



Société Française de
Pharmacologie et de Thérapeutique

Groupe de Travail Méthodologie

Livre blanc SFPT

De la nécessité de la méthodologie
dans l'évaluation des médicaments

Document compagnon

Dossier 8 - La balance bénéfico-risk

Groupe de rédaction / relecture (Par ordre alphabétique)

- Theodora Angoulvant
- Laurent Bertolotti
- Jean-Luc Cracowski
- Michel Cucherat
- Dominique Deplanque
- Guillaume Grenet
- François Gueyffier
- Behrouz Kassai
- Charles Khouri
- Silvy Laporte
- Bruno Laviolle
- Jean-Christophe Lega
- Clara Locher
- Florian Naudet
- Edouard Ollier
- Antoine Pariente
- Matthieu Roustit
- Tabassome Simon



[Licence Creative Commons](#)

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International

Vous êtes autorisé à :

- Partager — copier, distribuer et communiquer le matériel par tous moyens et sous tous formats
- Adapter — remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale.

Table des matières

1	Introduction.....	7
2	L'impossible démonstration de l'absence d'effet indésirable	8
2.1	Problématique.....	8
2.2	Utilisation de l'intervalle de confiance.....	9
2.3	Incertitude liée à la taille des essais.....	10
3	Les essais de non-infériorité de safety.....	12
4	Conclure à l'existence d'un effet indésirable	14
4.1	Le risque d'inflation des faux positifs en safety	14
5	Comment décider en présence d'un surcroit d'effets indésirables	16
6	L'approche quantitative de la balance bénéfice risque	19
6.1	Appréciation du bénéfice clinique net par les différences des risques.....	19
6.2	Graphique de Cates	24
6.3	Appréciation du bénéfice clinique net par les NNT et NNH.....	25
6.4	Critère composite de bénéfice clinique net	26
6.5	Problématique de l'utilisation du bénéfice net en non-infériorité	27
6.5.1	La mortalité totale comme critère de bénéfice clinique net.....	28
6.5.2	Utilisation d'un critère de bénéfice net comme critère de jugement principal	29
7	Variation de la balance bénéfice risque dans les strates de patients	30
8	Approche qualitative de la balance bénéfice risque	34
9	Références.....	35

1 Introduction

Ce dossier concerne uniquement le questionnement sur la safety tel qu'il se pose à la fin des études pivots d'un nouveau produit. Il s'agit de l'évaluation initiale de la balance bénéfice risque pour déterminer si celle-ci est favorable. Comme il est possible que des effets indésirables soient identifiés après utilisation du produit (pharmacovigilance), l'appréciation de la balance bénéfice risque est susceptible d'évoluer dans le temps.

Pour un traitement il ne suffit pas d'être efficace, encore faut-il qu'il soit sûr. La balance bénéfice risque est le principal critère d'intérêt clinique du traitement. L'efficacité, évaluée de façon isolée, ne suffit pas pour conclure à un réel intérêt clinique d'un nouveau traitement.

L'évaluation de la balance bénéfice risque est une étape cruciale avant d'introduire un nouveau traitement dans la stratégie thérapeutique. De nombreuses molécules échouent, non pas en raison d'un manque d'efficacité, mais bien en raison d'un profil de tolérance questionnable. Cependant, malgré la simplicité de l'adage qu'un traitement doit être efficace et sûr, l'évaluation de la balance bénéfice risque est très souvent complexe (cf. section 2). Des lignes directrices claires sont cependant identifiables pour guider la lecture critique sur ce point.

Pour les traitements majeurs, apportant des bénéfices cliniquement pertinents dans des maladies graves, seuls les événements indésirables de même gravité clinique que les événements évités/retardés par le traitement sont en mesure de contrebalancer le bénéfice (cf. section 6). En revanche, avec des traitements plus à visée symptomatique, tous les types d'effets indésirables sont susceptibles de remettre en cause l'intérêt médical du traitement (cf. section 8).

L'évaluation de la safety et de la balance bénéfice risque a fait l'objet de nombreux travaux théoriques et propositions de frameworks, de méthodes, de métriques, etc. [1]. Le projet européen PROTECT en a fait une revue exhaustive (<https://protectbenefitrisk.eu/index.html>). En pratique, relativement peu de ces propositions sont utilisées et l'évaluation actuelle de la balance bénéfice risque repose finalement sur une approche relativement simple qui raisonne sur un plan quantitatif [2] et un plan qualitatif. C'est ce que nous développerons dans ce dossier.

2 L'impossible démonstration de l'absence d'effet indésirable

Classiquement, la question de la balance bénéfice risque implique de montrer que le traitement est sûr (en suivant l'adage qu'il ne suffit pas qu'un traitement soit efficace, il faut qu'il soit aussi sûr).

2.1 Problématique

La conclusion à l'absence d'effet indésirable notable lors de l'analyse d'un essai est souvent faites de façon rudimentaire, à partir du constat que la fréquence du critère considérée est similaire entre les 2 groupes (non augmentée dans le groupe traité par rapport au groupe contrôle), sans plus de formalisme.

Dans l'essai TRILOGY ACS [3] comparant à la phase précoce des syndromes coronariens aigus le prasugrel au clopidogrel, il est conclu que « rates of severe and intracranial bleeding were similar in the two groups » à partir des résultats de safety suivant :

	Prasugrel (N=3590)		Clopidogrel (N=3590)		Hazard Ratio (95% CI)	P Value
	Patients with Event	Event Rate at 30 Mo	Patients with Event	Event Rate at 30 Mo		
	no. (%)	% (95% CI)	no. (%)	% (95% CI)		
GUSTO criteria						
Severe or life-threatening	13 (0.4)	0.9 (0.1-1.7)	14 (0.4)	0.6 (0.3-1.0)	0.94 (0.44-1.99)	0.87
Severe or life-threatening or moderate	52 (1.4)	2.5 (1.5-3.4)	35 (1.0)	1.7 (1.0-2.3)	1.50 (0.98-2.30)	0.06
TIMI criteria†						
Major	39 (1.1)	2.1 (1.1-3.0)	30 (0.8)	1.5 (0.9-2.1)	1.31 (0.81-2.11)	0.27
Life-threatening	16 (0.4)	0.9 (0.1-1.6)	17 (0.5)	0.8 (0.4-1.2)	0.95 (0.48-1.87)	0.88
Fatal	4 (0.1)	0.5 (0.0-1.2)	4 (0.1)	0.2 (0.0-0.5)	1.01 (0.25-4.05)	0.99
Intracranial hem-orrhage	8 (0.2)	0.7 (0.0-1.5)	12 (0.3)	0.5 (0.2-0.8)	0.67 (0.28-1.65)	0.39
Major or minor	70 (1.9)	3.3 (2.3-4.4)	46 (1.3)	2.1 (1.4-2.8)	1.54 (1.06-2.23)	0.02

Cette conclusion semble donc avoir été faite à partir du constat que les estimations ponctuelles de la fréquence des saignements sévères (1^{ère} ligne de résultat) sont « similaires » entre les 2 groupes (0.4% et 0.4%)

Cependant l'incertitude statistique existant en général sur les données de safety rend ce raisonnement extrêmement fragile et susceptible de conclure facilement à l'absence de surcroît d'évènements indésirable à tort. En effet, compte tenu de la rareté des évènements considérés, ne pas constater de différence numérique notable n'apporte à peu près aucune garantie de l'absence réelle de surrisque.

Dans d'autres cas, un test statistique est effectué et un résultat non statistiquement significatif est alors invoqué pour conclure à l'absence de surrisque.

de confiance représente la meilleure façon de communiquer l'incertitude existant sur l'interprétation des données de safety

2.3 Incertitude liée à la taille des essais

À l'issue des essais pivots, compte tenu de la petitesse relative des effectifs, il est possible que des effets indésirables particuliers ne soient pas identifiés. Une certaine incertitude existe sur le risque lors de l'appréciation initiale de la balance bénéfice risque. Il est donc nécessaire à ce stade que les données d'efficacité et de risque issues des essais pivots assurent que la balance bénéfice risque est très favorable, afin qu'elle puisse le rester même en cas de découverte ultérieure de nouvelles problématiques de sécurité. Pour cela il faut un bénéfice cliniquement pertinent parfaitement démontré et s'assurer que les effets indésirables attendus (effets latéraux attendus) ou déjà connus ne sont pas en mesure de contrebalancer ce bénéfice.

Par exemple avec une immunothérapie en oncologie, qui est susceptible de donner des effets indésirables immunomédiés potentiellement mortels, une simple démonstration sur la survie sans progression ne permet pas de garantir une balance bénéfice risque favorable. Cependant, si une amélioration de la survie globale est démontrée, la balance B/R est assurée d'être positive (si le traitement n'apporte pas une dégradation importante de la qualité de vie).

Lorsque ces conditions sont remplies, la question des effets indésirables non détectée n'a alors presque plus d'importance. Les effets indésirables non détectés dans ces études pivots en raison de la petitesse relative des effectifs sont de ce fait trop peu fréquents pour pouvoir remettre en cause quantitativement la balance bénéfice risque (cf. section 6). Dans ces conditions, il est assuré que si un effet indésirable rare est identifié à postériori, cette découverte ne remettra pas en cause la balance bénéfice risque et ne fera pas regretter à postériori l'adoption du traitement dans la stratégie thérapeutique. En revanche, si le bénéfice est mal établi (sur un critère intermédiaire et non pas un critère clinique par exemple), la même situation mettra en évidence que, par manque d'évaluation initiale, l'introduction prématurée du traitement dans la stratégie thérapeutique n'aura pas été profitable aux patients.

Si l'essai avait comme objectif de démontrer un bénéfice clinique pertinent, le manque de puissance pour la recherche des événements indésirables rares n'est pas très problématique, car les effets indésirables qui pourrait remettre en cause quantitativement ce bénéfice doivent d'être de la fréquence que le critère d'efficacité pour conduire à une différence de risque suffisamment importante pour contrecarrer celle obtenue sur l'efficacité. Ou si leur fréquence est plus faible, l'augmentation doit être très importante pour conduire à une différence des risques de même grandeur. Dans ce cas, il est probable que la précision statistique soit suffisante avec la calibration faite pour l'efficacité. Par contre si la calibration de l'étude a été faite sur un critère peu pertinent, ce raisonnement ne tient plus, car l'essai ne permettra pas de documenter le réel bénéfice clinique (pertinent).

Un exemple historique de ce point est apporté par la cévistatine et les rhabdomyolyses [5]. Cette statine avait été introduite sur le marché sans essai de morbi-mortalité, sur la base de son effet sur le LDL cholestérol principalement. La survenue de rhabdomyolyses a conduit au retrait de ce produit, car il était impossible de savoir si le bénéfice permettait éventuellement de rendre acceptable cet effet indésirable grave. L'observation du même effet indésirable avec les statines ayant parfaitement bien démontré leur

3 Les essais de non-infériorité de safety

La recherche d'une conclusion formelle d'absence d'effet indésirable passe, en théorie, par la réalisation d'un essai de non-infériorité. Cette approche reste plutôt rare et elle n'a été utilisée de manière systématique que pour les nouveaux antidiabétiques afin d'établir leur sécurité cardiovasculaire² [6, 7, 8].

