5): 577-85.

69. Pros N, Cambou JP, Aboyans V, Malloizel Delaunay J, Constans J, Lacroix P, et al. A hospital discharge risk score for 1-year all-cause mortality or non-fatal cardiovascular events in patients with lower-extremity peripheral artery disease, with and without revascularisation. *Eur J Vasc Endovasc Surg*. 2013; 45 (5): 488-96.

70. Hackl G, Belaj K, Gary T, Rief P, Deutschmann H, Seinost G, et al. COPART Risk Score Predicts Long-term Mortality in Peripheral Arterial Occlusive Disease. *Eur J Vasc Endovasc Surg*. 2015; 50 (1): 94-100.

71. Mills JL, Conte MS, Armstrong DG, Pomposelli FB, Schanzer A, Sidawy AN, et al. The Society for Vascular Surgery Lower Extremity Threatened Limb Classification System: risk stratification based on wound, ischemia, and foot infection (WIFI). *J Vasc Surg*. 2014; 59 (1): 220-234.e1-2.

72. Mathioudakis N, Hicks CW, Canner JK, Sherman RL, Hines KF, Lum YW, et al. The Society for Vascular Surgery Wound, Ischemia, and foot Infection (WIFI) classification system predicts wound healing but not major amputation in patients with diabetic foot ulcers treated in a multidisciplinary setting. *J Vasc Surg*. 2017; 65 (6): 1698-1705.e1.

73. Hicks CW, Canner JK, Mathioudakis N, Sherman R, Malas MB, Black JH, et al. The Society for Vascular Surgery Wound, Ischemia, and foot Infection (WIFI) classification independently predicts wound healing in diabetic foot ulcers. *J Vasc Surg*. 2018; 68 (4): 1096-103.

74. Aujesky D, Obrosky DS, Stone RA, Auble TE, Perrier A, Cornuz J, et al. Derivation and validation of a prognostic model for pulmonary embolism. *Am J Respir Crit Care Med*. 2005; 172 (8): 1041-6.

75. Jiménez D, Aujesky D, Moores L, Gómez V, Lobo JL, Uresandi F, et al. Simplification of the pulmonary embolism severity index for prognostication in patients with acute symptomatic pulmonary embolism. *Arch Intern Med*. 2010; 170 (15): 1383-9.

76. Kearon C, Akl EA, Ornelas J, Blaivas A, Jimenez D, Bounameaux H, et al. Antithrombotic Therapy for VTE Disease: CHEST Guideline and Expert Panel Report. *Chest*. 2016; 149 (2): 315-52.
77. Sanchez O, Benhamou Y, Bertolotti L, Constant J, Couturaud F, Delluc A, et al. [Recommendations of good practice for the management of thromboembolic venous disease in adults. Short version]. *Rev Mal Respir*. 2019; 36 (2): 249-83.
78. Kearon C, Ageno W, Cannegieter SC, Cosmi B, Geersing G-J, Kyrle PA, et al. Categorization of patients as having provoked or unprovoked venous thromboembolism: guidance from the SSC of ISTH. *J Thromb Haemost*. 2016; 14 (7): 1480-3.
79. Iorio A, Kearon C, Filippucci E, Marcucci M, Macura A, Pengo V, et al. Risk of recurrence after a first episode of symptomatic venous thromboembolism provoked by a transient risk factor: a systematic review. *Arch Intern Med*. 2010; 170 (19): 1710-6.
80. Prandoni P, Noventa F, Ghirarduzzi A, Pengo V, Bernardi E, Pesavento R, et al. The risk of recurrent venous thromboembolism after discontinuing anticoagulation in patients with acute proximal deep vein thrombosis or pulmonary embolism. A prospective cohort study in 1,626 patients. *Haematologica*. 2007; 92 (2): 199-205.
81. Khan F, Rahman A, Carrier M, Kearon C, Weitz JI, Schulman S, et al. Long term risk of symptomatic recurrent venous thromboembolism after discontinuation of anticoagulant treatment for first unprovoked venous thromboembolism event: systematic review and meta-analysis. *BMJ*. 2019; 366: 14363.
82. Heit JA, Silverstein MD, Mohr DN, Petterson TM, O'Fallon WM, Melton LJ. Risk factors for deep vein thrombosis and pulmonary embolism: a population-based case-control study. *Arch Intern Med*. 2000; 160 (6): 809-15.
83. Horsted F, West J, Grainge MJ. Risk of venous thromboembolism in patients with cancer: a systematic review and meta-analysis. *PLoS Med*. 2012; 9 (7): e1001275.
84. Napolitano M, Saccullo G, Malato A, Sprini D, Ageno W, Imberti D, et al. Optimal duration of low molecular weight heparin for the treatment of cancer-related deep vein thrombosis: the Cancer-DACUS Study. *J Clin Oncol*. 2014; 32 (32): 3607-12.
85. Jara-Palomares L, Solier-Lopez A, Elias-Hernandez T, Asensio-Cruz MI, Blasco-Esquivias I, Sanchez-Lopez V, et al. D-dimer and high-sensitivity C-reactive protein levels to

predict venous thromboembolism recurrence after discontinuation of anticoagulation for cancer-associated thrombosis. *Br J Cancer*. 2018; 119 (8): 915-21.

