



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 7 - Les surrogates endpoints

Comité de rédaction et relecture (par ordre alphabétique)

Jean Luc Cracowski

Michel Cucherat

Dominique Deplanque

Behrouz Kassai

Charles Khouri

Silvy Laporte

Jean-Christophe Lega

Clara Locher

Florian Naudet

Edouard Ollier

Matthieu Roustit



[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	Pourquoi une action sur le critère intermédiaire peut ne pas apporter de bénéfice clinique	13
2.1	Les effets off-target	13
2.2	Effets indésirables	14
2.3	Complexité des mécanismes reliant critères intermédiaires et critères cliniques	15
2.4	Autres raisons.....	16
3	Méthode de validation des surrogates	17
3.1	Corrélation des effets.....	17
3.2	Autres approches statistiques.....	22
4	Limites des études de validation de surrogate [new]	23
4.1	Limites liées à la nature rétrospective des études, HARKing, p hacking.....	23
4.2	Le biais écologique [new].....	25
5	Extrapolabilité/transposabilité.....	26
6	Méta-épidémiologie, situation actuelle	27
7	Conclusion	29

1 Introduction

Avant d'être utilisé en pratique, un nouveau traitement doit apporter la preuve qu'il améliore le devenir des patients sur des critères cliniquement pertinents (démonstration d'un bénéfice cliniquement pertinent).

Par rapport à la simple mise en évidence d'un effet pharmacologique sur un critère intermédiaire, cette démonstration demande un plus grand nombre de patients ou une durée de suivi des patients plus importante.

Par exemple, la mise en évidence de l'effet d'un traitement contre l'ostéoporose sur la prévention des fractures du col du fémur chez la personne âgée nécessite un suivi de plusieurs années, le but du traitement étant de ralentir la perte de densité osseuse survenant au fil des années après la ménopause pour retarder le moment où le seuil fracturaire est atteint

D'où l'idée de chercher si certains critères intermédiaires ne pourraient pas être des surrogates (critères de substitution), c'est-à-dire des critères pour lesquels l'observation d'un effet d'un traitement à leur niveau impliquerait forcément un effet sur le critère clinique approprié. Dans ce cas, la simple mise en évidence d'un effet pharmacologique sur ce type de critère équivaldrait à la démonstration du bénéfice clinique, sans devoir le démontrer directement dans l'essai sur le critère clinique.

Avec l'exemple précédent, cela revient à s'interroger si la mise en évidence d'un effet sur la densité osseuse ne suffirait pas à conclure que le nouveau traitement réduit la fréquence des fractures. Si c'est le cas, un essai portant la densité osseuse demandera nettement moins de sujets et de suivi qu'un essai sur les fractures, mais montrera la valeur du traitement pour la prévention des fractures (sans avoir besoin de le démontrer par lui-même).

Cependant, tous les critères intermédiaires ne sont pas des surrogates, et il est nécessaire de le démontrer spécifiquement, au cas par cas et par une approche rigoureuse qu'un critère intermédiaire a bien les propriétés nécessaires pour être un surrogate.

Critères cliniques. Les critères cliniques sont des critères de jugement qui correspondent directement avec la problématique clinique que pose la maladie. La mise en évidence d'un effet du traitement à leur niveau assure que celui-ci apporte bien le bénéfice clinique recherché. Par exemple dans l'hypertension artérielle, les critères cliniques sont les événements cardiovasculaires (infarctus du myocarde ou AVC) mortels et non mortels ou les décès de cause cardiovasculaire ; pour les cancers au stade avancé ou métastatique, le décès ; pour l'ostéoporose les fractures vertébrales ou des os longs ; etc.

Critères intermédiaires. Les critères intermédiaires sont de la catégorie des signes des maladies ou de leur mécanisme d'action. Ils ne mesurent pas directement les conséquences de la maladie. Ils servent à montrer une activité biologique des traitements (activité anticancer, action pharmacologique). La mise en évidence de l'effet d'un traitement à leur niveau est insuffisante,

car rien ne garantit que cet effet se traduit par un bénéfice clinique. C'est par exemple la pression artérielle systolique dans l'hypertension artérielle, la réponse tumorale avec un traitement du cancer ou la densité osseuse dans l'ostéoporose.

Pour être critère de substitution (*surrogate*) et, ainsi, être utilisé dans un essai randomisé à la place d'un critère clinique, un critère intermédiaire doit garantir que l'observation d'un **effet** d'un traitement à son niveau **prédit avec certitude** un **effet** de ce traitement au niveau du **critère clinique** [1].

*D'emblée il faut insister qu'il s'agit ici de la prédiction d'un effet traitement (objectivé par un hazard ratio, ou une différence de moyenne par exemple) par un autre effet traitement (objectivé lui aussi par un hazard ou une différence de moyenne par exemple) et non pas, de la prédiction individuelle de la probabilité de survenue d'un évènement clinique par un facteur pronostique ou un facteur de risque (cf. section **Erreur ! Source du renvoi introuvable.**)*

La FDA définit les surrogates de la manière suivante¹ :

« a surrogate endpoint is a marker, such as a laboratory measurement, radiographic image, physical sign, or other measure that is thought to predict clinical benefit, but is not itself a measure of clinical benefit »

Le critère de substitution est une problématique d'évaluation des traitements : trouver un substitut aux critères cliniques afin de permettre la réalisation d'essais randomisés plus courts et nécessitant peu de patients. Il ne s'agit donc pas d'une problématique médicale, comme comprendre par quel mécanisme passe le bénéfice clinique apporté par un traitement. Cette recherche d'explication est le domaine de la « *mediation analysis* » [3].

Deux guidelines de rédaction SPIRIT et CONSORT précisent les informations indispensables à mentionner dans les protocoles (SPIRIT-surrogate [2]) et les publications (CONSORT-surrogate [3]) des essais cliniques ayant comme critère de jugement principal un surrogate.

On peut faire le constat que dans la littérature, les termes de proxy, critère intermédiaire et de surrogate sont souvent interchangeable, ce qui témoigne de la confusion dans les concepts. La Food and Drug Administration a produit une liste de critères intermédiaires acceptés pour l'enregistrement des médicaments en utilisant le terme surrogate (<https://www.fda.gov/drugs/development-resources/surrogate-endpoint-resources-drug-and-biologic-development>). Une analyse de la pertinence de ces critères a montré que ces critères ne relevaient pas de la surrogacy dans les champs médicaux non-oncologiques [4]. L'association entre critère intermédiaire et critère clinique ne fait pas un surrogate

L'existence d'une association entre un critère intermédiaire (biomarqueurs) et un critère clinique (décès, évènement) n'est pas suffisante pour faire de ce critère intermédiaire un surrogate. Cette association (qui peut être observée au niveau épidémiologique par exemple) ne garantit pas qu'une modification provoquée du critère intermédiaire par le traitement se traduira automatiquement par un bénéfice proportionné au niveau du critère clinique. Cette association démontre seulement la

¹ FDA Guidance for Industry Expedited Programs for Serious Conditions—Drugs and Biologics

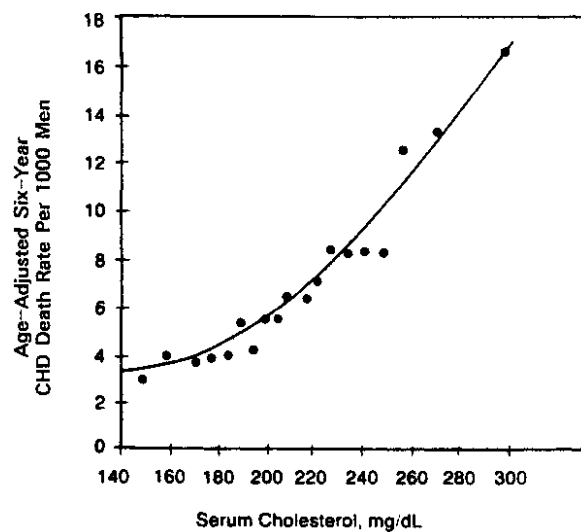
valeur pronostique du critère intermédiaire. Dans le jargon des méthodes de validation des surrogates elle est dénommée « *subject level correlation* ».

En effet, par définition, une association de ce type montre qu'il existe une relation statistique à travers les patients entre les valeurs du critère intermédiaire (par exemple la pression artérielle systolique) et celles du critère clinique (par exemple la probabilité de faire un AVC à 5 ans). Plus un sujet a une valeur élevée (ou l'inverse) au niveau du critère intermédiaire, plus il est probable, en moyenne, qu'il aura une valeur élevée sur le critère clinique (d'où l'appellation de subject level correlation).

*La définition du surrogate se réfère quant à elle à une association entre les effets traitements et non pas entre les critères. Comme les effets traitements ne peuvent être mesurés qu'à l'aide d'un essai, cette association sera dénommée « trial level correlation » (cf. section **Erreur ! Source du renvoi introuvable.**).*

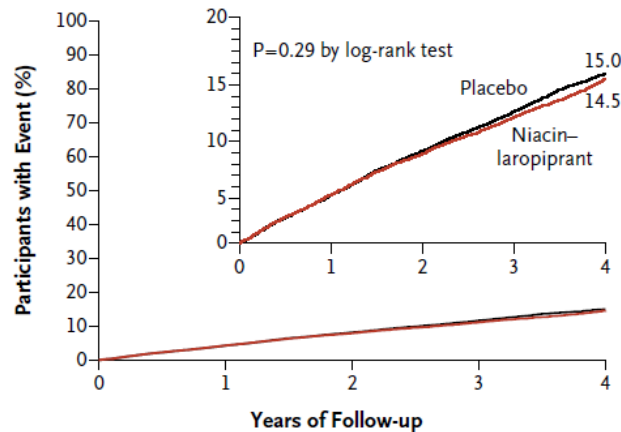
En effet, de très nombreux exemples montrent qu'une association entre le critère intermédiaire et le critère clinique n'est pas suffisante pour garantir que tout traitement agissant sur le critère intermédiaire provoquera un bénéfice sur le critère clinique.

Epidémiologiquement, il existe une forte relation au niveau de sujets, entre le niveau de cholestérolémie et le risque cardiovasculaire [5][6]. Cette relation a été très bien identifiée dans les études épidémiologiques historiques comme MRFIT ou l'étude de Framingham :



Association statistique entre la cholestérolémie et le risque de décès coronariens identifiée dans l'étude épidémiologique MRFIT [7].