Ces essais permettent de démontrer que le nouveau traitement n'entraîne pas une augmentation d'un effet indésirable supérieure à une certaine valeur, la limite de non-infériorité³. Cette limite devrait correspondre au surrisque le plus important acceptable/tolérable compte tenu du bénéfice du produit. L'essai permettra de conclure à la non-infériorité lorsque la borne supérieure de l'intervalle de confiance sera inférieure à cette limite.

Les essais de non-infériorité ne permettent pas stricto sensu de conclure à l'absence d'augmentation du risque, mais ils permettent de conclure que, si augmentation du risque il y a, cette augmentation reste acceptable compte tenu du bénéfice du produit.

Le maillon faible de cette approche réside dans la fixation de la limite de non-infériorité qui est arbitraire et repose sur un choix de valeur « quelle augmentation de cet événement indésirable accepter compte tenu du bénéfice ? ». La réponse à cette question n'est jamais évidente et aucune valeur standard ne peut être proposée. Le choix repose sur un raisonnement clinique au cas par cas.

⋈ Dans le cas des essais de safety cardiovasculaire des nouveaux antidiabétiques demandés par la FDA, cette limite a été fixée à un risque ratio de 1.3 (sur le critère composite d'évènements cardiovasculaires, infarctus et AVC, mortels et non mortels). Avec ce choix, un traitement sera donc considéré comme sûr dès que l'essai de non-infériorité permettra d'exclure qu'il n'entraîne une augmentation des évènements cardiovasculaires de plus de 30%. Se pose alors la question de la signification clinique d'une valeur aussi importante. Peut-on vraiment considérer comme sûr un traitement qui est susceptible d'entraîner une telle augmentation d'évènements graves ? Faire courir ce risque potentiel aux patients est-il justifiable par le bénéfice qu'apportent ces traitements, surtout qu'un bénéfice recherché par le contrôle glycémique est la réduction des évènements cardiovasculaire qui est alors loin d'être démontrée si la limite supérieure de l'intervalle de confiance permet juste d'écarter un risque ratio de 1.3.

L'utilisation d'un traitement actif à la place du placebo comme comparateur est aussi possible si la question est celle d'une sécurité relative par rapport au traitement standard, pour écarter formellement la possibilité que le nouveau traitement puisse être moins sûr que le standard qu'il vise à remplacer.

⋈ L'essai PRO₂TECT a évalué la safety cardiovasculaire du vadadustat par rapport à la darbepoetin alfa chez des patients porteurs d'une pathologie rénale sans dialyse (non-dialysis-dependent chronic kidney disease, NDD-CKD) [9]. Deux critères de jugement ont été utilisés : les évènements cardiovasculaires (MACE) et un critère composite MACE et hospitalisation. Une limite de non-infériorité de 1.25, non justifiée dans la publication, a été utilisée La non-infériorité n'a pas été démontrée sur le premier critère de safety avec un hazard ratio de 1.17 (IC 95% 1.01 ; 1.36) dont l'intervalle de confiance excède la limite de 1.25. Ce résultat a bloqué l'enregistrement de ce produit dans cette indication par la FDA.

² <https://www.fda.gov/media/71297/download>

³ Cf. [Dossier 6 – L'essai de non-infériorité](#)

L'essai CARES a été entrepris pour confirmer la bonne sécurité cardiovasculaire du febuxostat dans la prévention de la crise de goutte [10]. Cette étude a été réalisée dans le but de lever un doute sur la sécurité cardiovasculaire du febuxostat induit par son essai pivot qui faisait craindre une augmentation des événements cardiovasculaires par rapport à l'allopurinol [11, 12]. L'essai CARES permet de conclure à la non-infériorité recherchée (limite supérieure de l'intervalle de confiance du hazard ratio à 1.23 pour une limite de non-infériorité de 1.3) mais une surmortalité cardiovasculaire (hazard ratio 1.34 IC95% [1.03 ; 1.73]) et de toute cause (1.22 IC95% [1.01 ; 1.47];) est observée ne permettant pas de conclure à la sécurité de cette molécule par rapport à l'allopurinol.

D'autres exemples d'application de cette approche formelle d'évaluation de la safety existent [13, 14, 15, 16] (y compris dans le domaine des dispositifs médicaux ou de la chirurgie [14, 17, 18]) mais pour l'instant (novembre 2022) nous ne disposons d'aucune étude de méta-recherche sur l'utilisation de cette approche et la description des pratiques.

Même si cette approche basée sur l'essai de non-infériorité est assez peu employée, elle met en exergue l'intérêt du raisonnement de non-infériorité pour solidifier les conclusions sur la safety dans les essais thérapeutiques. Comme il est impossible de conclure à l'absence d'effet, il convient de toujours relativiser l'incertitude existant sur l'absence de surrisque par rapport au bénéfice démontré apporté par le traitement. Ainsi il ne convient plus de chercher à conclure à l'absence d'effet indésirable, mais bien de s'assurer que l'on peut exclure la possibilité d'un surrisque qui ne serait pas acceptable compte tenu du bénéfice qu'apporte le traitement. En d'autres termes, il s'agit de s'assurer qu'au pire, le surcroît non excluible d'événements indésirables ne contrebalance pas en totalité le bénéfice apporté. C'est la base du raisonnement quantitatif, basé sur le bénéfice clinique net, qui apparaît de plus en plus comme étant la voie à suivre pour rendre moins arbitraires les conclusions des essais cliniques sur la balance bénéfice risque. Cette approche est développée section 6.

4 Conclure à l'existence d'un effet indésirable

Conclure à l'existence d'un effet indésirable⁴ est bien plus aisé que de conclure à son absence. En effet, pour la safety, le principe de précaution prévaut. À partir du moment où il existe un doute suffisamment important sur la possibilité d'un surcroît d'un évènement indésirable gênant, la décision est prise, comme si ce surrisque était certain.

En effet, contrairement à l'efficacité où des preuves formelles de l'effet bénéfique doivent être disponibles avant de décider d'utiliser le traitement, au niveau de la safety, il ne serait pas raisonnable de ne prendre aucune décision vis-à-vis d'une suspicion d'effet indésirable en prétendant que rien n'est formellement démontré.

Par exemple, il serait très difficile de recommander un traitement pour lequel des décès toxiques sont suspectés dans l'essai pivot sous prétexte que ce surrisque n'est pas formellement prouvé et qu'il faut attendre des données le démontrant parfaitement (que cet essai ne pourra pas apporter, car il n'a pas été conçu et calibré pour cela).

Il existe donc une asymétrie complète dans le mode de raisonnement entre efficacité et safety, car les craintes ne sont pas les mêmes. Pour l'efficacité, la crainte est de conclure à tort à une efficacité qui n'existe pas. Ce sont les résultats faux positifs qui sont redoutés. Pour la safety, la crainte est de passer à côté d'un effet indésirable et d'utiliser in fine un traitement dont les risques dépassent le bénéfique. La situation redoutée est alors celle des résultats faux négatifs. L'efficacité est une question de risque alpha, la safety une question de risque beta.

Cette asymétrie se retrouve aussi en termes de démarche scientifique. Pour l'efficacité, l'évaluation doit impérativement s'inscrire dans une démarche hypothético-déductive, en particulier afin d'éviter d'utiliser un traitement sur la base d'une découverte fortuite⁵. En revanche, pour la safety, une démarche inductive est suffisante pour conclure à un surcroît d'évènements indésirables et décider de ne pas utiliser le traitement. Sur le plan éthique, le recours à la démarche hypothético-déductive ne serait d'ailleurs pas possible, car elle reviendrait à mettre en place un essai spécifiquement dans le but de montrer que le traitement entraîne des évènements indésirables ! Un tel objectif d'étude n'est, à l'évidence, pas envisageable.

4.1 Le risque d'inflation des faux positifs en safety

Malgré les aspects légitimes de la gestion de la safety des nouveaux traitements suivant le principe de précaution, la laxité méthodologique qu'il autorise fait émerger une problématique d'inflation des faux positifs. Pour réduire le risque de faux négatif, la moindre observation en faveur d'un risque potentiel est souvent prise en considération et conduit à la non-utilisation du traitement même dans le cas d'observation peut robuste. De ce fait, le moindre faux positif peut être érigé en fait « probant » devant conduire à une décision. Comme la démarche sur laquelle est appliqué le principe de précaution est par essence exploratoire et inductive, des signaux de défaut de safety sont cherchés tous azimuts, entraînant une très forte multiplicité de comparaisons. Le risque de faux positifs devient alors considérable. Ainsi les problématiques de p hacking, de découvertes fortuites par fouille intensive

⁴ Dans le cadre d'une comparaison statistique permettant de conclure à une relation de causalité. Ne concerne pas le raisonnement basé sur les cas anecdotiques et l'évaluation de l'imputabilité de la pharmacovigilance.

⁵ Cf. [Dossier 1 – La démarche hypothético-déductive et les résultats post hoc](#)

des données, de cherry picking, de selective reporting, de HARKing⁶ sont toutes aussi présentes que lors de l'évaluation de l'efficacité, mais ne peuvent pas être considérées pour disqualifier les signaux identifiés (sans être accusé de favoriser les intérêts du traitement évalué par exemple).

Cette problématique des faux positifs est aussi amplifiée par le fait que l'analyse retrospective de la safety (à partir de base de données, de méta-analyses, etc.) est facilement réalisable par les chercheurs académiques et offre ainsi de nombreuses opportunités à publication. Les problématiques liées à l'infodémie et au changement de finalité de la recherche apparaissent ainsi à ce niveau (cf. section [3.4 « Infodémie » et course à la publication](#) du livre blanc).

Le juste équilibre nécessaire entre risque de faux négatifs et risque de faux positifs est difficile à obtenir. Depuis que cette problématique est discutée, quelques exemples de remise en cause de l'existence de certains effets indésirables considérés pourtant comme acquis apparaissent comme avec les statines et les symptômes musculaires [19], les pneumopathies avec les inhibiteurs de la pompe à protons [20], etc.

⁶ Cf. le dossier [De la nécessité de la méthodologie dans l'évaluation des médicaments](#), [section 3. Distorsion de la connaissance scientifique](#)

5 Comment décider en présence d'un surcroît d'effets indésirables

La conception initiale qu'un traitement doit être efficace et sûr pour avoir une balance bénéfice risque favorable s'avère complètement désuet dans de nombreux domaines thérapeutiques concernant des pathologies graves (oncologie, hématologie, cardiologie aigüe, neurologie aigüe, etc.) où les traitements sont susceptibles d'induire de réels effets délétères graves (comme des décès). Il s'agit par exemple des fibrinolytiques à la phase aigüe de l'infarctus du myocarde qui sont pourvoyeur éventuellement d'hémorragies fatales (par hémorragies intra-crâniennes entre autres). Très fréquemment les traitements anticoagulants entraînent une augmentation statistiquement significative de la fréquence des hémorragies majeures. De nombreux traitements anticancer sont aussi susceptibles d'entraîner des décès toxiques.

L'essai TRITON [21] comparant le prasugrel au clopidogrel dans le traitement initial des syndromes coronariens aigus. Les données de safety rapportées dans le tableau ci-dessous montrent que l'utilisation du prasugrel s'accompagne d'une augmentation statistiquement significative de la fréquence des saignements majeurs de +32%, et des saignements mettant en jeu le pronostic vital de +52%. Il est donc clair que le prasugrel induit un surrisque d'hémorragies majeures. Cependant cette molécule réduit la fréquence des événements ischémiques mortels et non mortels. Se pose alors la question de décider de l'intérêt clinique de ce traitement en présence d'effets indésirables graves.