86. Rodger MA, Kahn SR, Wells PS, Anderson DA, Chagnon I, Le Gal G, et al. Identifying unprovoked thromboembolism patients at low risk for recurrence who can discontinue anticoagulant therapy. *CMAJ*. 2008; 179 (5): 417-26.

87. Eichinger S, Heinze G, Jandeck LM, Kyrle PA. Risk assessment of recurrence in patients with unprovoked deep vein thrombosis or pulmonary embolism: the Vienna prediction model. *Circulation*. 2010; 121 (14): 1630-6.

88. Tosetto A, Iorio A, Marcucci M, Baglin T, Cushman M, Eichinger S, et al. Predicting disease recurrence in patients with previous unprovoked venous thromboembolism: a proposed prediction score (DASH). *J Thromb Haemost*. 2012; 10 (6): 1019-25.

89. Ensor J, Riley RD, Jowett S, Monahan M, Snell KI, Bayliss S, et al. Prediction of risk of recurrence of venous thromboembolism following treatment for a first unprovoked venous thromboembolism: systematic review, prognostic model and clinical decision rule, and economic evaluation. *Health Technol Assess*. 2016; 20 (12): i-xxxiii, 1-190.

90. Rodger MA, Le Gal G, Anderson DR, Schmidt J, Pernod G, Kahn SR, et al. Validating the HERDOO2 rule to guide treatment duration for women with unprovoked venous thrombosis: multinational prospective cohort management study. *BMJ*. 2017; 356: j1065.

91. Louzada ML, Carrier M, Lazo-Langner A, Dao V, Kovacs MJ, Ramsay TO, et al. Development of a clinical prediction rule for risk stratification of recurrent venous thromboembolism in patients with cancer-associated venous thromboembolism. *Circulation*. 2012; 126 (4): 448-54.

92. Bikdeli B, Jimenez D, Hawkins M, Ortíz S, Prandoni P, Brenner B, et al. Rationale, Design and Methodology of the Computerized Registry of Patients with Venous Thromboembolism (RIETE). *Thromb Haemost*. 2018; 118 (1): 214-24.

93. Fuentes HE, Tafur AJ, Caprini JA, Alatri A, Trujillo-Santos J, Farge-Bancel D, et al. Prediction of early mortality in patients with cancer-associated thrombosis in the RIETE Database. *Int Angiol*. 2019; 38 (3): 173-84.

94. den Exter PL, Gómez V, Jiménez D, Trujillo-Santos J, Muriel A, Huisman MV, et al. A clinical prognostic model for the identification of low-risk patients with acute symptomatic pulmonary embolism and active cancer. *Chest*. 2013; 143 (1): 138-45.
95. Trujillo-Santos J, Herrera S, Page MA, Soto MJ, Raventós A, Sánchez R, et al. Predicting adverse outcome in outpatients with acute deep vein thrombosis. findings from the RIETE Registry. *J Vasc Surg*. 2006; 44 (4): 789-93.
96. Mora D, Nieto JA, Mateo J, Bikdeli B, Barco S, Trujillo-Santos J, et al. Machine Learning to Predict Outcomes in Patients with Acute Pulmonary Embolism Who Prematurely Discontinued Anticoagulant Therapy. *Thromb Haemost*. 2021 Jun 9. doi: 10.1055/a-1525-7220.
97. Anand SS, Dagenais GR, Mohan V, Diaz R, Probstfield J, Freeman R, et al. Glucose levels are associated with cardiovascular disease and death in an international cohort of normal glycaemic and dysglycaemic men and women: the EpiDREAM cohort study. *Eur J Prev Cardiol*. 2012; 19 (4): 755-64.
98. Rawshani A, Rawshani A, Franzén S, Eliasson B, Svensson A-M, Miftaraj M, et al. Mortality and Cardiovascular Disease in Type 1 and Type 2 Diabetes. *N Engl J Med*. 2017; 376 (15): 1407-18.
99. Tancredi M, Rosengren A, Svensson A-M, Kosiborod M, Pivodic A, Gudbjörnsdottir S, et al. Excess Mortality among Persons with Type 2 Diabetes. *N Engl J Med*. 2015; 373 (18): 1720-32.
100. Fowkes FGR, Rudan D, Rudan I, Aboyans V, Denenberg JO, McDermott MM, et al. Comparison of global estimates of prevalence and risk factors for peripheral artery disease in 2000 and 2010: a systematic review and analysis. *Lancet*. 2013; 382 (9901): 1329-40.
101. Althouse AD, Abbott JD, Forker AD, Bertolet M, Barinas-Mitchell E, Thurston RC, et al. Risk factors for incident peripheral arterial disease in type 2 diabetes: results from the Bypass Angioplasty Revascularization Investigation in type 2 Diabetes (BARI 2D) Trial. *Diabetes Care*. 2014; 37 (5): 1346-52.
102. Mohammedi K, Woodward M, Hirakawa Y, Zoungas S, Williams B, Lisheng L, et al. Microvascular and Macrovascular Disease and Risk for Major Peripheral Arterial Disease in Patients With Type 2 Diabetes. *Diabetes Care*. 2016; 39 (10): 1796-803.