Cependant plusieurs molécules hypocholestérolémiantes n'ont pas montré de bénéfice clinique sur les événements coronariens comme l'acide nicotinique dans l'essai de grande taille HPS2-THRIVE [8] :



No. at Risk						
Niacin-laropiprant	12,838	12,232	11,517	7672	4978	
Placebo	12,835	12,247	11,523	7643	5036	

Et cela malgré un effet pharmacologique sur les différents paramètres lipidiques :

Table S2: Effects of niacin/laropiprant on lipids

Year of follow-up	Lipid parameter (mg/dL)					
	Total cholesterol	LDL cholesterol	HDL cholesterol	Triglycerides	Apo A1	Apo B
<1	-8 (1.3)	-12 (1.1)	6 (0.4)	-35 (3.4)	5 (0.8)	-9 (0.8)
≥1 <2	-6 (0.4)	-10 (0.4)	7 (0.2)	-33 (1.3)	7 (0.3)	-8 (0.3)
≥2 <3	-5 (0.9)	-9 (0.7)	6 (0.3)	-31 (2.3)	7 (0.6)	-7 (0.5)
≥3 <4	-3 (0.9)	-7 (0.7)	6 (0.3)	-33 (2.6)	7 (0.6)	-6 (0.5)
≥4	-3 (0.6)	-7 (0.5)	6 (0.2)	-32 (1.8)	7 (0.4)	-5 (0.4)
Study average	-5 (0.4)	-10 (0.3)	6 (0.1)	-33 (1.1)	7 (0.3)	-7 (0.2)

Absolute differences (with standard error) in lipid fraction during the trial. Study averages are weighted for participant-years at risk within region. Follow-up periods are as in supplementary table 1. LDL: low density lipoprotein. HDL: high density lipoprotein. Apo A1: apolipoprotein A1. Apo B: apolipoprotein B.

Un autre essai similaire de grande taille de la niacine, AIM-HIGH [7], donne le même résultat.

Le fluorure de sodium stimule la formation osseuse et accroît la densité osseuse. En raison de ces effets, il a été envisagé comme traitement préventif de l'ostéoporose de la femme ménopausée. Dans un essai randomisé [9] contre placebo réalisé chez 202 femmes ménopausées et suivies pendant 4 ans, le fluorure de sodium augmente la densité osseuse de la colonne lombaire de façon hautement significative ($p < 0,001$). Cependant, des nouvelles fractures vertébrales surviennent plus fréquemment sous fluorure de sodium que sous placebo (163 comparé à 136) ainsi que des fractures non vertébrales (72 versus 24, $p = 0,01$). Cet essai conclut que la forme de fluorure utilisé augmente certains aspects de la densité osseuse, mais rend les os plus fragiles. L'utilisation de la densité osseuse comme critère de substitution n'aurait pas permis de mettre en évidence ces effets délétères qui enlèvent pourtant tout intérêt au traitement.

La simple existence de ces contre-exemples montre qu'une association au niveau des sujets entre les critères intermédiaire et clinique (subject level correlation) n'implique pas que le critère intermédiaire soit un surrogate et qu'il soit possible d'inférer un bénéfice clinique à partir d'un effet sur le critère intermédiaire. Le Tableau 1 liste une série d'exemples où des essais ont réfuté la valeur de surrogacy de simple association entre critères au niveau sujet pourtant bien établies, en ne montrant pas de bénéfice clinique de molécule impactant pourtant favorablement le critère intermédiaire.

Tableau 1 - Exemples de réfutation de la valeur de surrogacy d'une association entre critères au niveau sujet pourtant bien établie par des molécules ayant bien une action sur le biomarqueur, mais qui ont échouées à mettre en évidence un bénéfice clinique.

Biomarqueurs ayant un lien avec les critères cliniques et susceptibles d'être des surrogates potentiels	Essai qui ont échoué à mettre en évidence un bénéfice clinique avec une molécule ayant une action sur le biomarqueur
HbA1 et événements cardiovasculaires (macrovasculaires)	Essai d'intensification PREVENT [ref]
Vitamine D et événements cardiovasculaires	Méta-analyse de 21 essais randomisés [10]
Vitamine D et événements osseux en l'absence de carence	Essai VITAL [11]
LDL et événements cardiovasculaires	HPS2-THRIVE [8] et AIM-HIGH [7] avec la niacine
HDL et événements cardiovasculaires	ILLUMINATE [6] avec le torcétrapib
Triglycérides et événements cardiovasculaires	FIELD avec le fénofibrate [12]. PROMINENT avec le pemafibrate [13]
PFS dans les cancers du sein avec mutation BRCA	ARIEL 4 avec le rucaparib en 3ème ligne suggérant une augmentation de la mortalité conduisant au retrait de l'AMM ² pour cette indication, AMM donnée sur la base des résultats de PFS [14].

Il existe des contre-exemples même pour des facteurs qui peuvent raisonnablement être considérés comme causaux (en raison de la force de l'association, de l'existence d'une relation dose effet et d'une plausibilité biologique forte ou en raison d'une démonstration de causalité apportée par une étude de randomisation mendélienne robuste). Sur ce point, il est souvent assez difficile, cognitivement, d'accepter qu'une relation causale n'est pas suffisant pour établir un surrogate. La suppression d'une cause entraîne forcément la disparition des conséquences de la cause. Certes, mais l'action qui supprime la cause (ou plus vraisemblablement la réduit) peut avoir d'autres actions qui s'avèrent délétères et impacte négativement le critère clinique : effets indésirables graves (cf. section 2.2), effets off-target (cf. section 2.1) ou mécanisme plus complexe qu'une cause unique et directe (cf. section 2.3). De ce fait, même en supprimant une cause d'un événement clinique, rien ne garantit, avant que cela ne soit parfaitement démontré, que cette action aura un bilan positif sur le critère clinique et apportera, au total, un bénéfice clinique.

Ainsi, tout facteur pronostique ou tout facteur de risque n'est pas, de-facto, un surrogate pour l'évaluation des traitements. Les surrogates doivent vérifier des propriétés qui leur sont propres et qui vont au-delà de la simple association statistique critère intermédiaire – critère clinique.

Parfois, un effet pharmacologique est considéré comme étant un surrogate à partir du moment où il existe un essai clinique qui a montré un bénéfice sur le critère clinique avec une molécule produisant cet effet pharmacologique [15]. Par exemple, comme les essais des statines montrent une réduction de la morbi-mortalité cardiovasculaire et que les statines entraînent une baisse du LDL cholestérol, le LDL est un surrogate des événements cardiovasculaires. Ce raisonnement ne peut pas être retenu par manque de

² <https://www.ema.europa.eu/en/news/ema-recommends-restricting-use-cancer-medicine-rubraca>

représentativité (est-ce que toutes les molécules de différentes classes produisant cet effet pharmacologie ont conduit à un bénéfice sur le critère clinique ?) et en raison de nombreux facteurs de confusion possible (bénéfice apporté par un autre mécanisme que l'effet pharmacologique considéré). De nombreux contre-exemples existent de molécules ayant échoué à montrer un bénéfice clinique alors qu'elles produisent le même effet pharmacologique que d'autres molécules ayant, quant à elles, parfaitement démontré un bénéfice clinique (comme les exemples de la niacine, du fluorure de sodium développé précédemment).

2 Pourquoi une action sur le critère intermédiaire peut ne pas apporter de bénéfice clinique

2.1 Les effets off-target

Agir sur un facteur causal du critère clinique n'implique pas automatiquement un bénéfice clinique. En effet, supprimer la cause d'un événement clinique ne garantit pas le bénéfice à coup sûr, car le traitement peut entraîner par lui-même (indépendamment de son action sur la cause) un effet délétère par un effet off-target. Il s'agit d'un effet sur le critère clinique autre que celui recherché avec le médicament, souvent inconnu avant la réalisation des essais, et qui va contrebalancer le « bénéfice » apporté par l'action sur la cause de l'événement.

Ces effets off-target sont inconnus avant la réalisation des essais cliniques et sont la principale justification de la réalisation des essais de confirmation.

Historiquement, le recours aux essais thérapeutiques a été motivé par le constat que les raisonnements basés sur les mécanismes d'action, les effets pharmacologiques, étaient faillibles et qu'ils ne permettaient pas de prédire avec suffisamment de fiabilité le réel bénéfice clinique des candidats traitements. Un de ces premiers exemples fût des antiarythmiques de classe Ic [4]. Il est donc apparu **indispensable de vérifier de manière expérimentale, « grandeur nature »**, que les nouveaux médicaments apportaient bien le bénéfice que l'on attendait en théorie et non pas l'inverse. C'est le fondement de la demande d'essais cliniques pour la régulation et de l'evidence based medicine. Même avec les progrès réalisés dans la connaissance du vivant, cette vérification reste toujours indispensable aujourd'hui, car en moyenne elle ne confirme pas la théorie dans environ 50% des cas [16].

L'exemple du torcetrapib illustre parfaitement cette problématique.

Le taux de HDL cholestérol est un facteur de risque protecteur de la survenue des événements cardiovasculaires et la mortalité cardiovasculaire comme le montre de nombreuses études épidémiologiques.

Les inhibiteurs de la CETP (cholesteryl ester transfer protein) comme le torcetrapid entraînent une augmentation notable du HDL cholestérol (et une diminution du LDL cholestérol). Il était donc attendu une réduction des événements cardiovasculaires et de la mortalité induite, étant donné que ce produit avait une action sur deux déterminants majeurs de l'athérosclérose. Entre autres, des études de randomisation mendélienne évoquent un rôle causal du LDL et du HDL dans l'athérosclérose.

Cependant, la vérification de cette théorie par l'essai pivot ILLUMINATE [6] n'a pas confirmé cette attente. Bien au contraire, l'essai ILLUMINATE a mis en évidence une augmentation de la mortalité cardiovasculaire et totale. Une augmentation de 72.1% du HDL cholestérol a été obtenue à 12 mois, mais cela s'est traduit par une augmentation des événements cardiovasculaires (hazard ratio 1.25 [1.09 ; 1.44] et des décès de toutes causes (HR 1.58 [1.14 ; 2.19])

Cette augmentation paradoxale de la mortalité est apparue, a posteriori, être en rapport à des effets off target de la molécule, non anticipés, comme la production de particules de HDL proathérogènes conduisant à l'effet inverse de celui escompté, avec un effet délétère sur la pression artérielle et des perturbations des électrolytes.