Table 3. Thrombolysis in Myocardial Infarction (TIMI) Bleeding End Points in the Overall Cohort at 15 Months.*

End Point	Prasugrel (N=6741)	Clopidogrel (N=6716)	Hazard Ratio for Prasugrel (95% CI)	P Value
<i>no. of patients (%)</i>				
Non-CABG-related TIMI major bleeding (key safety end point)	146 (2.4)	111 (1.8)	1.32 (1.03–1.68)	0.03
Related to instrumentation	45 (0.7)	38 (0.6)	1.18 (0.77–1.82)	0.45
Spontaneous	92 (1.6)	61 (1.1)	1.51 (1.09–2.08)	0.01
Related to trauma	9 (0.2)	12 (0.2)	0.75 (0.32–1.78)	0.51
Life-threatening†	85 (1.4)	56 (0.9)	1.52 (1.08–2.13)	0.01
Related to instrumentation	28 (0.5)	18 (0.3)	1.55 (0.86–2.81)	0.14
Spontaneous	50 (0.9)	28 (0.5)	1.78 (1.12–2.83)	0.01
Related to trauma	7 (0.1)	10 (0.2)	0.70 (0.27–1.84)	0.47
Fatal‡	21 (0.4)	5 (0.1)	4.19 (1.58–11.11)	0.002
Nonfatal	64 (1.1)	51 (0.9)	1.25 (0.87–1.81)	0.23
Intracranial	19 (0.3)	17 (0.3)	1.12 (0.58–2.15)	0.74
Major or minor TIMI bleeding	303 (5.0)	231 (3.8)	1.31 (1.11–1.56)	0.002
Bleeding requiring transfusion§	244 (4.0)	182 (3.0)	1.34 (1.11–1.63)	<0.001
CABG-related TIMI major bleeding¶	24 (13.4)	6 (3.2)	4.73 (1.90–11.82)	<0.001

Ainsi, le vieil adage « *primum non nocere* », toujours enseigné et fréquemment cité, tient plus d'une certaine scolastique universitaire que de la réalité actuelle de la thérapeutique de nombreuses pathologies. Traiter dans ces situations peut nuire au patient, mais avec une probabilité bien plus faible que celle de lui apporter un bénéfice. Au total, la balance bénéfice risque est très positive. Si la priorité

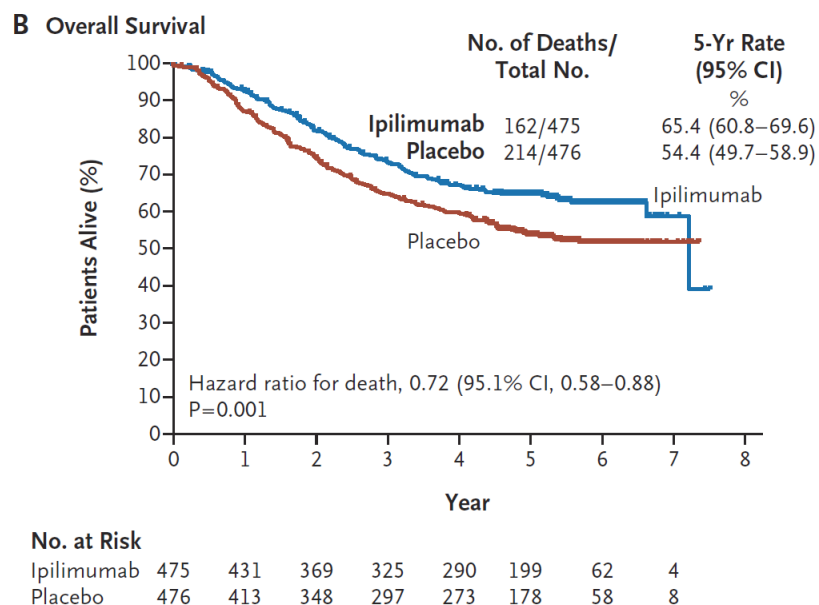
était d'abord d'éviter de pouvoir nuire aux patients, ce type de traitements ne pourrait pas être envisagé compte tenu de leur potentialité de nuisance (d'effets indésirables graves). Pourtant ces traitements apportent un réel bénéfice clinique (cf. section suivante) et illustrent bien le fait qu'actuellement, la **question n'est pas l'absence d'effet indésirable, mais une probabilité de bénéfice supérieure à la probabilité d'un évènement délétère**. La problématique est quantitative et peut être résolue par l'intégration quantitative du bénéfice et du risque à l'aide de différents outils comme le calcul du bénéfice clinique net (cf. section 6).

Au stade précoce du mélanome, après résection chirurgicale satisfaisante, la pratique standard était une expectative attentive, sans traitement adjuvant. L'ipilimumab en adjuvant a été évalué versus placebo dans l'essai randomisé EORTC 18071 [22]. Cinq décès toxiques (grade 5) par effet indésirable immunologique ont été observés dans le groupe ipilimumab, ce qui au premier abord semble remettre en question l'intérêt de ce traitement instauré dans une situation où jusqu'à présent la règle était l'abstention thérapeutique.

Table 3. Immune-Related Adverse Events.*

Event	Ipilimumab (N=471)				Placebo (N=474)			
	Any Grade	Grade 3	Grade 4	Grade 5	Any Grade	Grade 3	Grade 4	Grade 5
Any immune-related adverse event	426 (90.4)	169 (35.9)	27 (5.7)	5 (1.1)	188 (39.7)	12 (2.5)	1 (0.2)	0

Cependant, malgré ces décès toxiques, l'essai démontre une réduction de la mortalité :



Le nombre de décès évités par le traitement est nettement supérieur à celui des décès toxiques, faisant que la probabilité à priori que le patient bénéficie du traitement plutôt qu'il n'en pâtisse est plutôt favorable, comme en témoigne la réduction relative du risque instantané de décéder de 28% estimée par le hazard ratio. Cet exemple montre la piste du raisonnement du bénéfice clinique net : intégrer quantitativement le risque avec le bénéfice pour montrer que le bilan net est positif. C'est ce que fait ici la mortalité totale en intégrant les décès toxiques avec les décès évités par l'efficacité du traitement. Au total le nombre net de décès est réduit. D'autres traitements (anti PD-1 et l'association

6 L'approche quantitative de la balance bénéfice risque

Comme nous l'avons vu dans la section précédente, de plus en plus de traitements induisent des effets délétères (par exemple les hémorragies, parfois mortelles, avec les anticoagulants et antiagrégants plaquettaires). Paradoxalement, plus un traitement est « puissant », plus il est susceptible d'apporter un bénéfice, mais aussi des effets délétères (on peut aussi prendre comme exemple le domaine des anticoagulants avec les stratégies d'intensification par augmentation de dose ou association d'antiagrégants ou antiagrégant et anticoagulant). Le terme « balance bénéfice risque » prend alors toute sa dimension dans l'appréciation in fine de l'intérêt clinique du traitement et la prise de décision. Ces situations (de plus en plus fréquentes) montrent aussi que l'on ne peut plus se contenter d'une simple recherche de traitement (complètement) sûr.

Ce qui pourrait justifier d'utiliser un traitement présentant des risques avérés est que, lorsque le patient débute le traitement, la probabilité qu'il en tire un bénéfice clinique soit supérieure à la probabilité qu'il pâtit d'un effet délétère (et à condition aussi qu'il n'existe pas de meilleur traitement, plus sûr).

La question se pose en probabilité, car, comme pour tout le domaine de l'évaluation des thérapeutiques, il existe une variabilité inter-sujets irréductible qui fait que les événements délétères ne surviennent que chez quelques patients sans qu'il soit possible de les identifier à priori, et que les événements que cherche à éviter le traitement surviennent également de manière tout aussi erratique et imprévisible. Ainsi le modèle probabiliste permet de gérer ce caractère imprévisible et aléatoire pour formaliser le processus décisionnel dans un univers aussi incertain à priori.

Au moment de la décision d'instaurer un traitement pour un patient donné, il est impossible de savoir de quoi l'avenir sera exactement fait : est-ce que le patient va faire un effet indésirable ? Est-ce qu'il ferait l'évènement que l'on cherche à prévenir sans traitement ? Ne va-t-il pas malgré le traitement faire cet évènement ? La gestion de cette incertitude⁷ repose alors sur les probabilités. Ainsi, avec ces traitements efficaces, mais aussi pourvoyeur d'effets indésirables graves, l'instauration d'un traitement présente un intérêt tant que la probabilité à priori (au moment de la prescription) du bénéfice est supérieure à la probabilité de l'effet indésirable.

En d'autres termes le traitement gardera un intérêt tant que les événements indésirables qu'il induit ne viennent pas contrebalancer en totalité le bénéfice qu'il apporte. Les événements indésirables qu'il induit ne doivent donc pas compenser numériquement les événements qu'il permet d'éviter.

Cette évaluation peut se faire de différentes manières, soit à partir des tailles d'effet sur les critères d'efficacité et de safety produite par l'essai [2, 24], soit directement par un critère de jugement composite.

6.1 Appréciation du bénéfice clinique net par les différences des risques

La balance bénéfice risque peut-être appréciée quantitativement en comparant les différences de risques, aussi appelée bénéfice absolu (*Absolute Risk Reduction*, ARR) correspondant à l'effet du traitement sur le critère d'efficacité et sur le critère de safety.

⁷ liée à l'absence d'outils prédictifs fiables pour la situation clinique considérée ou à une nature réellement aléatoire des événements cliniques

Exemple d'une balance bénéfice risque non favorable quantitativement

L'essai PEGASUS [25] a comparé l'association ticagrelor + aspirine à l'aspirine seule pour la prévention des évènements cardiovasculaires ischémiques à distance de l'épisode aigu illustre parfaitement bien cette problématique. Ce traitement à la dose de 60mg multiplie par 2.3 (hazard ratio de 2.32) le risque de saignements majeurs (critère de safety primaire) soit une augmentation relative de +132%.

Table 3. Safety End Points as 3-Year Kaplan–Meier Estimates.*

End Point	Ticagrelor, 90 mg (N = 6988)	Ticagrelor, 60 mg (N = 6958)	Placebo (N = 6996)	Ticagrelor, 90 mg vs. Placebo		Ticagrelor, 60 mg vs. Placebo	
				Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
<i>number (percent)</i>							
Bleeding							
TIMI major bleeding	127 (2.60)	115 (2.30)	54 (1.06)	2.69 (1.96–3.70)	<0.001	2.32 (1.68–3.21)	<0.001

Cependant il permet de prévenir davantage la survenue d'un évènement ischémique grave (infarctus ou AVC mortel ou non, critère d'efficacité primaire de l'essai) en réduisant leur fréquence de 16% en relatif (hazard ratio de 0.84).

Table 2. Efficacy End Points as 3-Year Kaplan–Meier Estimates.