103. Malyar NM, Freisinger E, Meyborg M, Lüders F, Gebauer K, Reinecke H, et al. Amputations and mortality in in-hospital treated patients with peripheral artery disease and diabetic foot syndrome. *J Diabetes Complicat.* 2016; 30 (6): 1117-22.
104. Vrsalovic M, Vucur K, Vrsalovic Presecki A, Fabijanic D, Milosevic M. Impact of diabetes on mortality in peripheral artery disease: a meta-analysis. *Clin Cardiol.* 2017; 40 (5): 287-91.
105. Humphries MD, Brunson A, Hedayati N, Romano P, Melnkow J. Amputation Risk in Patients with Diabetes Mellitus and Peripheral Artery Disease Using Statewide Data. *Ann Vasc Surg.* 2016; 30: 123-31.
106. Agnelli G, Belch JFF, Baumgartner I, Giovass P, Hoffmann U. Morbidity and mortality associated with atherosclerotic peripheral artery disease: A systematic review. *Atherosclerosis.* 2020; 293: 94-100.
107. Low Wang CC, Blomster JI, Heizer G, Berger JS, Baumgartner I, Fowkes FGR, et al. Cardiovascular and Limb Outcomes in Patients With Diabetes and Peripheral Artery Disease: The EUCLID Trial. *J Am Coll Cardiol.* 2018; 72 (25): 3274-84.
108. Frank U, Nikol S, Belch J, Boc V, Brodmann M, Carpentier PH, et al. ESVM Guideline on peripheral arterial disease. *Vasa.* 2019; 48 (Suppl 102): 1-79.
109. Society for Vascular Surgery Lower Extremity Guidelines Writing Group, Conte MS, Pomposelli FB, Clair DG, Geraghty PJ, McKinsey JF, et al. Society for Vascular Surgery practice guidelines for atherosclerotic occlusive disease of the lower extremities: management of asymptomatic disease and claudication. *J Vasc Surg.* 2015; 61 (3 Suppl): 2S-41S.
110. Lapébie F-X, Bura-Rivière A, Lacroix P, Constans J, Boulon C, Messas E, et al. Impact of angiotensin receptor blockers on mortality after hospitalization for symptomatic lower extremity artery disease. *Eur Heart J Cardiovasc Pharmacother.* 2020 Apr 9; pvaa026. doi: 10.1093/ehjcvp/pvaa026.
111. Coudene A, Lapébie F-X, Desormais I, Lacroix P, Aurillac V, Mangin M, et al. Evolution of Major Amputation Risk in Patients Hospitalized in France for Critical Limb Ischemia: The COPART Registry. *Angiology.* 2021; 72 (4): 315-21.
112. Alberti KGMM, Eckel RH, Grundy SM, Zimmet PZ, Cleeman JI, Donato KA, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes

Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. *Circulation*. 2009; 120 (16): 1640-5.

113. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009; 150 (9): 604-12.

114. Levey AS, Eckardt K-U, Dorman NM, Christiansen SL, Cheung M, Jadoul M, et al. Nomenclature for Kidney Function and Disease: Executive Summary and Glossary From a Kidney Disease: Improving Global Outcomes Consensus Conference. *Transplantation*. oct 2020;104(10):1986-94.

115. Caro J, Migliaccio-Walle K, Ishak KJ, Proskorovsky I. The morbidity and mortality following a diagnosis of peripheral arterial disease: long-term follow-up of a large database. *BMC Cardiovasc Disord*. 2005; 5: 14.

116. Mueller T, Hinterreiter F, Luft C, Poelz W, Haltmayer M, Dieplinger B. Mortality rates and mortality predictors in patients with symptomatic peripheral artery disease stratified according to age and diabetes. *J Vasc Surg*. 2014; 59 (5): 1291-9.

117. Abu Dabrh AM, Steffen MW, Undavalli C, Asi N, Wang Z, Elamin MB, et al. The natural history of untreated severe or critical limb ischemia. *J Vasc Surg*. 2015; 62 (6): 1642-1651.e3.

118. Reinecke H, Unrath M, Freisinger E, Bunzemeier H, Meyborg M, Lüders F, et al. Peripheral arterial disease and critical limb ischaemia: still poor outcomes and lack of guideline adherence. *Eur Heart J*. 2015; 36 (15): 932-8.