Le développement de l'acide nicotinique (niacine) comme traitement préventif cardiovasculaire documente aussi cette problématique des effets off-target inconnu.

La niacine est une molécule entraînant une baisse du LDL et une augmentation du HDL. Ces propriétés suggèrent que cette molécule pourrait prévenir les événements cardiovasculaires. Cette hypothèse a été vérifiée en 2011 et 2014 dans deux essais randomisés de morbi-mortalité de grande ampleur : AIM-HIGH[7] et HPS2-THRIVE [8]. Les deux ont échoué à montrer une réduction des événements cardiovasculaires.

Récemment sont apparus des résultats d'analyses métabolomiques montrant que des métabolites terminaux de la niacine induiraient une inflammation vasculaire et augmenterait ainsi le risque cardiovasculaire [9]. Ce phénomène pourrait ainsi expliquer l'échec des deux essais malgré les propriétés pharmacologiques apparemment bénéfiques au premier abord de la niacine.

Si les essais HPS2-THRIVE et AIM-HIGH avaient été remplacés par un raisonnement sur la nature causale de ces deux paramètres biologiques et la corrélation entre les effets sur le LDL et la réduction des événements coronariens observée avec les statines, ce raisonnement n'aurait pas inclus ces effets off-target, connus qu'en 2024 et aurait probablement conclu, à tort, au bénéfice clinique de la niacine.

Au-delà d'invalider la nature causale d'un critère intermédiaire comme seule justification de sa valeur de surrogacy, cette problématique a un autre retentissement sur le concept de surrogate. En effet, même si un critère surrogate a été validé à partir des essais randomisés sur critères cliniques de produits d'une certaine classe, rien ne garantit que de nouvelles molécules ou des molécules d'une nouvelle classe n'induiront pas des effets off-target non présent avec les molécules précédentes et qui remettent en cause la transposabilité de la relation de surrogacy à de nouvelles molécules. Ces effets off-target n'étant pas de réels effets indésirables potentiellement détectables par une approche de pharmacovigilance, mais de véritables effets délétères sur le critère clinique d'efficacité lui-même, il sera très difficile de les identifier a posteriori après commercialisation.

2.2 Effets indésirables

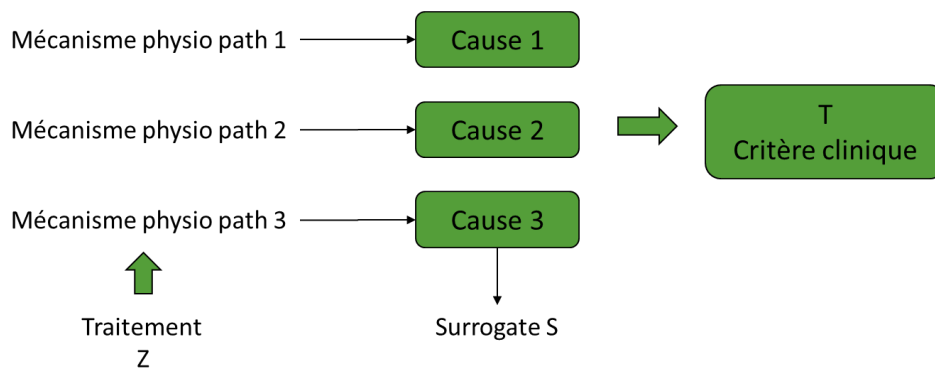
La notion de facteur de substitution reste aussi assez conceptuelle, car un critère de substitution ne prendra jamais en compte les effets délétères spécifiques des molécules à venir (cf. exemple du fluorure de sodium). En pratique, un essai sur critère clinique vérifie directement que les éventuels effets délétères d'un nouveau traitement ne contrebalancent pas le bénéfice. Une extrapolation du bénéfice clinique à partir d'un surrogate ne permet pas de faire cela, car les effets indésirables seront mal documentés et quantifiés (peu de patients, suivi court).

Plusieurs exemples illustrent ces réserves. La dissociation entre mortalité et effet sur les critères intermédiaires hémodynamiques observés dans l'insuffisance cardiaque avec les inhibiteurs de la phosphodiesterase III [17] en est un exemple. Les effets délétères de ces produits qui augmentent la mortalité étaient indépendants des mécanismes que reflétaient les critères intermédiaires.

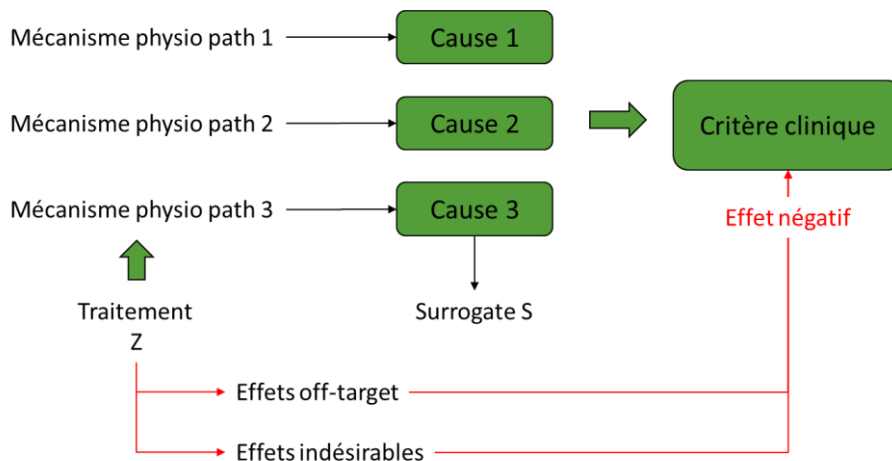
2.3 Complexité des mécanismes reliant critères intermédiaires et critères cliniques

Les mécanismes entre un critère intermédiaire et l'issue clinique de la maladie sont parfois complexes et ne correspondent pas forcément à un chemin de causalité unique et direct. De nombreuses situations différentes peuvent exister faisant que l'association entre les critères au niveau de sujet ne préjuge en rien que le critère intermédiaire puisse être un surrogate (ou que même une corrélation forte des effets (cf. section **Erreur ! Source du renvoi introuvable.**) implique que le critère intermédiaire soit un surrogate)

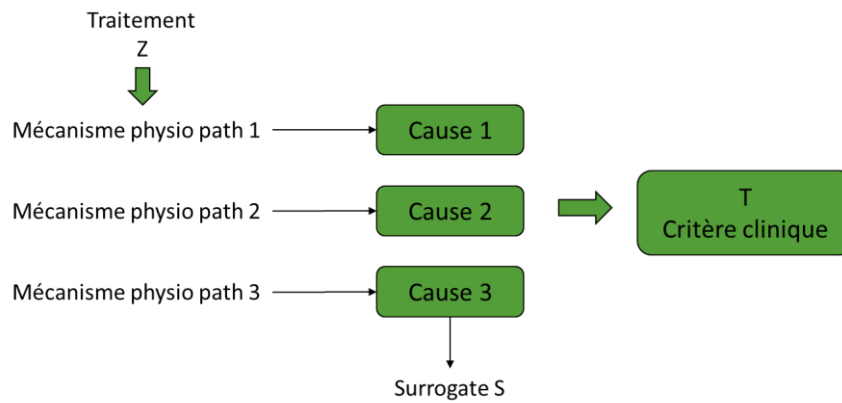
Un critère intermédiaire ne pourra être un vrai surrogate que s'il est directement lié à la cause de la réduction des événements cliniques qu'apporte le traitement :



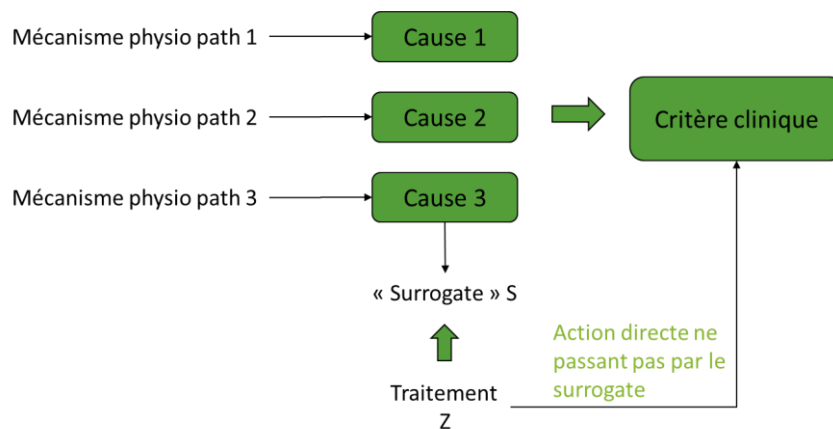
Le traitement ne doit pas induire des effets délétères sur le critère clinique passant par des effets off-target ou des effets délétères graves :



Mais il se peut très bien que le nouveau traitement apporte son bénéfice par une voie complètement différente de celle correspondant au traitement ayant permis d'évaluer la surrogacy :



Un autre cas de figure est une action des premiers traitements sur le surrogate par une voie qui n'impacte pas le critère clinique. Le traitement aura ainsi une double action indépendante sur le critère clinique et sur le surrogate. Un nouveau traitement impactant que le surrogate et sans effet sur la cause des évènements n'apportera pas de bénéfice malgré son apparente « efficacité sur le surrogate ».



La complexité de la relation entre les biomarqueurs et l'effet thérapeutique, souvent méconnue dans les raisonnements cliniques ou l'évaluation des produits de santé, plaident pour la réalisation de graphes acycliques dirigés causaux (causal directed acyclic graphs, DAG) afin d'explicitier les mécanismes (<https://www.nature.com/articles/s41390-018-0071-3>).

2.4 Autres raisons

Dans d'autres cas, ces contre-exemples peuvent simplement provenir du fait qu'il n'existe pas en réalité de relation entre le critère intermédiaire et le critère clinique. La relation précédemment observée provenait en réalité soit d'études peu robustes en raison de limites méthodologiques liées au HARKing, au p-hacking ou à des biais ; soit était due à autre chose que le facteur étudié (biais de confusion).

3 Méthode de validation des surrogates

Avant d'être utilisé en remplacement d'un critère clinique, un candidat surrogate doit démontrer qu'il possède un certain nombre de propriétés bien particulières. Cette démonstration passe classiquement par une approche de type méta-analyse³.

3.1 Corrélation des effets

La validation d'un critère de substitution repose actuellement sur la méthode proposée par M Buyse [18] et ses développements ultérieurs [19, 20, 21].