End Point	Ticagrelor, 90 mg (N = 7050)	Ticagrelor, 60 mg (N = 7045)	Placebo (N = 7067)	Ticagrelor, 90 mg vs. Placebo		Ticagrelor, 60 mg vs. Placebo	
				Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
<i>number (percent)</i>							
Cardiovascular death, myocardial infarction, or stroke	493 (7.85)	487 (7.77)	578 (9.04)	0.85 (0.75–0.96)	0.008	0.84 (0.74–0.95)	0.004

Au premier abord, il semble que l'augmentation des évènements hémorragique (+132%) outrepassent la réduction des évènements cardiovasculaire (-16%). Cependant il convient de remarquer qu'il s'agit d'effets relatifs qui ne sont pas directement comparables, car affectant des évènements qui n'ont pas le même risque de base. En effet, dans le groupe placebo la fréquence des saignements majeurs est 1.06% tandis que celle des évènements cardiovasculaires est bien plus élevée, 9.04%. Pour cette raison, il convient de comparer les augmentations ou réductions absolues et non pas relatives.

La différence de risque est obtenue en faisant simplement la différence de la fréquence des évènements considérés entre le groupe traité et le groupe contrôle. Cette différence absolue permet de quantifier le nombre d'évènements évités ou induits par le traitement de 100 patients⁸ (lorsqu'elle est exprimée en pourcentage ou pour 1000 patients traités lorsque cette différence est inférieure à 1%).

Appliqué à l'essai PEGASUS [25] (cf. ci-dessus), la différence de risque pour la dose 60mg sur le critère d'efficacité est de 7.77%-9.04% = -1.27%, soit 12.7 évènements cardiovasculaires évités pour 1000 patients traités durant la durée de l'étude. Pour le critère de jugement principal de safety, les saignements majeurs, la différence des risques est 2.30%-1.06%= +1.24%, soit 12.4 hémorragies de plus pour 1000 patients traités, avec la même dose.

⁸ Sur une durée de traitement identique à celle de l'étude ou au suivi effectué dans l'étude

La balance bénéfique apparaît ainsi non favorable à l'utilisation du traitement étant donné que le nombre d'hémorragies induites compense numériquement la totalité des événements cardiovasculaires évités. Sur 1000 patients traités, il y aura moins d'événements ischémiques qu'avec le traitement contrôle mais il y aura le même nombre d'hémorragies majeures en plus. La balance bénéfique risque n'est donc pas favorable (cf. avis de la commission de transparence https://www.has-sante.fr/jcms/c_2744502/).

Exemple de balance bénéfique risque quantitative favorable

L'essai PLATO [26] a évalué l'administration durant 1 an après l'événement aigu du ticagrelor par rapport au clopidogrel pour la prévention de la récurrence des événements cardiovasculaires ischémiques. Une réduction des événements cardiovasculaires a été observée.

End Point	Ticagrelor Group	Clopidogrel Group	Hazard Ratio for Ticagrelor Group (95% CI)	P Value†
Primary end point: death from vascular causes, MI, or stroke — no./total no. (%)	864/9333 (9.8)	1014/9291 (11.7)	0.84 (0.77–0.92)	<0.001‡

Les résultats, pour le critère principal de safety, ne montrent pas d'augmentation significative des saignements majeurs.

End Point	Ticagrelor Group	Clopidogrel Group	Hazard or Odds Ratio for Ticagrelor Group (95% CI)†	P Value
Primary safety end points — no./total no. (%)				
Major bleeding, study criteria	961/9235 (11.6)	929/9186 (11.2)	1.04 (0.95–1.13)	0.43

La différence des risques pour l'efficacité est $9.8\% - 11.7\% = -1.9\%$. Celle pour la safety est $11.6\% - 11.2\% = +0.4\%$. La balance bénéfique risque est donc largement favorable. Le bénéfice obtenu au niveau des événements ischémiques n'est pas contrebalancé par l'effet du traitement sur les saignements majeurs. La différence de ces 2 différences permet, éventuellement, d'estimer un bénéfice net de $-1.9\% - 0.4\% = -1.5\%$, correspondant à 15 événements ischémiques évités, net de saignements majeurs, pour 1000 patients traités durant la durée de l'étude. La balance bénéfique risque est ainsi favorable et permet de positionner le traitement dans la stratégie thérapeutique concernant ces patients (cf. avis de la commission de transparence https://www.has-sante.fr/jcms/c_1172422/).

L'approche de quantification de la balance bénéfique risque par l'approche des différences de risque est aussi utilisée en méta-analyse.

L'effet global des inhibiteurs SGLT2 sur la fonction rénale pour différents types de patients a été estimé par une méta-analyse sur données individuelles [27]. Cette méta-analyse regroupant 13 essais incluant 90 413 patients montre plusieurs bénéfices sur le rein de ces molécules ainsi que de possibles effets indésirables en termes d'acidocétose et d'amputation (principalement chez le diabétique). Dans ce contexte, une analyse fine de la balance bénéfique risque a été entreprise en confrontant les différences de risques des principaux bénéfices avec celles des effets indésirables augmentées de manière statistiquement significative qui montre « Based on estimates of absolute effects, the absolute benefits of SGLT2 inhibition outweighed any serious hazards of ketoacidosis or amputation ». Les résultats sont

présentés sous forme graphique permettant une appréciation visuelle directe du bénéfice clinique net (pour les différents types de patients considérés).

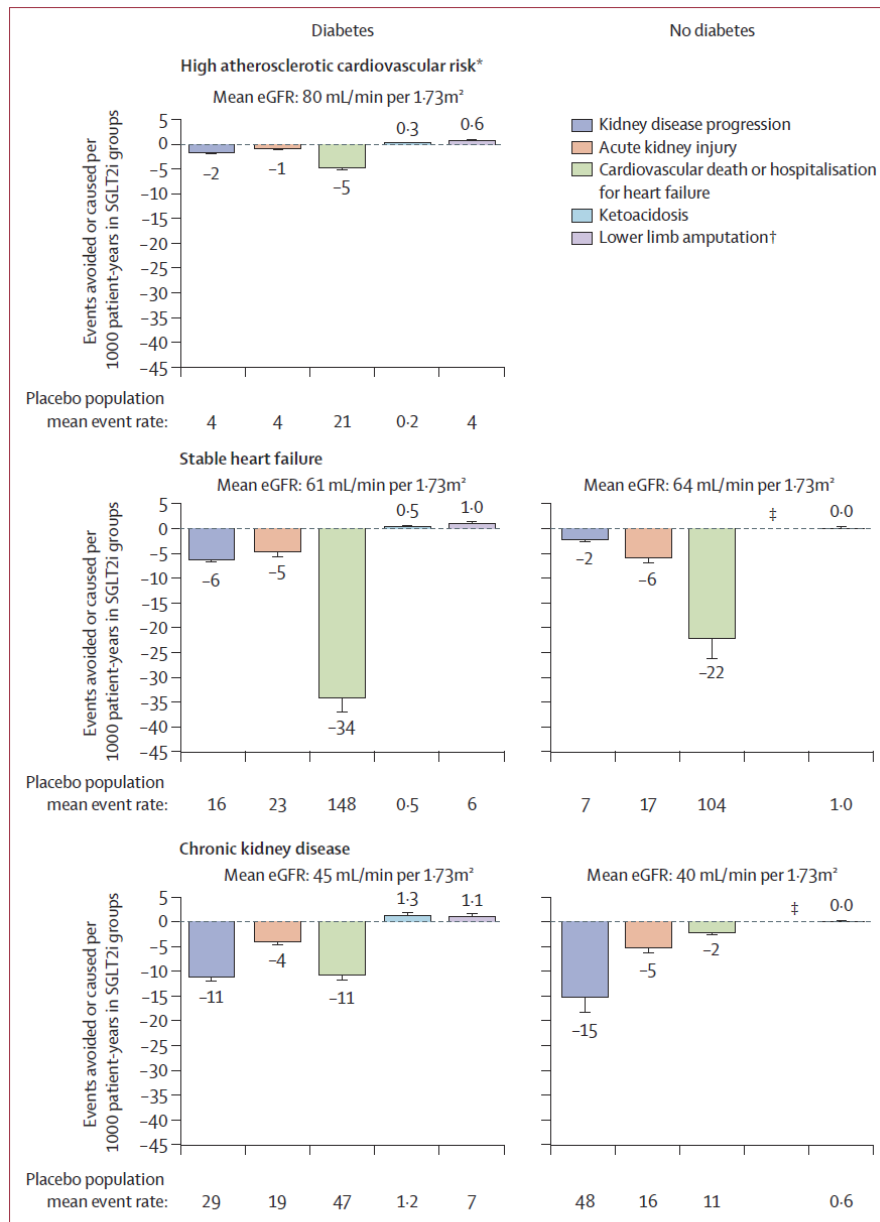


Figure 5: Absolute benefits and harms of SGLT2 inhibition per 1000 patient-years by diabetes status and patient group
 Patient group specific absolute effects estimated by applying the diabetes subgroup specific RR to the average event rate in the placebo arms (first event only). Negative numbers indicate events avoided by SGLT2 inhibition per 1000 patient-years. Error bars represent SE in the numbers of events avoided or caused, estimated from uncertainty in the RRs. Mean eGFR values are given for combined trial populations by patient group and diabetes status. Placebo population mean event rates are the absolute numbers of events per 1000 patient-years in the placebo groups of all trials in the relevant subpopulation. eGFR=estimated glomerular filtration rate. RR=relative risk. SGLT2i=sodium glucose co-transporter-2 inhibitor. *Additionally, two (SE 0.5) fewer myocardial infarctions per 1000 patient-years of SGLT2i treatment were observed in the diabetes and high atherosclerotic cardiovascular risk group. †RRs to determine absolute effects for lower limb amputation included CANVAS. ‡Too few ketoacidosis events to estimate absolute effects.

Ces calculs soulèvent plusieurs points de discussion. Le premier concerne l'équivalence clinique des évènements d'efficacité et des évènements de safety pris en compte. Ils doivent représenter des

événements de la même sévérité clinique étant donné que l'occurrence d'un événement de safety contrecarre exactement l'évitement d'un événement d'efficacité dans ces calculs. Ainsi la gravité, le pronostic, le niveau d'embêtement pour le patient (hospitalisation, traitement) doit être similaire entre ces deux éléments du calcul. C'est d'ailleurs la définition du critère principal de safety. En effet, dans ce terme, critère principal de safety, le mot « principal » ne renvoie pas à l'aspect statistique du contrôle du risque alpha global comme avec le critère de jugement principal (d'efficacité), car le risque alpha n'est pas pertinent pour la safety (cf. section 2.1). Le mot « principal » signifie que ce critère de safety représente des événements qui sont à même de remettre en question le bénéfice obtenu sur le critère de jugement principal (d'efficacité). Il s'agit donc des événements choisis par les investigateurs de l'essai pour faire cette confrontation dans l'évaluation de la balance bénéfice risque.

Il a été proposé de pondérer les événements pour obtenir cette équivalence clinique. Cette pondération se base que sur des considérations cliniques et doit être faite à priori (comme est fait à priori le choix du critère de jugement principal de safety). En pratique cette pondération, n'est pas effectuée de manière numérique, mais en sélectionnant pour le critère principal de safety des événements de même gravité.

Ce point de l'équivalence clinique conduit souvent à des discussions infinies lorsque la balance bénéfice n'apparaît pas favorable [28], mais repose le plus souvent sur des arguments arbitraires, choisis post-hoc pour l'opportunité qu'ils donnent de récuser cette appréciation défavorable. Il existe ainsi un besoin de standardisation des critères de safety dans tous les domaines pour éviter ces discussions, comme cela a été le cas dans le domaine des antithrombotiques avec la définition de saignements majeurs dont la survenue est cliniquement problématique (avec au moins 3 définitions différentes ISTH, TIMI ou GUSTO).