119. Beckman JA, Creager MA, Libby P. Diabetes and atherosclerosis: epidemiology, pathophysiology, and management. *JAMA*. 2002; 287 (19): 2570-81.

120. Golledge J, Quigley F, Velu R, Walker PJ, Moxon JV. Association of impaired fasting glucose, diabetes and their management with the presentation and outcome of peripheral artery disease: a cohort study. *Cardiovasc Diabetol*. 2014; 13: 147.

121. Anand SS, Caron F, Eikelboom JW, Bosch J, Dyal L, Aboyans V, et al. Major Adverse Limb Events and Mortality in Patients With Peripheral Artery Disease: The COMPASS Trial. *J Am Coll Cardiol*. 2018; 71 (20): 2306-15.

122. Paraskevas KI, de Borst GJ, Veith FJ. Why randomized controlled trials do not always reflect reality. *J Vasc Surg.* 2019; 70 (2): 607-614.e3.
123. Sartipy F, Sigvant B, Lundin F, Wahlberg E. Ten Year Mortality in Different Peripheral Arterial Disease Stages: A Population Based Observational Study on Outcome. *Eur J Vasc Endovasc Surg.* 2018; 55 (4): 529-36.
124. Visonà A, De Paoli A, Fedeli U, Tonello D, Zalunardo B, Zanatta N, et al. Abnormal ankle-brachial index (ABI) predicts primary and secondary cardiovascular risk and cancer mortality. *Eur J Intern Med.* 2020; 77: 79-85.
125. Fiotti N, Altamura N, Cappelli C, Schillan M, Guarnieri G, Giansante C. Long term prognosis in patients with peripheral arterial disease treated with antiplatelet agents. *Eur J Vasc Endovasc Surg.* 2003; 26 (4): 374-80.
126. Rantner B, Kollerits B, Pohlhammer J, Stadler M, Lamina C, Peric S, et al. The fate of patients with intermittent claudication in the 21st century revisited - results from the CAVASIC Study. *Sci Rep.* 2017; 8: 45833.
127. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. *Circulation.* 2016; 133 (11): 1104-14.
128. Raghavan S, Vassy JL, Ho Y-L, Song RJ, Gagnon DR, Cho K, et al. Diabetes Mellitus-Related All-Cause and Cardiovascular Mortality in a National Cohort of Adults. *J Am Heart Assoc.* 2019; 8 (4): e011295.
129. Aboyans V, Criqui MH, Denenberg JO, Knoke JD, Ridker PM, Fronck A. Risk factors for progression of peripheral arterial disease in large and small vessels. *Circulation.* 2006; 113 (22): 2623-9.
130. Ferraresi R, Mauri G, Losurdo F, Troisi N, Brancaccio D, Caravaggi C, et al. BAD transmission and SAD distribution: a new scenario for critical limb ischemia. *J Cardiovasc Surg (Torino).* 2018; 59 (5): 655-64.
131. Ahmad J. The diabetic foot. *Diabetes Metab Syndr.* 2016; 10 (1): 48-60.
132. Spreen MI, Gremmels H, Teraa M, Sprengers RW, Verhaar MC, Stadius van Eps RG, et al. Diabetes Is Associated With Decreased Limb Survival in Patients With Critical Limb Ischemia: Pooled Data From Two Randomized Controlled Trials. *Diabetes Care.* 2016; 39 (11): 2058-64.

133. Lee H-F, Chen S-W, Liu J-R, Li P-R, Wu L-S, Chang S-H, et al. Major adverse cardiovascular and limb events in patients with diabetes and concomitant peripheral artery disease treated with sodium glucose cotransporter 2 inhibitor versus dipeptidyl peptidase-4 inhibitor. *Cardiovasc Diabetol*. 2020; 19 (1): 160.
134. Bonaca MP, Bauersachs RM, Anand SS, Debus ES, Nehler MR, Patel MR, et al. Rivaroxaban in Peripheral Artery Disease after Revascularization. *N Engl J Med*. 2020; 382 (21): 1994-2004.
135. Kaplovitch E, Eikelboom JW, Dyal L, Aboyans V, Abola MT, Verhamme P, et al. Rivaroxaban and Aspirin in Patients With Symptomatic Lower Extremity Peripheral Artery Disease: A Subanalysis of the COMPASS Randomized Clinical Trial. *JAMA Cardiol*. 2021; 6 (1): 21-9.
136. Paul SK, Bhatt DL, Montvida O. The association of amputations and peripheral artery disease in patients with type 2 diabetes mellitus receiving sodium-glucose cotransporter type-2 inhibitors: real-world study. *Eur Heart J*. 2021; 42 (18): 1728-38.
137. Patoulias D, Papadopoulos C, Doumas M. Updated meta-analysis assessing the risk of amputation with sodium-glucose co-transporter-2 inhibitors in the hallmark cardiovascular and renal outcome trials. *Diabetes Obes Metab*. 2021; 23 (4): 1063-5.
138. Song P, Rudan D, Zhu Y, Fowkes FJI, Rahimi K, Fowkes FGR, et al. Global, regional, and national prevalence and risk factors for peripheral artery disease in 2015: an updated systematic review and analysis. *Lancet Glob Health*. 2019; 7 (8): e1020-30.
139. Heart Outcomes Prevention Evaluation Study Investigators, Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, et al. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. *N Engl J Med*. 2000; 342 (3): 145-53.
140. Bangalore S, Fakhri R, Toklu B, Ogedegbe G, Weintraub H, Messerli FH. Angiotensin-Converting Enzyme Inhibitors or Angiotensin Receptor Blockers in Patients Without Heart Failure? Insights From 254,301 Patients From Randomized Trials. *Mayo Clin Proc*. 2016; 91 (1): 51-60.
141. Vlachopoulos C, Terentes-Printzios D, Aboyans V, Brodmann M, De Carlo M, Tousoulis D. Angiotensin converting enzyme inhibitors and walking distance: Have we walked the whole distance? *Atherosclerosis*. 2016; 252: 199-200.