Le principe général est de montrer qu'il existe une très forte corrélation entre les **effets** des traitements sur le critère de substitution candidat et le critère clinique. En complément, il doit aussi exister une corrélation entre les valeurs du critère de substitution candidat et les valeurs du critère clinique (il s'agit du lien épidémiologique classique de type facteur pronostique ou facteur de risque, appelé dans ce contexte « *individual-level correlation* ») habituellement obtenue par les analyse à l'échelle individuelle de cohortes ou d'essais.

Cette approche repose sur la méta-analyse des données agrégées des essais randomisés du domaine et présente l'avantage d'être facilement mise en œuvre (contrairement à d'autres approches nécessitant des données individuelles d'essais).

Chaque essai donne une valeur de l'effet du traitement étudié par rapport au contrôle sur les 2 critères sous la forme d'un risque relatif, d'un hazard ratio ou d'une différence de moyenne, suivant la nature du critère de jugement. Classiquement ces données sont représentées sur un graphique bivarié (Figure 1) avec en abscisse l'effet sur le critère de substitution candidat et en ordonnée celui sur le critère clinique. Pour les effets relatifs, une échelle log doit être utilisée. Chaque essai est représenté par un point dont la taille est proportionnelle à l'effectif de l'étude. La droite de régression linéaire est tracée et le coefficient de corrélation r (ou son carré r^2 , noté aussi R^2) est calculé (appelé dans ce contexte « *trial-level correlation* »).

³ D'autres approches utilisant les données individuelles d'un essai sont aussi disponibles, cf. section ???

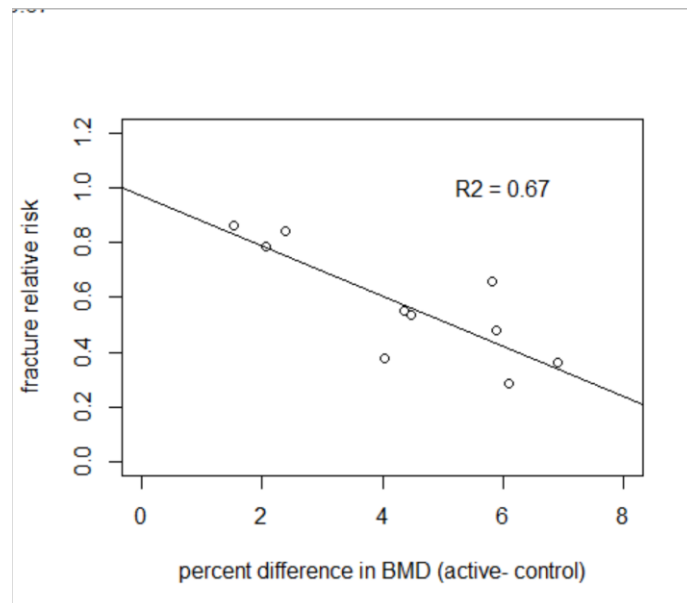


Figure 1 – Illustration de l’approche de validation d’un critère de substitution.

Chaque essai de la méta-analyse est représenté par un point (exemple inventé). La magnitude de l’effet observé sur le critère de substitution est représentée en abscisse. Ici il s’agit de la différence inter-groupe de l’augmentation de la densité osseuse. L’ordonnée représente l’effet du traitement sur le critère clinique. Ici il s’agit du risque relatif de fracture (rapport de la fréquence des fractures entre les 2 groupes de l’essai).

En effet, s’il existe une forte corrélation, c’est-à-dire un nuage de point très peu dispersé autour de la droite de régression, il y aura peu d’incertitude sur l’effet sur le critère clinique pour un effet sur le critère intermédiaire donné.

À l’opposé, en cas de faible corrélation, la variabilité des points autour de la droite est importante. Ainsi pour une valeur donnée d’effet sur le critère intermédiaire, l’effet sur le critère clinique ne pourra pas être prédit avec certitude du fait de cette variabilité. Il s’agit d’une variabilité résiduelle, non expliquée par l’effet sur le critère intermédiaire. L’effet sur le critère clinique dépend alors d’autres choses que l’effet sur le critère intermédiaire. L’effet sur le critère intermédiaire n’explique pas à lui seul l’effet sur le critère clinique et ne peut donc pas être utilisé comme surrogat.

La validité du critère de substitution se juge sur la valeur de ce coefficient de corrélation, qui doit être supérieur à 0.8 pour le coefficient de corrélation r (ou 0.65 pour le coefficient de détermination R^2)[22]. Plus récemment une contrainte supplémentaire a été introduite par l’IQWiG [23] en basant le raisonnement non pas sur la valeur du coefficient, mais sur la limite inférieure de son intervalle de confiance à 95%. Ainsi actuellement un critère de substitution ne sera formellement validé que si cette **borne inférieure de r est supérieure à 0.85 ($R^2 > 0.72$)**.

La prise en considération de l’intervalle de confiance est fondamentale pour s’assurer que l’estimation de cette corrélation est suffisamment précise pour garantir qu’il existe bien une vraie corrélation importante entre les effets à travers les essais. Compte tenu des enjeux des surrogates, il convient d’apporter une démonstration, donc, comme pour la mise en évidence de l’effet du traitement dans un essai, il est nécessaire d’avoir un haut degré de certitude statistique, qui est apporté par un intervalle de confiance étroit, assurant que, même au pire, la corrélation est forte [24].

En cas de forte corrélation (idéalement $r=1$), la valeur de l'effet sur le critère clinique est proportionnelle à la valeur de l'effet sur le critère de substitution, mais le plus souvent avec un facteur d'atténuation (il est classique d'observer que les magnitudes des effets traitement sur les critères intermédiaires sont plus importantes que celle des effets traitement sur les critères cliniques). De ce fait il est possible qu'un nouveau traitement ait un petit effet sur le critère de substitution et que celui-ci soit insuffisant pour induire un effet sur le critère clinique. Dans ce cas le traitement serait adopté à tort. Cette problématique est solutionnée en utilisant le « Surrogate threshold effect » (Figure 2) qui est le seuil d'effet minimal sur le critère de substitution qu'il faut obtenir avec le nouveau traitement pour être assuré d'un effet non nul sur le critère clinique [25].

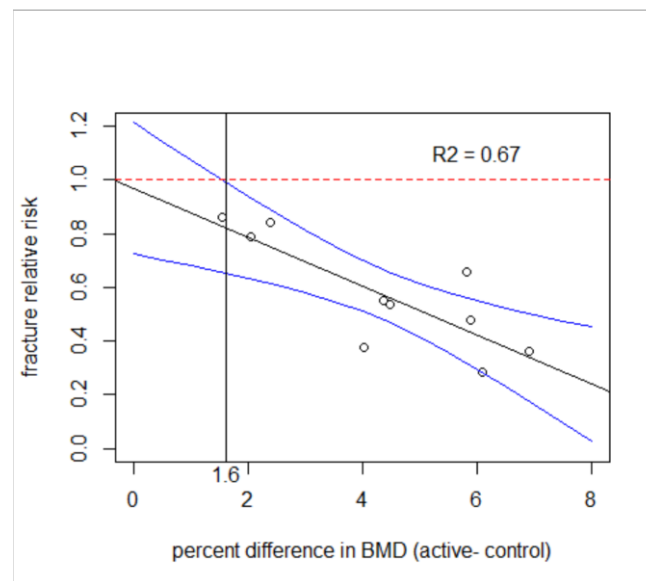


Figure 2 – Détermination du « Surrogate threshold effect » à partir de l'intervalle de prédiction.

Le seuil correspond à l'abscisse de l'intersection de la borne péjorative de cet intervalle de prédiction avec la ligne horizontale de l'effet nul sur le critère clinique (ligne discontinue rouge). Ici cette valeur est de 1.6 (ligne verticale noire). Si l'effet sur le critère de substitution est inférieur à une différence intergroupe de 1.6%, il n'est pas garanti que l'effet sur le critère clinique soit en faveur d'une réduction de la fréquence des fractures. Cependant, lorsque l'effet sur la densité osseuse est supérieur à une différence intergroupe de 1.6% (en faveur du groupe traité), le risque relatif sur les fractures est assuré d'être inférieur à 1, donc en faveur d'une réduction de la fréquence des fractures (avec un degré de certitude de 95%).

Ce seuil est déterminé à l'aide de l'intervalle de prédiction de la droite de régression. Il correspond à la valeur de x pour laquelle la borne péjorative de **l'intervalle de prédiction** passe en dessous du trait horizontal correspond à l'absence d'effet traitement sur le critère clinique. C'est donc le plus petit d'effet traitement sur le critère de substitution qui commence à se traduire par un effet non nul sur le critère clinique.

Lorsque le critère de substitution est validé, il pourra être conclu qu'un nouveau traitement apporte un bénéfice sur le critère clinique qu'à la condition qu'il démontre un effet sur ce surrogate au moins aussi important en taille que le seuil précédemment identifié pour ce surrogate. Pour cela il faudra que la borne supérieure de l'intervalle de confiance à 95% de l'effet sur le surrogate soit inférieure à ce seuil (ou qu'un calcul adapté soit effectué pour prendre en compte les incertitudes statistiques).

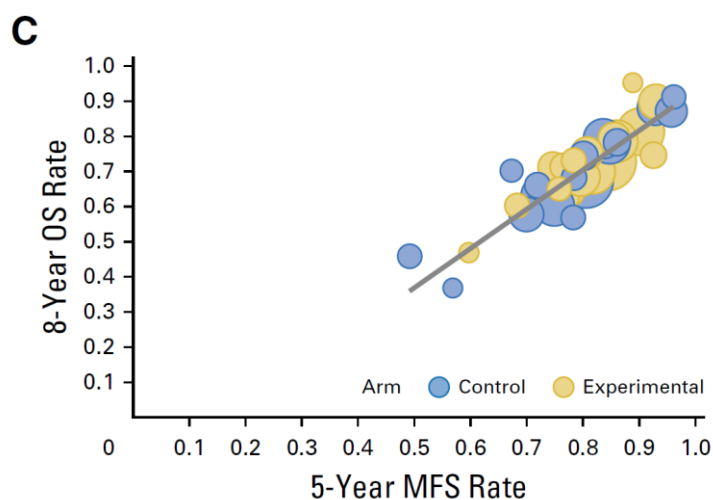
Le concept de critère de substitution est souvent confondu par erreur avec d'autres concepts. Par exemple avec l'association entre 2 variables en termes de risques ou de pronostic. De même, la mise en évidence dans une cohorte de patients, tous traités avec le traitement d'intérêt, d'une association entre un critère intermédiaire et le critère clinique ne permet pas de valider un critère de substitution. Cette association montre seulement une valeur pronostique sous traitement du critère intermédiaire (qui peut exister indépendamment du traitement). Comme tous les patients sont traités, il est impossible d'évaluer l'association entre l'effet du traitement sur le critère intermédiaire et sur le critère clinique. Le changement avant-après du critère intermédiaire ne peut pas être utilisé non plus comme effet du traitement, car sujet à un phénomène de confusion lié entre autres à l'évolution naturelle de la maladie et à la régression à la moyenne.