Il a aussi été proposé d'utiliser une pondération numérique pour tenir compte des différences de gravité clinique entre les événements d'efficacité et de sécurité. Cette pondération est alors fixée de manière arbitraire, post hoc et autorise ainsi à produire le résultat escompté [29].

Le deuxième point de discussion que soulève ce calcul est celui de la cooccurrence des 2 événements chez certains patients. Les données rapportées dans les publications ne permettent pas de prendre en compte ces co-occurrences. Les patients concernés sont alors comptés deux fois, au niveau du critère d'efficacité et aussi au niveau du critère de safety. La prise en compte de ces co-occurrences nécessite le recours aux données individuelles, avec la constitution d'un critère composite de bénéfice clinique net (cf. section 6.4).

Le dernier point de discussion concerne l'absence de prise en considération des incertitudes statistiques existant sur la connaissance du réel bénéfice et du réel effet indésirable. Le calcul de l'intervalle de confiance de la différence des différences de risques nécessite de faire une hypothèse d'indépendance en probabilité des deux événements. L'utilisation d'un critère composite de bénéfice clinique net permet quant à lui de prendre cette double incertitude sans faire cette hypothèse et permet de savoir s'il est possible de conclure à une balance bénéfice risque quantitative favorable de manière statistiquement significative (cf. section 6.4). D'autres approches sont aussi disponibles comme les analyses par paire de *net benefit* ou de *win ratio* [30, 31, 32], mais elles nécessitent les données individuelles. De plus, ces approches très récentes ne sont pas encore utilisées couramment dans les essais.

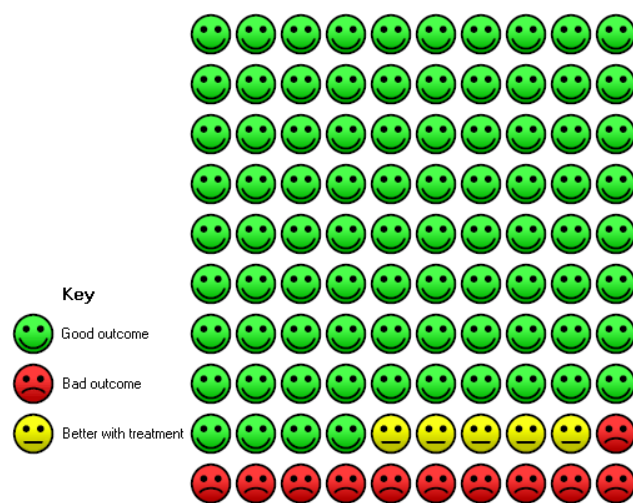
En pratique, l'absence de la prise en compte de l'incertitude statistique n'est gênante que si l'on cherche des certitudes (statistiques) sur le fait que la balance bénéfice risque soit favorable. Dans ce cas, il faudrait pouvoir s'assurer que cette balance reste favorable même dans la situation du pire (borne supérieure de l'IC du bénéfice net calculé). Cependant une telle exigence (démonstration formelle que la balance bénéfice risque est favorable) n'est pas vraiment exigée à l'heure actuelle à

aucun niveau de l'évaluation des nouveaux traitements, montrant que, finalement, la question de la safety est considérée de manière plutôt secondaire par rapport à l'efficacité. La méta-recherche a documenté à de nombreuses reprises que dans les essais de phases 3, l'évaluation de la safety passer au second plan par rapport à celle de l'efficacité, comme aussi dans la prise en compte de ces données pour la décision [33, 34, 35, 36, 37, 38].

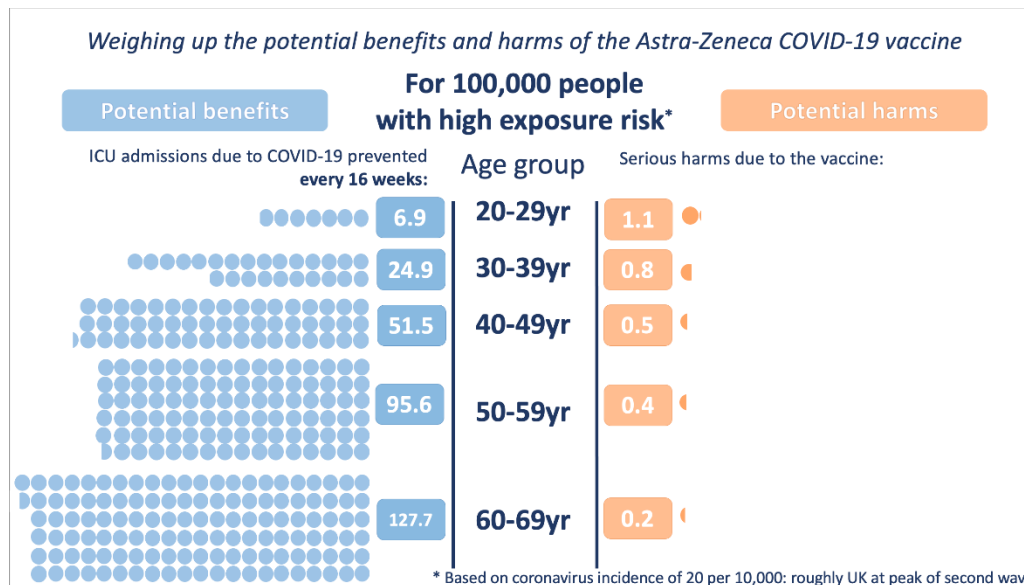
En revanche, lorsque le calcul suggère une balance bénéfice risque non favorable ou défavorable, la question de l'incertitude statistique ne se pose presque plus en application du principe de précaution.

6.2 Graphique de Cates

L'appréciation de la balance bénéfice risque par les différences de risque peut faire l'objet d'une représentation graphique connue entre autres sous le nom de « *Cates graph* » [39].



Cette représentation est devenue populaire et elle est utilisée aussi dans la presse grande publique, par exemple lors des discussions autour de la balance bénéfice risque des différents vaccins COVID-19.



Ref <https://wintoncentre.maths.cam.ac.uk/news/latest-data-mhra-blood-clots-associated-astra-zeneca-covid-19-vaccine/>

Elle est utilisée dans l'outil MagicAPP (<https://app.magicapp.org/>) comme outil de médiation de ces informations à destination du patient dans le cadre de la décision médicale partagée

6.3 Appréciation du bénéfice clinique net par les NNT et NNH

Il a été proposé d'effectuer l'appréciation qualitative de la balance bénéfice risque avec le ratio NNT/NNH [40]. Le NNT (*Number Needed to Treat*) est le nombre de sujets nécessaire de traiter pour éviter un évènement⁹. Il se calcule par l'inverse de la différence des risques. Le NNH (*Number Needed to Harm*) est le nombre de sujets traités qui emmènera à un effet indésirable de plus que dans le groupe contrôle. Il se calcule de la même façon que le NNT. Le nom change uniquement pour signifier que ce paramètre concerne un effet indésirable.

Dans un essai d'un antithrombotique en prévention primaire cardiovasculaire, la fréquence des évènements cardiovasculaires a été de 6% dans le groupe traité contre 8% dans le groupe contrôle. Le NNT est de 50, obtenu par $1/(0.06-0.08)$. Le signe négatif de la différence de risque, qui témoigne d'une réduction de la fréquence de l'évènement considéré par le traitement, est oublié étant donné que le terme NNT signifie un effet bénéfique (« pour éviter »). Si un effet indésirable était observé sur les évènements cardiovasculaires (une augmentation de leur fréquence par le traitement censé les prévenir) il faudrait utiliser le terme de NNH.

Dans cet essai, une augmentation des hémorragies majeures est observée dont la fréquence passe de 0.4% dans le groupe contrôle à 1.2% sous traitement. Le NNH (il y a augmentation, donc effet indésirable) est de 125, obtenu par $1/(0.012-0.004)$.

Le rapport NNT/NNH est de $50/125 = 0.4$ signifiant que 0.4 hémorragies sont induites par évènement ischémique évité (ou 4 hémorragies induites pour 10 évènements ischémiques évités).

⁹ Voir le dossier à venir sur les indices d'efficacité

L'inverse de ce rapport, le NNH/NNT est peut-être plus intelligible. Dans cet exemple il est de $125/50 = 2.5$, signifiant que 2.5 évènements ischémiques sont évités par hémorragie majeure induite.

Bien qu'étant une des premières méthodes proposées pour l'appréciation quantitative de la balance bénéfice risque [40], cette méthode semble de moins en moins utilisée au profit de l'approche par les différences de risque.

6.4 Critère composite de bénéfice clinique net

L'appréciation quantitative de la balance bénéfice risque peut aussi s'effectuer à l'aide d'un critère composite de bénéfice clinique net (*net clinical benefit*). Ce critère composite regroupe un ou plusieurs critères d'efficacité avec un ou plusieurs critères de safety. Ainsi la réduction éventuelle du nombre d'évènements d'efficacité qu'entraînera le traitement (par rapport au groupe contrôle) sera compensée, dans le décompte global, par le nombre d'évènements indésirables qu'il entraînera. Si ce critère de bénéfice clinique net est réduit dans le groupe traité par rapport au groupe contrôle, cela témoigne d'une balance bénéfice risque favorable. La taille d'effet mesure alors, directement, le bénéfice clinique net, c'est-à-dire le bénéfice, net d'effet indésirable (comme en comptabilité où le chiffre d'affaires déduit des coûts, donne un bénéfice net). Bien sûr, il convient que les évènements regroupés pour faire ce critère composite soient de gravité clinique similaires (cf. discussion de ce point en section 6.1).

Pour illustrer ce concept, nous pouvons prendre le cas d'un antithrombotique en prévention des évènements cardiovasculaires artériels. Dans un premier cas de figure, le traitement réduit la fréquence des évènements cardiovasculaire de 10% dans le groupe contrôle à 5% et n'induit aucune hémorragie de même gravité (3% dans les 2 groupes).

	Antithrombotique évalué N=1000	Contrôle N=1000	Risque ratio	Différence des risques
Évènements cardiovasculaires ischémiques (efficacité)	50 (5%)	100 (10%)	0.50	-5%
Hémorragies (safety)	30 (3%)	30 (3%)	1.00	0%
Critère composite de bénéfice clinique net ¹⁰	80 (8%)	130 (13%)	0.62	-5%

Si les hémorragies sont augmentées et passent de 3% dans le groupe contrôle à 6% avec le traitement évalué, le critère composite reste toujours en faveur du traitement évalué (au moins en tendance, sans tenir compte de la significativité statistique) :

¹⁰ Ce calcul fait l'hypothèse qu'aucune co-occurrences des 2 évènements n'est survenue. En pratique, ce critère est mesuré dans l'essai et décompte les patients ayant fait un évènement ischémique, ou une hémorragie, ou les deux, en ne comptant un même patient qu'une seule fois (patients ayant fait au moins un évènement).

	Antithrombotique évalué N=1000	Contrôle N=1000	Risque ratio	Différence des risques
Événements cardiovasculaires ischémiques (efficacité)	50 (5%)	100 (10%)	0.50	-5%
Hémorragies (safety)	60 (6%)	30 (3%)	2.00	+3%
Critère composite de bénéfice clinique net	110 (11%)	130 (13%)	0.85	-2%

Par rapport au calcul simple des différences des différences des risques (cf. section précédente), l'approche du critère composite permet de gérer les co-occurrences des 2 critères chez les mêmes patients. En effet, dans les critères composites, les patients ne sont comptés qu'une seule fois (souvent en fonction du premier évènement qui survient).