142. Potier L, Roussel R, Elbez Y, Marre M, Zeymer U, Reid CM, et al. Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers in high vascular risk. *Heart*. 2017; 103 (17): 1339-46.
143. Bhatt DL, Steg PG, Ohman EM, Hirsch AT, Ikeda Y, Mas J-L, et al. International prevalence, recognition, and treatment of cardiovascular risk factors in outpatients with atherothrombosis. *JAMA*. 2006; 295 (2): 180-9.
144. Bangalore S, Fakhri R, Toklu B, Messerli FH. In Reply-The Different Effects of Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers on Mortality. *Mayo Clin Proc*. 2016; 91 (7): 972-5.
145. Li ECK, Heran BS, Wright JM. Angiotensin converting enzyme (ACE) inhibitors versus angiotensin receptor blockers for primary hypertension. *Cochrane Database Syst Rev*. 2014; (8): CD009096.
146. Barnett AH, Bain SC, Bouter P, Karlberg B, Madsbad S, Jervell J, et al. Angiotensin-receptor blockade versus converting-enzyme inhibition in type 2 diabetes and nephropathy. *N Engl J Med*. 2004; 351 (19): 1952-61.
147. Pfeffer MA, McMurray JJV, Velazquez EJ, Rouleau J-L, Køber L, Maggioni AP, et al. Valsartan, captopril, or both in myocardial infarction complicated by heart failure, left ventricular dysfunction, or both. *N Engl J Med*. 2003; 349 (20): 1893-906.
148. Messerli FH, Bangalore S, Bavishi C, Rimoldi SF. Angiotensin-Converting Enzyme Inhibitors in Hypertension: To Use or Not to Use? *J Am Coll Cardiol*. 2018; 71 (13): 1474-82.
149. Subherwal S, Patel MR, Kober L, Peterson ED, Bhatt DL, Gislason GH, et al. Peripheral artery disease is a coronary heart disease risk equivalent among both men and women: results from a nationwide study. *Eur J Prev Cardiol*. 2015; 22 (3): 317-25.
150. ONTARGET Investigators, Yusuf S, Teo KK, Pogue J, Dyal L, Copland I, et al. Telmisartan, ramipril, or both in patients at high risk for vascular events. *N Engl J Med*. 2008; 358 (15): 1547-59.
151. Osgood MJ, Harrison DG, Sexton KW, Hocking KM, Voskresensky IV, Komalavilas P, et al. Role of the renin-angiotensin system in the pathogenesis of intimal hyperplasia: therapeutic potential for prevention of vein graft failure? *Ann Vasc Surg*. 2012; 26 (8): 1130-44.

152. Schmieder RE, Delles C, Mimran A, Fauvel JP, Ruilope LM. Impact of telmisartan versus ramipril on renal endothelial function in patients with hypertension and type 2 diabetes. *Diabetes Care*. 2007; 30 (6): 1351-6.
153. Heinze G, Jüni P. An overview of the objectives of and the approaches to propensity score analyses. *Eur Heart J*. 2011; 32 (14): 1704-8.
154. Eikelboom JW, Connolly SJ, Bosch J, Dagenais GR, Hart RG, Shestakovska O, et al. Rivaroxaban with or without Aspirin in Stable Cardiovascular Disease. *N Engl J Med*. 2017; 377 (14): 1319-30.
155. Krzywinski M, Altman N. Points of significance: error bars. *Nat Methods*. 2013; 10 (10): 921-2.
156. Knuuti J, Wijns W, Saraste A, Capodanno D, Barbato E, Funck-Brentano C, et al. 2019 ESC Guidelines for the diagnosis and management of chronic coronary syndromes. *Eur Heart J*. 2020; 41 (3): 407-77.
157. Fox KAA, Anand SS, Aboyans V, Cowie MR, Debus ES, Zeymer U, et al. Xarelto plus Acetylsalicylic acid: Treatment patterns and Outcomes in patients with Atherosclerosis (XATO): Rationale and design of a prospective registry study to assess rivaroxaban 2.5 mg twice daily plus aspirin for prevention of atherothrombotic events in coronary artery disease, peripheral artery disease, or both. *Am Heart J*. 2020; 222: 166-73.
158. Darmon A, Bhatt DL, Elbez Y, Aboyans V, Anand S, Bosch J, et al. External applicability of the COMPASS trial: an analysis of the reduction of atherothrombosis for continued health (REACH) registry. *Eur Heart J*. 2018; 39 (9): 750-757a.
159. Würtz M, Olesen KKW, Thim T, Kristensen SD, Eikelboom JW, Maeng M. External applicability of the COMPASS trial: the Western Denmark Heart Registry. *Eur Heart J Cardiovasc Pharmacother*. 2019; 5 (4): 192-9.
160. Desperak P, Hudzik B, Gąsior M. Assessment of patients with coronary artery disease who may benefit from the use of rivaroxaban in the real world: implementation of the COMPASS trial criteria in the TERCET registry population. *Pol Arch Intern Med*. 2019; 129 (7-8): 460-8.