L'obtention d'une démonstration rigoureuse qu'un critère intermédiaire est un réel critère de substitution est indispensable, car il existe de nombreux exemples de traitements efficaces sur un potentiel surrogate, mais pour lesquels aucun bénéfice clinique n'a été mis en évidence dans les essais randomisés sur les critères cliniques. Par exemple, en prévention cardiovasculaire, le LDL cholestérol ne peut pas être considéré comme un surrogate universel, car l'acide nicotinique, qui baisse le taux de LDL cholestérol, a échoué à mettre en évidence une réduction des événements cardiovasculaires dans un essai incluant 25673 patients suivis 3.9 ans. De nombreux autres exemples de ce type existent dans tous les champs de la médecine [24, 25].

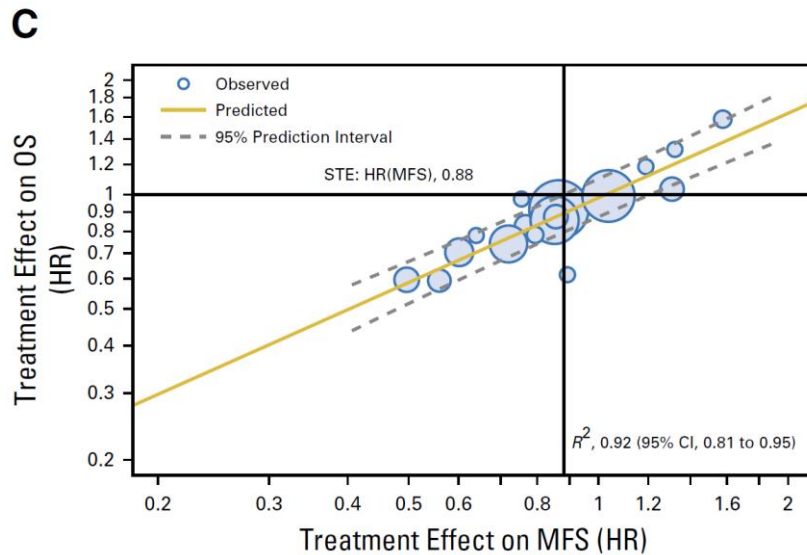
Exemple de démonstration d'un surrogate

La survie sans métastase (MFS) a été validée comme surrogate de de la survie globale (OS) dans le cancer de la prostate localisé [26]. Cette validation a été obtenue à partir des données de 28 essais randomisés regroupant plus de 28,000 sujets. Quatre candidats surrogates ont été explorés : disease free survival (DFS), time to disease recurrence (TDR), time to metastasis (TTM) et metastasis free survival (MFS).

Une corrélation entre la MFS et l'OS a été observée faisant que ce critère remplit cette première condition (individual level correlation). Cette corrélation est documentée par le graphique ci-dessous où l'abscisse représente le taux de survie sans métastase mesurée à 5 ans (estimée par la courbe de Kaplan Meier de la MFS) et l'ordonnée le taux de survie à 8 ans (estimée avec la courbe de Kaplan Meier de l'OS). Chaque bras des essais (code couleur) documente ces 2 valeurs (le nombre de points est le double de celui d'essai)



Une très bonne corrélation des effets est aussi observée. Le seuil est à 0.88 (abscisse du croisement de la borne supérieure de l'intervalle de prédiction de la droite de régression et le trait horizontal correspondant à un HR de 1 pour l'OS (absence d'effet sur l'OS)). Cette corrélation est documentée par le graphique ci-dessous où l'abscisse représente le hazard ratio obtenu sur la MFS et l'ordonnée celui correspondant à l'OS. Chaque point représente un essai (les hazard ratio quantifiant la différence entre groupe traité et groupe contrôle pour chaque essai).



Les résultats statistiques permettant de faire cette démonstration sont présentés dans le tableau suivant qui donne un bon exemple des résultats qui sont attendus dans ce type travaux :

Table 1. Two-Condition Surrogacy Analysis

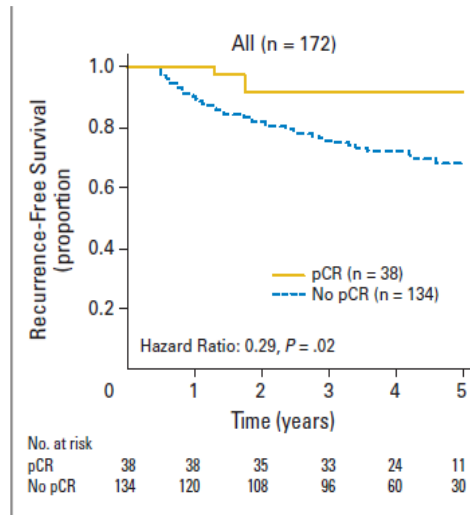
TE	ICE	No. of Trials	No. of Units*	No. of Patients	Condition 1 (TE and ICE are correlated)		Condition 2 (treatment effects on both end points are correlated)	
					Correlation at the Patient Level, Kendall's τ (95% CI)	Regression of 8-Year TE Rate v 5-Year ICE Rate [†] by Trial and Arm, R^2 (95% CI)	R^2 (95% CI)	Regression Equation
OS	DFS	24	31	21,140	0.85 (0.85 to 0.86)	0.86 (0.78 to 0.90)	0.73 (0.53 to 0.82)	$\text{Log(HR)}_{\text{OS}} = 0.035 + 0.605 \times \text{Log(HR)}_{\text{DFS}}$
DSS	TDR	21 [‡]	28	20,496 [‡]	0.68 (0.67 to 0.69)	0.80 (0.70 to 0.85)	0.63 (0.36 to 0.75)	$\text{Log(HR)}_{\text{DSS}} = 0.027 + 0.809 \times \text{Log(HR)}_{\text{TDR}}$
OS	MFS	19	21	12,712	0.91 (0.91 to 0.91)	0.83 (0.71 to 0.88)	0.92 (0.81 to 0.95)	$\text{Log(HR)}_{\text{OS}} = -0.021 + 0.740 \times \text{Log(HR)}_{\text{MFS}}$
DSS	TTM	16 [‡]	18	12,068 [‡]	0.91 (0.91 to 0.92)	0.86 (0.75 to 0.90)	0.89 (0.72 to 0.93)	$\text{Log(HR)}_{\text{DSS}} = -0.072 + 0.880 \times \text{Log(HR)}_{\text{TTM}}$

Abbreviations: DFS, disease-free survival; DSS, disease-specific survival; HR, hazard ratio; ICE, intermediate clinical end point; MFS, metastasis-free survival; OS, overall survival; TDR, time to disease recurrence; TE, true end point; TTM, time to metastasis.
^{*}Five trials were split according to the type of primary therapy or experimental arm (if two or more experimental arms).
[†]Eight-year TE rates and 5-year ICE rates were Kaplan-Meier estimates by trial and treatment arm, excluding three studies with median follow-up < 6 years.
[‡]Excluding three studies with the number of prostate cancer deaths < 3.

La corrélation « individual level » (condition 1) est forte avec un R^2 de 0.83. La corrélation des effets « trial-level » (condition2) est aussi forte avec un R^2 de 0.92 (borne inférieure de l'intervalle de confiance à 0.81).

Exemple d'un échec de démonstration de la surrogacy

La réponse pathologique complète (pCR) est un facteur de risque de récurrence dans le cancer du sein précoce.



Cependant il n'a pas été possible de démontrer qu'il s'agissait d'un surrogat dans le cadre du traitement néoadjuvant [63]. L'observation qu'un traitement impacte favorablement la pCR ne permet pas d'induire avec certitude qu'il impacte favorablement la récurrence. Ultérieurement, une étude [27] utilisant la bonne méthode (graphique ci-dessous) de surrogacy entre pCR et survie.

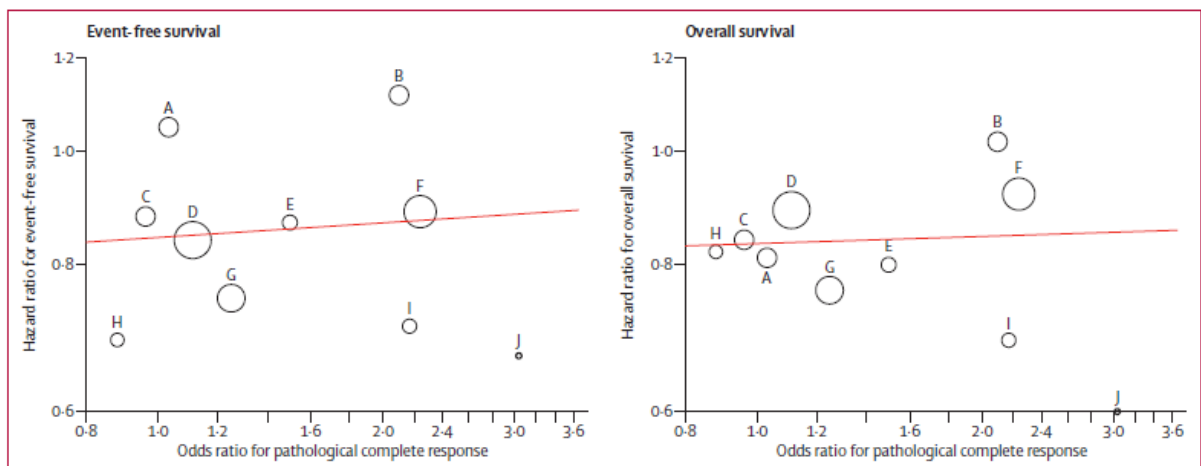


Figure 6: Trial-level correlation between treatment effect on pathological complete response and event-free survival or overall survival
 Each circle corresponds to one randomised comparison and the size of the circle represents the sample size. A=GeparQuattro (epirubicin plus cyclophosphamide followed by docetaxel then capecitabine vs epirubicin plus cyclophosphamide followed by docetaxel). B=GeparDuo. C=GeparQuattro (epirubicin plus cyclophosphamide followed by docetaxel and capecitabine vs epirubicin plus cyclophosphamide followed by docetaxel). D=EORTC 10994/BIG 1-00. E=PREPARE. F=NSABP B-27. G=responders in GeparTrio. H=non-responders in GeparTrio. I=AGO 1. J=NOAH.