L'essai ARISTOTLE [41] a comparé l'apixaban à la warfarine dans la fibrillation auriculaire. Deux critères composites de bénéfice clinique net, regroupant des entités différentes, sont présentés. Ces critères sont purement exploratoires dans cet essai et ne sont pas intégrés à la hiérarchie des critères décisionnels pour la démonstration du ou des bénéfices cliniques de l'apixaban. Une réduction statistiquement significative de ces deux critères est observée suggérant une réduction des AVC et embolie systémiques nette d'hémorragies majeures (pour le premier des deux).

Table 3. Bleeding Outcomes and Net Clinical Outcomes.*

Outcome	Apixaban Group (N=9088)		Warfarin Group (N=9052)		Hazard Ratio (95% CI)	P Value
	Patients with Event	Event Rate	Patients with Event	Event Rate		
	no.	%/yr	no.	%/yr		
Primary safety outcome: ISTH major bleeding†	327	2.13	462	3.09	0.69 (0.60–0.80)	<0.001
Intracranial	52	0.33	122	0.80	0.42 (0.30–0.58)	<0.001
Other location	275	1.79	340	2.27	0.79 (0.68–0.93)	0.004
Gastrointestinal	105	0.76	119	0.86	0.89 (0.70–1.15)	0.37
Major or clinically relevant nonmajor bleeding	613	4.07	877	6.01	0.68 (0.61–0.75)	<0.001
GUSTO severe bleeding	80	0.52	172	1.13	0.46 (0.35–0.60)	<0.001
GUSTO moderate or severe bleeding	199	1.29	328	2.18	0.60 (0.50–0.71)	<0.001
TIMI major bleeding	148	0.96	256	1.69	0.57 (0.46–0.70)	<0.001
TIMI major or minor bleeding	239	1.55	370	2.46	0.63 (0.54–0.75)	<0.001
Any bleeding	2356	18.1	3060	25.8	0.71 (0.68–0.75)	<0.001
Net clinical outcomes						
Stroke, systemic embolism, or major bleeding	521	3.17	666	4.11	0.77 (0.69–0.86)	<0.001
Stroke, systemic embolism, major bleeding, or death from any cause	1009	6.13	1168	7.20	0.85 (0.78–0.92)	<0.001

6.5 Problématique de l'utilisation du bénéfice net en non-infériorité

Les critères composites de bénéfice net sont parfois utilisés en non-infériorité. Cette utilisation est complètement inappropriée, car elle peut conduire à accepter un traitement qui n'est ni supérieur en

termes de sécurité ni en termes d'efficacité et qui ne représente de ce fait aucun progrès thérapeutique par rapport au traitement standard.

L'essai MASTER DAPT [42] a évalué une stratégie de désescalade thérapeutique de la double anti-agrégation plaquettaire après stent coronarien chez des sujets à haut risque de saignement. Une interruption de la double anti-agrégation plaquettaire après 1 mois de traitement était comparée à sa poursuite durant 2 mois supplémentaires (durée standard). Le but était de limiter les saignements majeurs sans trop perdre en termes de protection contre la sténose de stent. L'essai a utilisé une logique de non-infériorité sur le critère composite de bénéfice net regroupant saignement majeur et évènements cardiovasculaires mortels et non-mortels.

L'essai TROPICAL ACS [43] a évalué une stratégie de désescalade thérapeutique de la bithérapie antiagrégants plaquettaires après syndrome coronarien aigu. Le critère de jugement, évaluée en non-infériorité, était le critère composite hémorragies majeures et évènements coronariens.

La logique de la non-infériorité est rappelons-le (cf. [Dossier 6 – L'essai de non-infériorité](#)) d'accepter d'éventuellement perdre en efficacité avec le nouveau traitement, car il représentera malgré cela un progrès thérapeutique du fait d'un avantage important sur un autre plan que l'efficacité (tolérance, facilité d'utilisation, cout, etc.). Rapporté à la question de la balance bénéfice risque, cette logique correspond à accepter de perdre un peu en efficacité, car, au total, il a aura un progrès en raison d'une réduction des évènements indésirables. Mais si le raisonnement de non-infériorité est appliqué sur un critère de bénéfice net, il y a acceptation que le nouveau traitement puisse avoir une moins bonne balance bénéfice risque que le standard, sans qu'il y ait un avantage par ailleurs pouvant rendre acceptable cette perte potentielle étant donné que cet avantage, moins d'effets indésirables, a été intégré dans le raisonnement de non-infériorité ? Une non-infériorité sur un critère de bénéfice net signifie que l'on accepte d'être inférieur en termes de balance bénéfice risque et sans que le nouveau traitement présente un avantage par ailleurs (puisque celui-ci aurait été justement une meilleure sécurité, déjà pris en compte dans le critère composite).

Ainsi, si un nouveau traitement est potentiellement moins efficace (comme dans les exemples cités ci-dessus) mais moins pourvoyeur d'effets indésirables, le progrès que pourrait constituer ce traitement doit être mis en évidence soit par une démonstration de non-infériorité sur l'efficacité couplée à une démonstration de la supériorité en termes de tolérance soit par une démonstration de supériorité avec un critère de bénéfice net. Un critère composite de bénéfice net n'a de sens qu'en supériorité pour montre que le nouveau traitement à une meilleure balance bénéfice risque que le standard.

6.5.1 La mortalité totale comme critère de bénéfice clinique net

La mortalité totale est naturellement un critère de bénéfice clinique net dans les situations où le bénéfice recherché est la prévention d'évènements mortels (comme en oncologie) et où les effets indésirables susceptibles de remettre en cause ce bénéfice sont donc aussi de nature fatale. Une réduction de la mortalité totale assure d'un bénéfice net. Par extension, si un bénéfice est démontré sur la mortalité de toute cause, plus aucun effet indésirable n'est susceptible de remettre en cause la balance bénéfice risque du produit en général. La valeur médicale du bénéfice rend acceptables les évènements indésirables comme cela est en général le cas en oncologie (cf. chimiothérapie). Il reste cependant une discussion résiduelle quand la tolérance globale du produit remet en cause la qualité de vie et interroge sur la qualité de l'augmentation de survie apportée par le traitement.

7 Variation de la balance bénéfico risque dans les strates de patients

La balance bénéfico risque d'un même médicament peut être variable en fonction du groupe (de la strate) de patients. En effet, les nombres d'évènements évités ou induits pour un même nombre de patients traités vont varier en fonction du risque de base [45, 46, 47, 47, 48]. Pour une même réduction relative, le nombre d'évènements évités sera d'autant plus élevé que le risque de base est important [49, 50]. Si le risque de base d'évènement indésirable ne change pas avec le risque de base de l'évènement que le traitement cherche à éviter, la balance bénéfico risque sera plus favorable chez les patients à haut risque que chez les patients à bas risque. Cette propriété arithmétique fait que la balance bénéfico risque d'un traitement qui globalement n'est pas favorable peut le devenir dans une sous population à haut risque comme l'illustre l'exemple numérique présenté dans le tableau ci-dessous.

	Patients tous venants	Sous population à haut risque
Efficacité		
Risque de base	5%	20%
Risque relatif	0.80	0.80
Risque sous traitement	4%	16%
Différence des risques	-1%	-4%
Safety		
Risque de base	2%	2%
Risque relatif	2.00	2.00
Risque sous traitement	4%	4%
Différence des risques	+2%	+2%
Bénéfice net	+1%	-2%

Historiquement, cette approche a été illustrée par la question du traitement anticoagulant dans la fibrillation auriculaire en montrant que la décision de recourir à un anticoagulant devait se baser sur le nombre de facteurs de risque d'AVC ischémiques (ou d'un autre outil prédictif du risque d'AVC ischémiques), dans le but de n'exposer les patients au risque hémorragique que si leur propre risque d'AVC est suffisamment important [49, 51]. L'existence d'un seuil de décision [50] apparaît sur le graphique ci-dessous au niveau de risque où le nombre d'AVC évités devient supérieur au nombre d'AVC hémorragiques induits par le traitement anticoagulant.

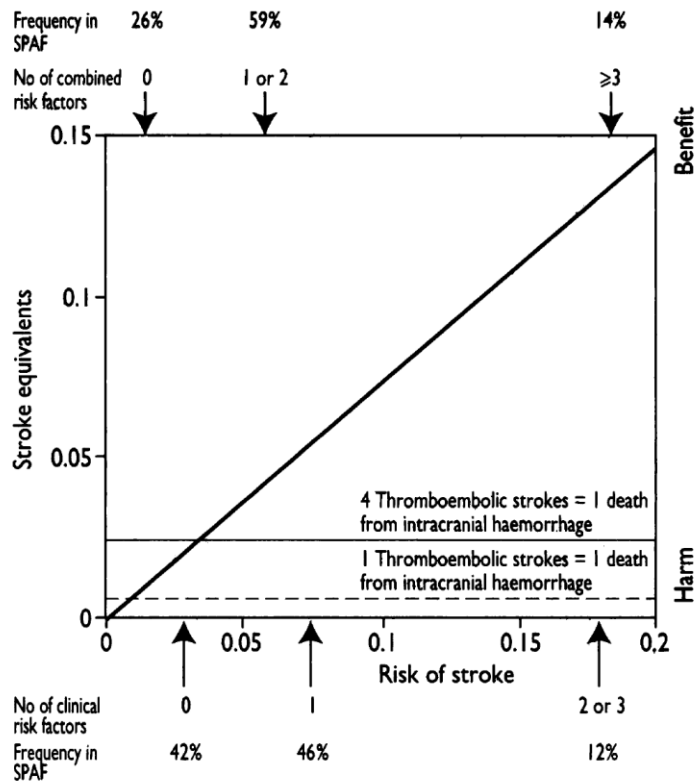


FIG 3—Benefit compared with harm for warfarin on basis of six trials. If one death from an intracranial haemorrhage is equivalent to four thromboembolic strokes, threshold is about 2% per year. This may be predicted from three clinical risk factors (previous embolism, recent cardiac failure, and hypertension; bottom axis) or three clinical plus two echocardiographic risk factors (clinical risk factors plus atrial size and left ventricular dysfunction; bottom axis). SPAF=stroke prevention in atrial fibrillation trial

L'essai ASCEND a évalué l'aspirine versus placebo en prévention primaire des évènements cardiovasculaires chez les personnes diabétiques [2]. Les résultats obtenus montrent une balance bénéfique risque défavorable quantitativement. Il a ensuite été procédé à une analyse en fonction du risque de base (figure ci-dessous) qui ne montre pas d'amélioration de cette balance chez les patients à haut risque.