161. Faria D, Santos M, Baptista SB, Ferreira J, Leal P, Abreu PFE, et al. Eligibility for extended antithrombotic therapy for secondary prevention of acute coronary syndrome. *Rev Port Cardiol.* 2020; 39 (9): 493-501.
162. Schiele F, Puymirat E, Ferrières J, Simon T, Fox KAA, Eikelboom J, et al. The FAST-MI 2005-2010-2015 registries in the light of the COMPASS trial: The COMPASS criteria applied to a post-MI population. *Int J Cardiol.* 2019; 278: 7-13.
163. Björck M, Earnshaw JJ, Acosta S, Bastos Gonçalves F, Cochenec F, Debus ES, et al. Editor's Choice - European Society for Vascular Surgery (ESVS) 2020 Clinical Practice Guidelines on the Management of Acute Limb Ischaemia. *Eur J Vasc Endovasc Surg.* 2020; 59 (2): 173-218.
164. Lapébie F-X, Aboyans V, Lacroix P, Constans J, Boulon C, Messas E, et al. Editor's Choice - External Applicability of the COMPASS and VOYAGER-PAD Trials on Patients with Symptomatic Lower Extremity Artery Disease in France: The COPART Registry. *Eur J Vasc Endovasc Surg.* 2021; 62 (3): 439-49.
165. Mahajan A, Brunson A, White R, Wun T. The Epidemiology of Cancer-Associated Venous Thromboembolism: An Update. *Semin Thromb Hemost.* 2019; 45 (4): 321-5.
166. Khorana AA, DeSancho MT, Liebman H, Rosovsky R, Connors JM, Zwicker J. Prediction and Prevention of Cancer-Associated Thromboembolism. *Oncologist.* 2021; 26 (1): e2-7.
167. Khorana AA, Connolly GC. Assessing risk of venous thromboembolism in the patient with cancer. *J Clin Oncol.* 2009; 27 (29): 4839-47.
168. Heit JA, Mohr DN, Silverstein MD, Petterson TM, O'Fallon WM, Melton LJ. Predictors of recurrence after deep vein thrombosis and pulmonary embolism: a population-based cohort study. *Arch Intern Med.* 2000; 160 (6): 761-8.
169. Abdulla A, Davis WM, Ratnaweera N, Szefer E, Ballantyne Scott B, Lee AYY. A Meta-Analysis of Case Fatality Rates of Recurrent Venous Thromboembolism and Major Bleeding in Patients with Cancer. *Thromb Haemost.* 2020; 120 (4): 702-13.
170. Trujillo-Santos J, Martos FM, Font C, Farge-Bancel D, Rosa V, Lorenzo A, et al. Analysis of clinical factors affecting the rates of fatal pulmonary embolism and bleeding in cancer patients with venous thromboembolism. *Heliyon.* 2017; 3 (1): e00229.