3.2 Autres approches statistiques

D'autres approches statistiques ont été proposées, mais ne sont pas considérées comme des approches standards pour l'instant [28, 29, 30, 31]. L'utilisation de ces nouvelles méthodes est à surveiller.

4 Limites des études de validation de surrogate [new]

4.1 Limites liées à la nature rétrospective des études, HARKing, p hacking

La démonstration d'une très forte corrélation des effets n'est pas suffisante pour faire accepter un surrogate. Ces travaux d'évaluation de la surrogacy sont, par définition, rétrospectifs, et sont donc exposés à plusieurs problématiques dans la méthodologie de leur réalisation et peuvent avoir de nombreuses limites.

Les études de validation des surrogates sont avant tout des revues systématiques et des méta-analyses. Elles héritent des problématiques de ces techniques liées au caractère rétrospectif du travail induisant la possibilité de p hacking [32, 33] et de HARKing [34].

Les résultats de l'analyse de corrélation dépendent directement des études retenues pour effectuer cette recherche. Il est facile, du fait du caractère rétrospectif, de tester toutes les combinaisons possibles d'études et d'écrire ensuite le protocole et les critères de sélection des études. Compte tenu de la complexité de la littérature, de la multiplicité des études, ces manipulations de données sont difficiles à détecter. La fréquence de ces pratiques fait que l'on estime que seuls 20% des revues systématiques et méta-analyses publiées (quel que soit leur objectif, pas seulement celles destinées à évaluer un candidat surrogate) sont correctes méthodologiquement [35][36].

Il existe aussi une explosion combinatoire dans la façon dont peuvent être conduites les analyses de surrogacy en raison de la multiplicité des choix d'analyse possibles :

- Choix du surrogate lui-même, c'est-à-dire de la variable explicative de la corrélation, avec souvent la possibilité de choisir entre plusieurs biomarqueurs (par exemple le cholestérol total, le LDL cholestérol ou le non HDL cholestérol pour la prévention cardiovasculaire ; la réponse objective, la réponse complète, la survie sans progression, la survie sans progression modifiée, le temps à progression en oncologie).
- Choix dans le critère clinique et ses éventuelles multiples combinaisons dans des critères composites (par exemple décès de toutes causes, décès coronariens, décès cardiovasculaires, événements coronariens mortels et non mortels en cardiologie).
- Choix dans les métriques de ces deux variables : différence absolue ou différence relative par rapport à la baseline, pente d'un modèle pour mesure multiple pour les critères intermédiaires continus (comme le LDL) ; risque ratio, odds ratio, hazard ratio, différence des risques pour les événements cliniques. Pour les effets traitement relatifs sur les critères binaires se pose aussi le choix de la transformation logarithmique.
- Choix des horizons temporels de mesure des effets sur les deux variables : par exemple effet sur le LDL à 1 mois après la randomisation, 6 mois, 1 an ou moyenne sur la durée totale de l'essai ; effet sur le critère clinique à la fin de l'étude ou standardisé sur un suivi fixe et dans ce cas quel suivi (1 an, 3 ans, 5 ans, etc.).
- Choix du niveau de corrélation suffisant pour conclure à la surrogacy : 0.8, 0.90, etc., en termes de r ou r^2 , en prenant l'estimation ponctuelle ou la borne inférieure de l'intervalle de confiance, etc.
- Méthode statistique elle-même (moindre carré simple, REML, etc.), sans pondération ou avec pondération, dans ce dernier cas, par la taille des études ou l'inverse de la variance de l'un des effets traitements (candidat surrogate ou critère final).

- Gestion des études avec des bras sans événements (zero event) : exclusion, correction de la solution de continuité
- Choix des études avec de nombreuses possibilités suivant les définitions retenues pour
 - les populations (prévention primaire ou secondaire pour la prévention cardiovasculaire ciblant le cholestérol par exemple),
 - les molécules (toutes les molécules ayant fait l'objet de RCT, que les molécules ayant montré un bénéfice clinique)
 - la méthodologie des essais : double aveugle ou ouvert, analyse en ITT ou en per protocole (peut être plus logique dans ce cadre d'analyse à vocation plus cognitive que décisionnelle), etc.

Ainsi, du fait de la nature rétrospective, rien n'empêche que de très nombreuses analyses soient réalisées avant d'obtenir l'analyse qui produit les résultats escomptés et qui sera finalement la seule présentée comme validation du surrogate.

En se basant sur ce qui a été observé de manière générale pour toutes les études rétrospectives (et en particulier les méta-analyses), il est raisonnable de conjecturer qu'il est possible d'obtenir dans ces travaux tous les résultats possibles en faveur ou en défaveur de la surrogacy. Ce point est d'ores et déjà illustré lorsque plusieurs travaux cherchant à établir la même surrogacy sont disponibles et donnent des résultats et des conclusions différents comme c'est le cas par exemple avec le cholestérol ou ses sous-fractions. [37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49].

Pour limiter un peu la flexibilité des analyses, des guidelines statistiques [50, 51] ont été élaborés et pré-séparent un certain nombre de principes d'analyse comme l'utilisation d'effet traitement relatif, l'utilisation de l'odds ratio avec les critères binaires (pour ses propriétés de symétrie et de domaine de définition), avec une transformation logarithmique pour retrouver les hypothèses de la régression (variables variant de $-\infty$ à $+\infty$, hétéroscédasticité). De même, le modèle linéaire s'impose en l'absence de justification biologique forte pour une autre forme de modèle, en raison du principe de parcimonie du rasoir d'Ockham.

Pour éviter ces travers et assurer une certaine crédibilité aux résultats produits, ces travaux doivent être réalisés suivant un protocole et un plan d'analyse statistique fixant a priori les choix possibles décrits ci-dessus et établis avant toute analyse exploratoire des données, enregistrés et accompagnés dans le rapport ou la publication d'une attestation que les résultats présentés sont conformes à l'analyse préétablie (ou une description des changements avec leur justification).

En particulier la recherche des essais participants à la recherche de la corrélation des effets doit être impérativement réalisée par une réelle revue systématique afin d'éviter toute sélection arbitraire des essais afin d'obtenir le résultat souhaité. La méthode de cette revue systématique doit être fixée au protocole et conforme à l'état de l'art (Cochrane Handbook par exemple). En particulier les critères de sélection des essais doivent être clairement énoncés à priori et justifiés. Une description des essais exclus avec la raison d'exclusion permettra de vérifier la rigueur de l'application de ces critères.

L'UE, dans le cadre du Joint Clinical Assessment (JCA), a précisé ses attentes⁴ sur la validation des surrogates [52].

⁴ HTA CG. Guidance on outcomes for joint clinical assessments. Localisation https://health.ec.europa.eu/publications/guidance-outcomes-joint-clinical-assessments_en

4.2 Le biais écologique [new]

La réalisation de ces travaux de validation de la surrogacy sur des données agrégées expose aussi à au « biais écologique » si les études sont différentes sur un facteur conditionnant le bénéfice clinique [53, 54, 55].

Cette situation apparaît par exemple si le bénéfice passe par un autre mécanisme d'action que l'effet sur le critère intermédiaire pour certaines molécules (ou classes). Une relation entre les effets pourra apparaître alors même si le critère intermédiaire n'est absolument pas un surrogat.

Imaginons un critère intermédiaire qui n'a aucune valeur de surrogacy. Quel que soit l'effet du traitement sur ce critère, son bénéfice clinique est identique (ce qui correspond à une droite de régression horizontale). Les essais ont été réalisés avec deux catégories de molécules : celle qui n'agit que sur le critère intermédiaire et qui est de ce fait dépourvue de bénéfice clinique, et d'autres qui en plus de leur action sur le critère intermédiaire apportent un bénéfice clinique direct par un tout autre mécanisme d'action (non médié par l'effet sur le critère intermédiaire). L'existence de ces 2 sous populations d'essais, se distinguant par l'existence ou non d'un effet direct sur le critère clinique, fera apparaître une fausse relation entre l'effet sur le critère intermédiaire et le bénéfice clinique (Figure 3). Seul un recours aux données individuelles avec la recherche de la relation au sein de chaque essai permettra de conclure correctement [55].

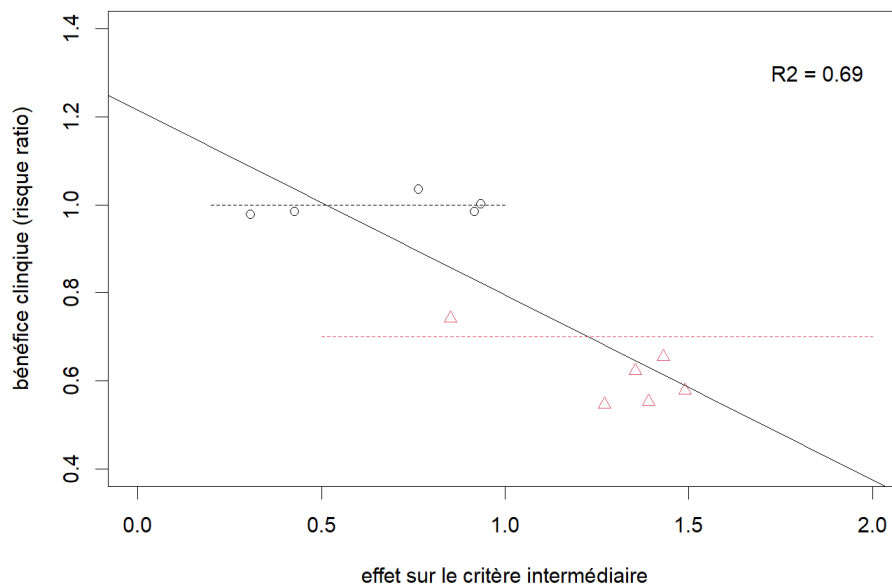


Figure 3 – Biais écologique dans la validation des surrogates.