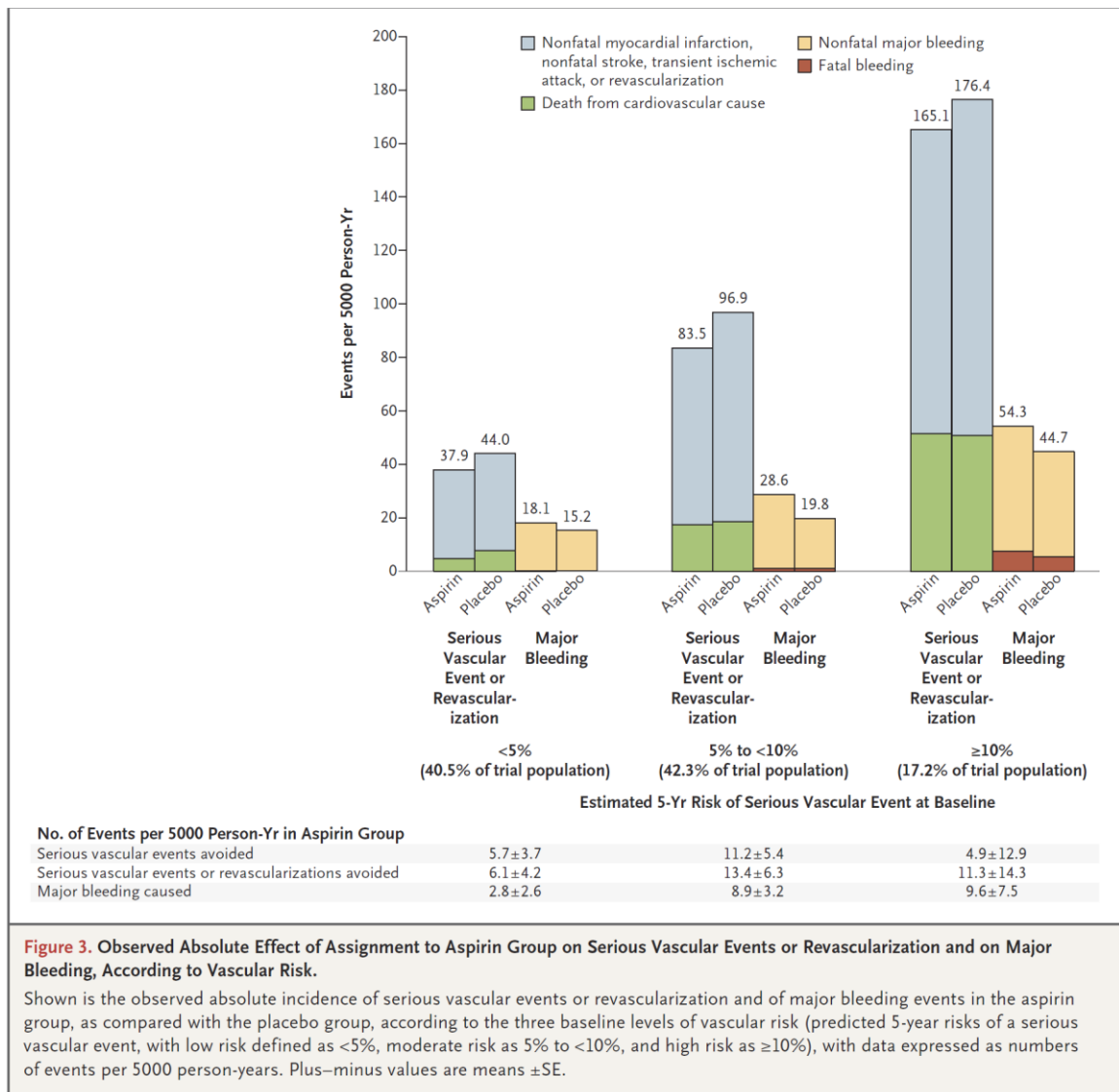


Figure 3. Observed Absolute Effect of Assignment to Aspirin Group on Serious Vascular Events or Revascularization and on Major Bleeding, According to Vascular Risk.

Shown is the observed absolute incidence of serious vascular events or revascularization and of major bleeding events in the aspirin group, as compared with the placebo group, according to the three baseline levels of vascular risk (predicted 5-year risks of a serious vascular event, with low risk defined as <5%, moderate risk as 5% to <10%, and high risk as ≥10%), with data expressed as numbers of events per 5000 person-years. Plus-minus values are means ±SE.

Cette approche est uniquement basée sur des propriétés arithmétiques de l'évolution des différences de risque en fonction du risque de base et ne dépend d'aucune autre hypothèse. Elle ne pose pas de problème de validité si ce n'est celle de la validité de l'outil utilisé pour prédire le risque de base [52]. Il faut à ce niveau remarquer que les outils prédictifs, même s'ils sont très fréquents dans la littérature, ne donnent que rarement toutes les garanties de validité nécessaire pour guider cette démarche. En particulier, la question de leur calibration, qui n'est pas centrale lorsqu'ils sont utilisés pour faire de la stratification, devient cruciale dans cet exercice de prédiction de la balance bénéfice risque. Le manque de validité de ces outils sur ce point conduit à des prises de décision individuelles non fondées, potentiellement dommageables pour les patients.

L'utilisation de cette propriété arithmétique est à distinguer des simples analyses en sous-groupe réalisées au sein d'un essai dont les résultats globaux ne sont pas en faveur d'une balance bénéfice risque favorable. Ces analyses cherchent alors si des variables ne modifieraient pas l'effet des traitements et, de ce fait, la balance bénéfice risque. Cette approche est purement exploratoire et ne peut être que générative d'hypothèse. Elle ne permet de baser les pratiques en raison des limites

habituelles des analyses en sous-groupes (exploratoires, absence d'hypothèse, post hoc, inflation du risque alpha et beta, etc.).

Dans l'essai TRITON [21] du prasugrel pour le traitement initial des syndromes coronariens aigus, la balance bénéfice risque n'apparaît pas favorable en raison d'un surcroît important d'hémorragies graves (y compris mortelles). La publication propose une analyse en sous-groupe à la recherche d'une sous population où la balance bénéfice risque (mesurée par un critère composite de bénéfice clinique net) deviendrait favorable.

Table 4. The Balance of Efficacy and Safety in Selected Subgroups.*

End Point	Prasugrel <i>no. of patients/total no. (%)</i>	Clopidogrel	Hazard Ratio for Prasugrel (95% CI)	P Value	P Value for Interaction†
History of stroke or TIA					
Death from cardiovascular causes, nonfatal MI, or nonfatal stroke (primary efficacy end point)	47/262 (19.1)	35/256 (14.4)	1.37 (0.89–2.13)	0.15	
Non-CABG-related TIMI major bleeding	14/257 (5.0)	6/252 (2.9)	2.46 (0.94–6.42)	0.06	
Death from any cause, nonfatal MI, nonfatal stroke, or non-CABG-related nonfatal TIMI major bleeding	57/262 (23.0)	39/256 (16.0)	1.54 (1.02–2.32)	0.04	
No history of stroke or TIA					
Death from cardiovascular causes, nonfatal MI, or nonfatal stroke (primary efficacy end point)	596/6551 (9.5)	746/6539 (12.0)	0.79 (0.71–0.88)	<0.001	0.02
Non-CABG-related TIMI major bleeding	132/6484 (2.3)	105/6464 (1.8)	1.26 (0.97–1.62)	0.08	0.22
Death from any cause, nonfatal MI, nonfatal stroke, or non-CABG-related nonfatal TIMI major bleeding	727/6551 (11.8)	854/6539 (13.8)	0.84 (0.76–0.93)	<0.001	0.006

Ces résultats suggèrent une interaction entre les antécédents d'AVC et l'efficacité du prasugrel, sans grande modification du surrisque hémorragique. Cette modification d'effet pouvant peut-être conduire à une balance bénéfice risque potentiellement favorable chez les patients sans antécédents, suggérée par une réduction du critère composite de bénéfice net possiblement de 16% (hazard ratio 0.84).

Cette analyse étant purement post-hoc, elle ne constitue pas un argument suffisamment fiable pour positionner ce produit dans la stratégie thérapeutique chez ces patients.

8 Approche qualitative de la balance bénéfice risque

La balance bénéfice risque d'un nouveau traitement peut aussi être remise en cause de manière qualitative lorsque les événements indésirables sont de gravité bien supérieure à la pathologie traitée. La balance bénéfice risque sera défavorable lorsque le risque encouru est disproportionné par rapport au bénéfice apporté. C'est le cas, par exemple, de la survenue de syndromes de Lyell potentiellement mortels avec la prise d'AINS pour soulager des traumatismes bénins. En revanche, avec des chimiothérapies anticancéreuses, un risque faible de syndrome de Lyell est acceptable compte tenu de la gravité de la maladie et du bénéfice attendu sur la mortalité.

En fait, la gravité des événements indésirables est à confronter, non pas à la gravité de la maladie, mais à la nature du bénéfice. Ainsi, la survenue de syndromes de Lyell avec un antiémétique est peu acceptable même dans un contexte oncologique.

Tableau 1 – Quelques situations où le risque lié aux effets indésirables a été jugé excessif par rapport au bénéfice apporté

Situation	Risque lié aux effets indésirables	Bénéfice prouvé
Alosetron dans le côlon irritable [53]	Infarctus mésentérique et occlusion intestinale (dont certains cas mortels). Ce traitement a été retiré de la commercialisation 9 mois après son introduction sur le marché aux USA à la suite du décès de 5 patients	Diminution de l'inconfort et des douleurs abdominales
Troglitazone dans le diabète de type 2	Insuffisance hépatique conduisant dans certains cas au décès ou à la transplantation. Le troglitazone a été retiré du marché 3 ans après sa commercialisation à la suite de 90 cas d'insuffisance hépatique dont 70 mortels ou ayant nécessités une transplantation.	Diminution des taux d'hémoglobine glycosylée (pas d'effet prouvé sur les complications du diabète)

Cette évaluation repose sur un jugement de valeur et va donc dépendre de l'appréciation qui est faite de la gravité des événements indésirables et de l'importance de leur fréquence de survenue. Cette appréciation repose sur des échelles de valeur qui peuvent être différentes suivants les médecins, les patients et qui dépendent des choix sociétaux. Ces éléments expliquent les différences d'appréciation des mêmes données qui se rencontrent parfois dans l'évaluation de la balance bénéfice risque.