171. Farge D, Frere C, Connors JM, Ay C, Khorana AA, Munoz A, et al. 2019 international clinical practice guidelines for the treatment and prophylaxis of venous thromboembolism in patients with cancer. *Lancet Oncol.* 2019; 20 (10): e566-81.
172. Key NS, Khorana AA, Kuderer NM, Bohlke K, Lee AYY, Arcelus JJ, et al. Venous Thromboembolism Prophylaxis and Treatment in Patients With Cancer: ASCO Clinical Practice Guideline Update. *J Clin Oncol.* 2020; 38 (5): 496-520.
173. Lyman GH, Carrier M, Ay C, Di Nisio M, Hicks LK, Khorana AA, et al. American Society of Hematology 2021 guidelines for management of venous thromboembolism: prevention and treatment in patients with cancer. *Blood Adv.* 2021; 5 (4): 927-74.
174. Prandoni P, Lensing AWA, Piccioli A, Bernardi E, Simioni P, Girolami B, et al. Recurrent venous thromboembolism and bleeding complications during anticoagulant treatment in patients with cancer and venous thrombosis. *Blood.* 2002; 100 (10): 3484-8.
175. Kyrle PA. Predicting recurrent venous thromboembolism in cancer: is it possible? *Thromb Res.* 2014; 133 Suppl 2: S17-22.
176. Ginsberg JS, Hirsh J, Julian J, Vander LaandeVries M, Magier D, MacKinnon B, et al. Prevention and treatment of postphlebotic syndrome: results of a 3-part study. *Arch Intern Med.* 2001; 161 (17): 2105-9.
177. 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; 3 (4): 692-4.
178. Walker AJ, Card TR, West J, Crooks C, Grainge MJ. Incidence of venous thromboembolism in patients with cancer - a cohort study using linked United Kingdom databases. *Eur J Cancer.* 2013; 49 (6): 1404-13.
179. Baraldi AN, Enders CK. An introduction to modern missing data analyses. *J Sch Psychol.* 2010; 48 (1): 5-37.
180. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med.* 1991; 10 (4): 585-98.
181. van der Hulle T, den Exter PL, van den Hoven P, van der Hoeven JJ, van der Meer FJM, Eikenboom J, et al. Cohort Study on the Management of Cancer-Associated Venous

Thromboembolism Aimed at the Safety of Stopping Anticoagulant Therapy in Patients Cured of Cancer. *Chest*. 2016; 149 (5): 1245-51.

182. Louzada ML, Majeed H, Dao V, Wells PS. Risk of recurrent venous thromboembolism according to malignancy characteristics in patients with cancer-associated thrombosis: a systematic review of observational and intervention studies. *Blood Coagul Fibrinolysis*. 2011; 22 (2): 86-91.

183. Donadini MP, Ageno W, Antonucci E, Cosmi B, Kovacs MJ, Le Gal G, et al. Prognostic significance of residual venous obstruction in patients with treated unprovoked deep vein thrombosis: a patient-level meta-analysis. *Thromb Haemost*. 2014; 111 (1): 172-9.

184. Becattini C, Giustozzi M, Cerdà P, Cimini LA, Riera-Mestre A, Agnelli G. Risk of recurrent venous thromboembolism after acute pulmonary embolism: Role of residual pulmonary obstruction and persistent right ventricular dysfunction. A meta-analysis. *J Thromb Haemost*. 2019; 17 (8): 1217-28.

185. Stain M, Schönauer V, Minar E, Bialonczyk C, Hirschl M, Weltermann A, et al. The post-thrombotic syndrome: risk factors and impact on the course of thrombotic disease. *J Thromb Haemost*. 2005; 3 (12): 2671-6.

186. Brunson A, Ho G, White R, Wun T. Inferior vena cava filters in patients with cancer and venous thromboembolism (VTE) does not improve clinical outcomes: A population-based study. *Thromb Res*. 2017; 153: 57-64.

187. Ejaz U, Akhtar F, Xue J, Wan X, Zhang T, He S. Review: Inhibitory potential of low molecular weight Heparin in cell adhesion; emphasis on tumor metastasis. *Eur J Pharmacol*. 2021; 892: 173778.

188. Mulder FI, Bosch FTM, Young AM, Marshall A, McBane RD, Zemla TJ, et al. Direct oral anticoagulants for cancer-associated venous thromboembolism: a systematic review and meta-analysis. *Blood*. 2020; 136 (12): 1433-41.