Le critère intermédiaire n'est pas un surrogat du bénéfice clinique. Mais parmi toutes les molécules testées dans des essais (les points du graphique), il existe des molécules qui apportent un bénéfice par un autre mécanisme d'action (triangles rouges) et des molécules sans aucun bénéfice clinique (ronds noirs). Au sein de chaque catégorie de molécule, il n'existe aucune relation entre le critère intermédiaire et le bénéfice clinique (ligne pointillée horizontale). Mais une corrélation apparaît si l'on considère tous les points sans distinction car, concomitamment mais de façon complètement indépendante, les molécules apportant un bénéfice clinique ont un effet plus important sur le biomarqueur que les molécules n'apportant pas de bénéfice clinique.

5 Extrapolabilité/transposabilité

La valeur de surrogacy est le plus souvent établie à partir d'essais de molécules de la même classe, ayant le même mécanisme d'action. La question qui se pose alors est celle de l'extrapolabilité de cette relation à une nouvelle classe, à un nouveau mécanisme d'action. En effet, le but du surrogate est d'éviter, pour les nouvelles molécules ayant un nouveau mécanisme d'action, de devoir faire des essais de morbi-mortalité long et demandant des effectifs importants. La question de l'universalité de la relation de surrogacy, quels que soient les mécanismes d'action, les profils de sécurité des molécules devient donc primordiale. Or de nombreux arguments théoriques et des exemples passés réfutent cette possibilité (cf. section 2).

En complément d'une validation statistique satisfaisante, sans biais, montrant une forte corrélation des effets sur le critère intermédiaire avec ceux sur le critère clinique, il est impératif de justifier la transposabilité de la relation de surrogacy identifiée. Pour cela il faudra justifier que :

- La nouvelle molécule (ou le nouveau mécanisme d'action) n'entraîne pas d'effet off-target connu ou inconnu (aie !)
- Le profil de sécurité de la nouvelle molécule (ou du nouveau mécanisme d'action) n'est pas susceptible de contrebalancer l'éventuel bénéfice sur le même critère clinique (comme avec la mortalité totale qui intègre bénéfices et effets délétères graves).
- La relation de surrogacy établie n'est pas due (en partie ou en totalité) à une autre action des précédentes molécules que leur seule action sur le critère intermédiaire (cf. exemple des statines et de leur effet pléiotrope possiblement médié par une action antiinflammatoire [56])

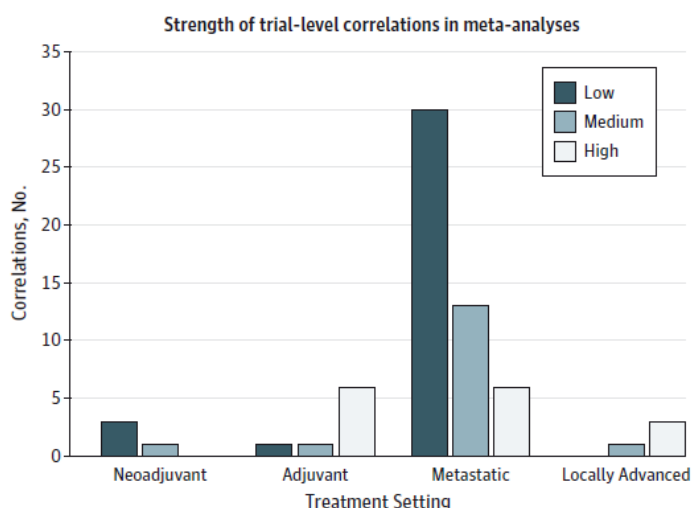
La question de la transposabilité se pose aussi lorsqu'une éventuelle relation de surrogacy est utilisée pour justifier le traitement d'une autre population que celles étudiées dans les essais précédents. Par exemple pour justifier l'utilisation de statine chez des sujets réellement âgés en se basant sur la corrélation des effets obtenus dans les essais des statines qui ont inclus des sujets plutôt jeunes. Dans cet exemple rien ne garantit que l'on retrouvera la réduction de mortalité totale par exemple, tout simplement en raison des risques compétitifs du risque de décès de cause cardiovasculaire, risques compétitifs très forts chez la personne âgée.

6 Méta-épidémiologie, situation actuelle

Depuis une dizaine d'années, les nouveaux traitements, particulièrement en oncologie, sont fréquemment évalués sur des « surrogates » qui n'ont pas fait l'objet de validation, à la suite de la possibilité offerte par la FDA⁵. De ce fait un doute persiste souvent sur le réel intérêt de ces traitements. Plusieurs auteurs se sont émus de cette situation [57, 58, 59, 60, 61, 62, 63, 64] qui a été bien étudié par diverses études de méta-épidémiologie.

En oncologie le critère clinique est la survie (ou la qualité de vie en montrant que le nouveau traitement s'accompagne d'une meilleure qualité de vie avec une survie identique au traitement standard). La PFS, plus fréquente et plus facilement impactée que l'OS, est souvent avancée comme étant un surrogate de l'OS. Les travaux étudiant la corrélation des effets sur la PFS avec ceux sur l'OS dans différent type de cancer et/ou différents types de traitements ont été colligés par Vinay Prasad et ses collègues en 2015 [65]. En majorité, ces travaux montrent une corrélation faible ou moyenne ne permettant pas de considérer la PFS comme surrogate de l'OS dans les situations correspondantes.

Figure 3. Correlations by Treatment Setting



We scored strength of trial-level correlation according to a modification to surrogate criteria proposed by the Institute of Quality and Efficiency in Health Care³⁴: low correlation ($r \leq 0.7$), medium strength correlation ($r > 0.7$ to $r < 0.85$), and high correlation ($r \geq 0.85$).

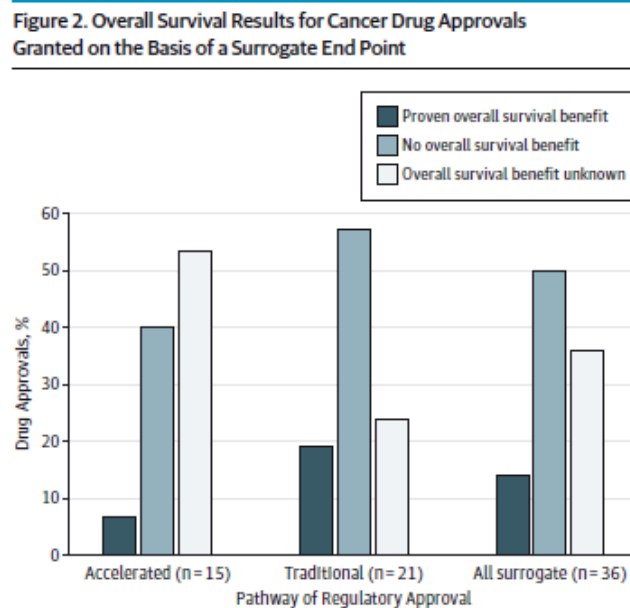
Les situations où la PFS peut être considéré comme un surrogate fiable de l'OS sont très peu nombreux et peuvent éventuellement être remis en cause quand des essais supplémentaires deviennent disponibles. Ce fût le cas avec les traitements du cancer du côlon métastatique pour lesquels la PFS

⁵ Sec. 314.510 approval based on a surrogate endpoint or on an effect on a clinical endpoint other than survival or irreversible morbidity (FDA TITLE 21, PART 314, SUBPART H)

“FDA may grant marketing approval for a new drug product on the basis of adequate and well-controlled clinical trials establishing that the drug product has an effect on a surrogate endpoint that is reasonably likely ... to predict clinical benefit...”

semblait être dans un 1^{er} temps un surrogate acceptable [66] mais un travail ultérieur a récusé ce résultat [67].

La conséquence directe de cela est que, 5 ans après leur enregistrement, des preuves du bénéfice sur l'OS ne sont disponibles que pour une faible proportion des nouveaux produits enregistrés sur des essais portant sur ces critères présentés comme étant des « surrogate » [68] ; et que pour un nombre non négligeable de produits, les essais de mortalité réalisés ne montrent pas de bénéfice sur l'OS :



D'autres études apportent des éclairages complémentaires et confirment la faiblesse des exigences dans l'enregistrement des nouveaux traitements en oncologie [61, 69, 70, 71].

Une scoping review publiée en 2023 liste les différentes définitions, critères d'acceptabilité, limitations mentionnées dans la littérature sur l'utilisation de surrogates dans les essais [72].

7 Conclusion

L'idée du critère de substitution est séduisante, mais il se pourrait qu'elle ne soit qu'une vaine chimère. En effet, une analyse approfondie de ce concept fait immédiatement apparaître plusieurs limites qui rendent l'existence de véritables critères de substitution assez improbable car très difficile à lever.