9 Références

- 1 Guo JJ, Pandey S, Doyle J, et al. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. *Value Health* 2010;13:657–66 doi:10.1111/j.1524-4733.2010.00725.x; PMID:20412543;
- 2 Bowman L, Mafham M, Wallendszus K, et al. Effects of Aspirin for Primary Prevention in Persons with Diabetes Mellitus. *N Engl J Med* 2018;379:1529–39 doi:10.1056/NEJMoa1804988; PMID:30146931;
- 3 Roe MT, Armstrong PW, Fox KAA, et al. Prasugrel versus clopidogrel for acute coronary syndromes without revascularization. *N Engl J Med* 2012;367:1297–309 doi:10.1056/NEJMoa1205512; PMID:22920930;
- 4 Altman DG, Bland JM. Absence of evidence is not evidence of absence. *BMJ* 1995;311:485 doi:10.1136/bmj.311.7003.485; PMID:7647644;
- 5 Furberg CD, Pitt B. Withdrawal of cerivastatin from the world market. *Curr Control Trials Cardiovasc Med* 2001;2:205–07 doi:10.1186/cvm-2-5-205; PMID:11806796;
- 6 FDA Announces New Recommendations On Evaluating Cardiovascular Risk In Drugs Intended To Treat Type 2 Diabetes 2022. Available at: <https://www.pharmaceuticalonline.com/doc/fda-announces-new-recommendations-on-0001> Accessed November 19, 2022.
- 7 FDA. Diabetes Mellitus — Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes: Guidance for Industry.
- 8 Low Wang CC, Everett BM, Burman KD, et al. Cardiovascular Safety Trials for All New Diabetes Mellitus Drugs? *Circulation* 2019;139:1741–43 doi:10.1161/CIRCULATIONAHA.118.038771; PMID:30933621;
- 9 Chertow GM, Pergola PE, Farag YMK, et al. Vadadustat in Patients with Anemia and Non-Dialysis-Dependent CKD. *N Engl J Med* 2021;384:1589–600 doi:10.1056/nejmoa2035938; PMID:33913637;
- 10 Mackenzie IS, Ford I, Nuki G, et al. Long-term cardiovascular safety of febuxostat compared with allopurinol in patients with gout (FAST): a multicentre, prospective, randomised, open-label, non-inferiority trial. *The Lancet* 2020;396:1745–57 doi:10.1016/S0140-6736(20)32234-0; PMID:33181081;
- 11 Becker MA, Schumacher HR, Wortmann RL, et al. Febuxostat compared with allopurinol in patients with hyperuricemia and gout. *N Engl J Med* 2005;353:2450–61 doi:10.1056/NEJMoa050373; PMID:16339094;
- 12 White WB, Chohan S, Dabholkar A, et al. Cardiovascular safety of febuxostat and allopurinol in patients with gout and cardiovascular comorbidities. *Am Heart J* 2012;164:14–20 doi:10.1016/j.ahj.2012.04.011; PMID:22795277;
- 13 Locatelli F, Hannedouche T, Fishbane S, et al. Cardiovascular Safety and All-Cause Mortality of Methoxy Polyethylene Glycol-Epoetin Beta and Other Erythropoiesis-Stimulating Agents in Anemia of CKD: A Randomized Noninferiority Trial. *Clin J Am Soc Nephrol* 2019;14:1701–10 doi:10.2215/CJN.01380219; PMID:31420350;
- 14 Austeng D, Morken TS, Bolme S, et al. Nurse-administered intravitreal injections of anti-VEGF: study protocol for noninferiority randomized controlled trial of safety, cost and patient satisfaction. *BMC Ophthalmol* 2016;16:169 doi:10.1186/s12886-016-0348-4; PMID:27716253;
- 15 Walsh M, Manns B, Garg AX, et al. The Safety of Eplerenone in Hemodialysis Patients: A Noninferiority Randomized Controlled Trial. *Clin J Am Soc Nephrol* 2015;10:1602–08 doi:10.2215/CJN.12371214; PMID:26138259;
- 16 Lemaire M, Islam QS, Shen H, et al. Iron-containing micronutrient powder provided to children with moderate-to-severe malnutrition increases hemoglobin concentrations but not the risk of infectious

- morbidity: a randomized, double-blind, placebo-controlled, noninferiority safety trial. *The American Journal of Clinical Nutrition* 2011;94:585–93 doi:10.3945/ajcn.110.009316; PMID:21715512;
- 17 Yamada S, Fujii T, Sonohara F, et al. Safety of Combined Division vs Separate Division of the Splenic Vein in Patients Undergoing Distal Pancreatectomy: A Noninferiority Randomized Clinical Trial. *JAMA Surg* 2021;156:418–28 doi:10.1001/jamasurg.2021.0108; PMID:33656542;
 - 18 Koster RW, Beenen LF, van der Boom EB, et al. Safety of mechanical chest compression devices AutoPulse and LUCAS in cardiac arrest: a randomized clinical trial for non-inferiority. *Eur. Heart J.* 2017;38:3006–13 doi:10.1093/eurheartj/ehx318; PMID:29088439;
 - 19 Reith C, Baigent C, Blackwell L, et al. Effect of statin therapy on muscle symptoms: an individual participant data meta-analysis of large-scale, randomised, double-blind trials. *The Lancet* 2022;400:832–45 doi:10.1016/S0140-6736(22)01545-8;
 - 20 Wang C-H, Li C-H, Hsieh R, et al. Proton pump inhibitors therapy and the risk of pneumonia: a systematic review and meta-analysis of randomized controlled trials and observational studies. *Expert Opinion on Drug Safety* 2019;18:163–72 doi:10.1080/14740338.2019.1577820; PMID:30704306;
 - 21 Wiviott SD, Braunwald E, McCabe CH, et al. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med* 2007;357:2001–15 doi:10.1056/NEJMoa0706482; PMID:17982182;
 - 22 Eggermont AMM, Chiarion-Sileni V, Grob J-J, et al. Prolonged Survival in Stage III Melanoma with Ipilimumab Adjuvant Therapy. *N Engl J Med* 2016;375:1845–55 doi:10.1056/NEJMoa1611299; PMID:27717298;
 - 23 Michielin O, van Akkooi ACJ, Ascierto PA, et al. Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology* 2019;30:1884–901 doi:10.1093/annonc/mdz411; PMID:31566661;
 - 24 Gail MH. Using absolute risks to assess the risks and benefits of treatment. *Thorax* 2014;69:604–05 doi:10.1136/thoraxjnl-2014-205175; PMID:24550059;
 - 25 Bonaca MP, Bhatt DL, Cohen M, et al. Long-term use of ticagrelor in patients with prior myocardial infarction. *N Engl J Med* 2015;372:1791–800 doi:10.1056/NEJMoa1500857; PMID:25773268;
 - 26 Wallentin L, Becker RC, Budaj A, et al. Ticagrelor versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med* 2009;361:1045–57 doi:10.1056/NEJMoa0904327; PMID:19717846;
 - 27 Baigent C, Emberson J, Haynes R, et al. Impact of diabetes on the effects of sodium glucose co-transporter-2 inhibitors on kidney outcomes: collaborative meta-analysis of large placebo-controlled trials. *The Lancet* 2022 doi:10.1016/S0140-6736(22)02074-8;
 - 28 Steg PG, Bhatt DL. Is There Really a Benefit to Net Clinical Benefit in Testing Antithrombotics? *Circulation* 2018;137:1429–31 doi:10.1161/CIRCULATIONAHA.117.033442; PMID:29610124;
 - 29 Barnett AS, Cyr DD, Goodman SG, et al. Net clinical benefit of rivaroxaban compared with warfarin in atrial fibrillation: Results from ROCKET AF. *International Journal of Cardiology* 2018;257:78–83 doi:10.1016/j.ijcard.2017.06.110; PMID:29506743;
 - 30 Pocock SJ, Ariti CA, Collier TJ, et al. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur. Heart J.* 2012;33:176–82 doi:10.1093/eurheartj/ehr352; PMID:21900289;
 - 31 Péron J, Giai J, Maucort-Boulch D, et al. The Benefit-Risk Balance of Nab-Paclitaxel in Metastatic Pancreatic Adenocarcinoma. *Pancreas* 2019;48:275–80 doi:10.1097/MPA.0000000000001234; PMID:30629024;
 - 32 Buyse M, Saad ED, Peron J, et al. The Net Benefit of a treatment should take the correlation between benefits and harms into account. *Journal of Clinical Epidemiology* 2021;137:148–58 doi:10.1016/j.jclinepi.2021.03.018; PMID:33774140;

- 33 Ioannidis JPA, Lau J. Improving safety reporting from randomised trials. *Drug Safety* 2002;25:77–84 doi:10.2165/00002018-200225020-00002; PMID:11888350;
- 34 Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437–43 doi:10.1001/jama.285.4.437; PMID:11242428;
- 35 Duffy J, Hirsch M, Pealing L, et al. Inadequate safety reporting in pre-eclampsia trials: a systematic evaluation. *BJOG* 2018;125:795–803 doi:10.1111/1471-0528.14969; PMID:29030992;
- 36 Favier R, Crépin S. The reporting of harms in publications on randomized controlled trials funded by the "Programme Hospitalier de Recherche Clinique," a French academic funding scheme. *Clin Trials* 2018;15:257–67 doi:10.1177/1740774518760565; PMID:29498543;
- 37 Sivendran S, Latif A, McBride RB, et al. Adverse event reporting in cancer clinical trial publications. *JCO* 2014;32:83–89 doi:10.1200/JCO.2013.52.2219; PMID:24323037;
- 38 Pitrou I, Boutron I, Ahmad N, et al. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756–61 doi:10.1001/archinternmed.2009.306; PMID:19858432;
- 39 Edwards A, Elwyn G, Mulley A. Explaining risks: turning numerical data into meaningful pictures. *BMJ* 2002;324:827–30 doi:10.1136/bmj.324.7341.827; PMID:11934777;
- 40 Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *New Engl J Med* 1988;318:1728–33 doi:10.1056/NEJM198806303182605; PMID:3374545;
- 41 Granger CB, Alexander JH, McMurray JJV, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365:981–92 doi:10.1056/NEJMoa1107039; PMID:21870978;
- 42 Valgimigli M, Frigoli E, Heg D, et al. Dual Antiplatelet Therapy after PCI in Patients at High Bleeding Risk. *N Engl J Med* 2021;385:1643–55 doi:10.1056/NEJMoa2108749; PMID:34449185;
- 43 Sibbing D, Aradi D, Jacobshagen C, et al. Guided de-escalation of antiplatelet treatment in patients with acute coronary syndrome undergoing percutaneous coronary intervention (TROPICAL-ACS): a randomised, open-label, multicentre trial. *The Lancet* 2017;390:1747–57 doi:10.1016/S0140-6736(17)32155-4; PMID:28855078;
- 44 Raskob GE, van Es N, Verhamme P, et al. Edoxaban for the Treatment of Cancer-Associated Venous Thromboembolism. *N Engl J Med* 2018;378:615–24 doi:10.1056/NEJMoa1711948; PMID:29231094;
- 45 Cates C. Confidence intervals for the number needed to treat. Pooling numbers needed to treat may not be reliable. *BMJ* 1999;318:1764–65 doi:10.1136/bmj.318.7200.1764c; PMID:10381732;
- 46 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21:1575–600 doi:10.1002/sim.1188; PMID:12111921;
- 47 Zhao Y, Slate EH, Xu C, et al. Empirical comparisons of heterogeneity magnitudes of the risk difference, relative risk, and odds ratio. *Syst Rev* 2022;11:26 doi:10.1186/s13643-022-01895-7; PMID:35151340;
- 48 Poole C, Shrier I, VanderWeele TJ. Is the Risk Difference Really a More Heterogeneous Measure? *Epidemiology* 2015;26:714–18 doi:10.1097/EDE.0000000000000354; PMID:26196684;
- 49 Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;311:1356–59 doi:10.1136/bmj.311.7016.1356; PMID:7496291;
- 50 Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *New Engl J Med* 1980;302:1109–17 doi:10.1056/NEJM198005153022003; PMID:7366635;
- 51 Aguilar MI, Hart R. Oral anticoagulants for preventing stroke in patients with non-valvular atrial fibrillation and no previous history of stroke or transient ischemic attacks. *Cochrane Database Syst Rev* 2005:CD001927 doi:10.1002/14651858.CD001927.pub2; PMID:16034869;

- 52 Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLOS Medicine* 2013;10:e1001381 doi:10.1371/journal.pmed.1001381; PMID:23393430;
- 53 Camilleri M, Northcutt AR, Kong S, et al. Efficacy and safety of alosetron in women with irritable bowel syndrome: a randomised, placebo-controlled trial. *The Lancet* 2000;355:1035–40 doi:10.1016/S0140-6736(00)02033-X; PMID:10744088;