Résumé en français

Les modèles statistiques multivariés ont des applications variées en médecine. Ils peuvent être descriptifs, capturant l'association entre une variable dépendante et des variables indépendantes, explicatifs, testant une hypothèse causale entre une exposition et un événement d'intérêt, ou prédictifs, calculant pour un individu la probabilité d'un diagnostic ou de la survenue d'un événement. Ces modèles sont majoritairement dérivés d'études de cohortes. Pour leur développement, le choix du type de modèle, la sélection des variables pour le modèle final et l'évaluation de sa performance (validation interne et externe) ont une importance particulière. Les données manquantes causent un biais dans l'estimation des paramètres et les tests de significativité, et limitent le nombre de patients pour l'analyse multivariée ; des méthodes d'imputation existent. Dans ce travail, ces différentes étapes ont été appliquées à deux cohortes de maladies vasculaires. Des modèles de survie ont été utilisés : le modèle de Cox, et le modèle de Fine et Gray, qui prend en compte les risques compétitifs. La cohorte COPART (cohorte de patients artériopathes) est une cohorte nationale multicentrique prospective s'intéressant aux patients hospitalisés pour la prise en charge d'une artériopathie athéromateuse des membres inférieurs symptomatique. Deux modèles explicatifs ont été dérivés : un premier évaluant l'impact du diabète sur le risque de décès de cause cardiovasculaire, d'infarctus du myocarde et d'accident vasculaire cérébral ischémique, et un second évaluant l'impact des antagonistes des récepteurs de l'angiotensine de type 2 sur le risque de décès. Un troisième modèle, descriptif, a été constitué, pour identifier les facteurs associés au risque d'amputation majeure du membre inférieur. La sélection des variables pour les modèles explicatifs a été réalisée sur les données de la littérature, indépendamment des données de la cohorte, avec une représentation des liens de causalité entre variables par un graphe acyclique dirigé, guidant ensuite la sélection. Les variables du modèle descriptif ont été sélectionnées sur les données de la cohorte, à l'aide d'un algorithme de sélection descendante. Enfin, la validité externe, notamment la transposabilité, de deux études contrôlées randomisées a été évaluée dans COPART. La cohorte RIETE (*registro informatizado de la enfermedad tromboembolica venosa*) est une cohorte internationale multicentrique prospective incluant les patients présentant une embolie pulmonaire ou une thrombose veineuse profonde. Le sous-groupe des patients présentant un événement thromboembolique veineux associé à un cancer a fait l'objet de ce travail. Une imputation multiple des données manquantes par équations chaînées a été réalisée. Les facteurs associés à la récurrence à l'arrêt d'un traitement anticoagulant d'une durée d'au moins 3 mois ont ensuite été identifiés, à l'aide d'une méthode mixte : dans la 1ère étape, les variables connues dans la littérature pour leur association au risque de récurrence ont été sélectionnées indépendamment des résultats dans la cohorte, puis dans la 2ème étape, une sélection descendante a été appliquée à partir du modèle plein, en forçant dans le modèle descriptif final les variables sélectionnées dans la 1ère étape. A partir de ces résultats, un modèle prédictif du risque de récurrence d'événement thromboembolique veineux a été sélectionné. La performance du modèle a été évaluée dans les données initiales avec imputation des valeurs manquantes, puis une validation interne a été réalisée par rééchantillonnage par *bootstrap* ; l'optimisme des différents paramètres de la validation interne (pente de la droite de calibration, statistique D de Royston et Sauerbrei pour la discrimination, R_D^2 de Royston et Sauerbrei pour la part de variance de la variable dépendante expliquée par le modèle) a ainsi été évalué, puis retranché à la performance apparente, tout en tenant compte de la présence de données imputées.

Résumé en anglais

Multivariate statistical models are widely used in medicine. Descriptive models capture the association between dependent and independent variables, explanatory models test a causal hypothesis between an exposition and an outcome, and predictive models predict a diagnosis or an outcome for a given patient. Multivariate statistical models are mainly derived from cohort studies. For their development, the choice of the type of model, variable selection for the final model, and the evaluation of its performance (internal and external validation) are of particular importance. Missing data can lead to biased estimates of parameters and incorrect significance tests, and limit the multivariate analysis to the complete cases; multiple imputation is a technique for handling missing data. In this work, these different steps were applied to two cohorts of vascular diseases. Survival analyses were used: Cox regression model, and Fine and Gray model, which handles competitive risks. The COPART (cohorte de patients artériopathes) cohort is an ongoing French, multicenter, observational, cohort study, prospectively collecting data on patients hospitalized for symptomatic lower extremity artery disease. Two explanatory models were derived: one to evaluate the relationship between diabetes and major adverse cardiovascular events, and one to assess the relationship between angiotensin-receptor blockers and mortality. A descriptive model was also derived to identify factors associated with lower limb major amputation. Background knowledge and a directed acyclic graph were used for the variable selection for explanatory models, instead of data-driven methods. Backward elimination variable selection algorithm starting was used for the descriptive model. Finally, external validity, especially transportability, of two randomized controlled trials was studied in COPART. The RIETE cohort (*Registro Informatizado de la Enfermedad ThromboEmbolica venosa*) is an ongoing, international, multicenter, prospective cohort study of patients diagnosed with symptomatic acute deep vein thrombosis or pulmonary embolism. Patients with cancer-associated thrombosis were studied. A multiple imputation by chained equations was performed on missing values. The factors associated with recurrence after a time-limited anticoagulant therapy of at least 3 months were identified with a two-tier predictor selection process: the first tier included variables chosen based on knowledge regarding their relationship with recurrence regardless the unadjusted analyses, and the variables of the second tiers were identified by multivariate Cox model with backward stepwise elimination. With these results, a multivariable prognostic model for the risk of recurrence after anticoagulant discontinuation was developed. The model performance was estimated directly from the data set with multiple imputation of missing data, also used to develop the prediction model. Then, a bootstrap internal validation approach was performed, with the computation of the optimism for the performance measures (slope of the calibration plot, Royston and Sauerbrei's D statistic for discrimination, and Royston and Sauerbrei's R_D^2 for explained variation). These optimisms were subtracted from the apparent performance measures obtained from the original data set, taking into account the multiple imputation of missing data.