Références

- 1 Shyr Y, Shyr D. What Constitutes a Valid Surrogate End Point in Cancer Clinical Trials? *JAMA Oncol* 2020 doi:10.1001/jamaoncol.2020.1847;
- 2 Manyara AM, Davies P, Stewart D, et al. Reporting of surrogate endpoints in randomised controlled trial protocols (SPIRIT-Surrogate): extension checklist with explanation and elaboration. *BMJ* 2024;386:e078525 doi:10.1136/bmj-2023-078525; PMID:38981624;
- 3 Manyara AM, Davies P, Stewart D, et al. Reporting of surrogate endpoints in randomised controlled trial reports (CONSORT-Surrogate): extension checklist with explanation and elaboration. *BMJ* 2024;386:e078524 doi:10.1136/bmj-2023-078524; PMID:38981645;
- 4 Group, ICECaP Working, Sweeney C, Nakabayashi M, et al. The Development of Intermediate Clinical Endpoints in Cancer of the Prostate (ICECaP). *JNCI Journal of the National Cancer Institute* 2015;107:djv261 doi:10.1093/jnci/djv261; PMID:26409187;
- 5 Gordon T, Castelli WP, Hjortland MC, et al. High density lipoprotein as a protective factor against coronary heart disease. The Framingham Study. *Am J Med* 1977;62:707–14 doi:10.1016/0002-9343(77)90874-9; PMID:193398;
- 6 Gordon DJ, Probstfield JL, Garrison RJ, et al. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation* 1989;79:8–15 doi:10.1161/01.cir.79.1.8; PMID:2642759;
- 7 LaRosa JC, Hunninghake D, Bush D, et al. The cholesterol facts. A summary of the evidence relating dietary fats, serum cholesterol, and coronary heart disease. A joint statement by the American Heart Association and the National Heart, Lung, and Blood Institute. The Task Force on Cholesterol Issues, American Heart Association. *Circulation* 1990;81:1721–33 doi:10.1161/01.cir.81.5.1721; PMID:2184951;
- 8 Landray MJ, Haynes R, Hopewell JC, et al. Effects of extended-release niacin with laropirant in high-risk patients. *N Engl J Med* 2014;371:203–12 doi:10.1056/NEJMoa1300955; PMID:25014686;
- 9 Riggs BL, Hodgson SF, O'Fallon WM, et al. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *N Engl J Med* 1990;322:802–09 doi:10.1056/NEJM199003223221203; PMID:2407957;
- 10 Barbarawi M, Kheiri B, Zayed Y, et al. Vitamin D Supplementation and Cardiovascular Disease Risks in More Than 83 000 Individuals in 21 Randomized Clinical Trials: A Meta-analysis. *JAMA Cardiol* 2019;4:765–76 doi:10.1001/jamacardio.2019.1870; PMID:31215980;
- 11 LeBoff MS, Chou SH, Ratliff KA, et al. Supplemental Vitamin D and Incident Fractures in Midlife and Older Adults. *N Engl J Med* 2022;387:299–309 doi:10.1056/NEJMoa2202106; PMID:35939577;
- 12 Keech A, Simes RJ, Barter P, et al. Effects of long-term fenofibrate therapy on cardiovascular events in 9795 people with type 2 diabetes mellitus (the FIELD study): randomised controlled trial. *The Lancet* 2005;366:1849–61 doi:10.1016/S0140-6736(05)67667-2; PMID:16310551;
- 13 Das Pradhan A, Glynn RJ, Fruchart J-C, et al. Triglyceride Lowering with Pemafibrate to Reduce Cardiovascular Risk. *N Engl J Med* 2022;387:1923–34 doi:10.1056/NEJMoa2210645; PMID:36342113;
- 14 Kristeleit R, Lisyanskaya A, Fedenko A, et al. Rucaparib versus standard-of-care chemotherapy in patients with relapsed ovarian cancer and a deleterious BRCA1 or BRCA2 mutation (ARIEL4): an international, open-label, randomised, phase 3 trial. *The Lancet Oncology* 2022;23:465–78 doi:10.1016/S1470-2045(22)00122-X; PMID:35298906;
- 15 Weintraub WS, Lüscher TF, Pocock S. The perils of surrogate endpoints. *Eur. Heart J.* 2015;36:2212–18 doi:10.1093/eurheartj/ehv164; PMID:25975658;

- 16 Hwang TJ, Carpenter D, Lauffenburger JC, et al. Failure of Investigational Drugs in Late-Stage Clinical Development and Publication of Trial Results. *JAMA Intern Med* 2016;176:1826–33 doi:10.1001/jamainternmed.2016.6008; PMID:27723879;
- 17 Amsallem E, Kasparian C, Haddour G, et al. Phosphodiesterase III inhibitors for heart failure. *Cochrane Database Syst Rev* 2005;2013:CD002230 doi:10.1002/14651858.CD002230.pub2; PMID:15674893;
- 18 Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006;5:173–86 doi:10.1002/pst.207; PMID:17080751;
- 19 Buyse M, Molenberghs G, Paoletti X, et al. Statistical evaluation of surrogate endpoints with examples from cancer clinical trials: Statistical evaluation of surrogate endpoints. *Biometrical Journal* 2016;58:104–32 doi:10.1002/bimj.201400049;
- 20 Ciani O, Buyse M, Drummond M, et al. Use of surrogate end points in healthcare policy: a proposal for adoption of a validation framework. *Nature Reviews Drug Discovery* 2016;15:516 doi:10.1038/nrd.2016.81;
- 21 Bujkiewicz S, Thompson JR, Spata E, et al. Uncertainty in the Bayesian meta-analysis of normally distributed surrogate endpoints. *Stat Methods Med Res* 2017;26:2287–318 doi:10.1177/0962280215597260;
- 22 Ciani O, Davis S, Tappenden P, et al. VALIDATION OF SURROGATE ENDPOINTS IN ADVANCED SOLID TUMORS: SYSTEMATIC REVIEW OF STATISTICAL METHODS, RESULTS, AND IMPLICATIONS FOR POLICY MAKERS. *International Journal of Technology Assessment in Health Care* 2014;30:312–24 doi:10.1017/S0266462314000300;
- 23 IQWiG. Validity of surrogate endpoints in oncology. Cologne 2011.
- 24 Norsworthy KJ, Gao X, Ko C-W, et al. Response Rate, Event-Free Survival, and Overall Survival in Newly Diagnosed Acute Myeloid Leukemia: US Food and Drug Administration Trial-Level and Patient-Level Analyses. *JCO* 2022;40:847–54 doi:10.1200/JCO.21.01548; PMID:34890212;
- 25 Lassere MN, Johnson KR, Schiff M, et al. Is blood pressure reduction a valid surrogate endpoint for stroke prevention? An analysis incorporating a systematic review of randomised controlled trials, a by-trial weighted errors-in-variables regression, the surrogate threshold effect (STE) and the Biomarker-Surrogacy (BioSurrogate) Evaluation Schema (BSES). *BMC Med Res Methodol* 2012;12:27 doi:10.1186/1471-2288-12-27; PMID:22409774;
- 26 Xie W, Regan MM, Buyse M, et al. Metastasis-Free Survival Is a Strong Surrogate of Overall Survival in Localized Prostate Cancer. *Journal of Clinical Oncology* 2017;JCO.2017.73.9987 doi:10.1200/JCO.2017.73.9987;
- 27 Cortazar P, Zhang L, Untch M, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *The Lancet* 2014;384:164–72 doi:10.1016/S0140-6736(13)62422-8; PMID:24529560;
- 28 Papanikos T, Thompson JR, Abrams KR, et al. A novel approach to bivariate meta-analysis of binary outcomes and its application in the context of surrogate endpoints 2020.
- 29 Branchoux S, Sofeu CL, Gaudin A-F, et al. Time to next treatment or death as a candidate surrogate endpoint for overall survival in advanced melanoma patients treated with immune checkpoint inhibitors: an insight from the phase III CheckMate 067 trial. *ESMO Open* 2021;7:100340 doi:10.1016/j.esmoop.2021.100340; PMID:34929616;
- 30 Baker SG. Five criteria for using a surrogate endpoint to predict treatment effect based on data from multiple previous trials. *Stat Med* 2018;37:507–18 doi:10.1002/sim.7561; PMID:29164641;
- 31 Sofeu CL, Emura T, Rondeau V. One-step validation method for surrogate endpoints using data from multiple randomized cancer clinical trials with failure-time endpoints. *Stat Med* 2019;38:2928–42 doi:10.1002/sim.8162; PMID:30997685;

- 32 Head ML, Holman L, Lanfear R, et al. The extent and consequences of p-hacking in science. *PLoS Biology* 2015;13:e1002106 doi:10.1371/journal.pbio.1002106; PMID:25768323;
- 33 Data dredging - Wikipedia 2021. Available at: https://en.wikipedia.org/wiki/Data_dredging Accessed August 30, 2021.
- 34 Kerr NL. HARKing: hypothesizing after the results are known. *Pers Soc Psychol Rev* 1998;2:196–217 doi:10.1207/s15327957pspr0203_4; PMID:15647155;
- 35 Siontis KC, Hernandez-Boussard T, Ioannidis JPA. Overlapping meta-analyses on the same topic: survey of published studies. *BMJ* 2013;347:f4501 doi:10.1136/bmj.f4501; PMID:23873947;
- 36 Ioannidis JPA. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly* 2016;94:485–514 doi:10.1111/1468-0009.12210; PMID:27620683;
- 37 Byrne P, Demasi M, Jones M, et al. Evaluating the Association Between Low-Density Lipoprotein Cholesterol Reduction and Relative and Absolute Effects of Statin Treatment: A Systematic Review and Meta-analysis. *JAMA Intern Med* 2022;182:474–81 doi:10.1001/jamainternmed.2022.0134; PMID:35285850;
- 38 Baigent C, Keech A, Kearney PM, et al. Efficacy and safety of cholesterol-lowering treatment: prospective meta-analysis of data from 90 056 participants in 14 randomised trials of statins. *The Lancet* 2005;366:1267–78 doi:10.1016/S0140-6736(05)67394-1; PMID:16214597;
- 39 Silverman MG, Ference BA, Im K, et al. Association Between Lowering LDL-C and Cardiovascular Risk Reduction Among Different Therapeutic Interventions: A Systematic Review and Meta-analysis. *JAMA* 2016;316:1289–97 doi:10.1001/jama.2016.13985; PMID:27673306;
- 40 Koskinas KC, Siontis GCM, Piccolo R, et al. Effect of statins and non-statin LDL-lowering medications on cardiovascular outcomes in secondary prevention: a meta-analysis of randomized trials. *Eur. Heart J.* 2018;39:1172–80 doi:10.1093/eurheartj/ehx566; PMID:29069377;
- 41 Holme I. An analysis of randomized trials evaluating the effect of cholesterol reduction on total mortality and coronary heart disease incidence. *Circulation* 1990;82:1916–24 doi:10.1161/01.cir.82.6.1916; PMID:2242517;
- 42 Gould AL, Rossouw JE, Santanello NC, et al. Cholesterol reduction yields clinical benefit. A new look at old data. *Circulation* 1995;91:2274–82 doi:10.1161/01.cir.91.8.2274; PMID:7697857;
- 43 Holme I. Relation of coronary heart disease incidence and total mortality to plasma cholesterol reduction in randomised trials: use of meta-analysis. *Br Heart J* 1993;69:S42-7 doi:10.1136/hrt.69.1_suppl.s42; PMID:8427764;
- 44 Holme I. Cholesterol reduction and its impact on coronary artery disease and total mortality. *Am J Cardiol* 1995;76:10C-17C doi:10.1016/s0002-9149(99)80465-2; PMID:7572677;
- 45 Simes RJ, Marschner IC, Hunt D, et al. Relationship between lipid levels and clinical outcomes in the Long-term Intervention with Pravastatin in Ischemic Disease (LIPID) Trial: to what extent is the reduction in coronary events with pravastatin explained by on-study lipid levels? *Circulation* 2002;105:1162–69 doi:10.1161/hc1002.105136; PMID:11889008;
- 46 Gould AL, Davies GM, Alemao E, et al. Cholesterol reduction yields clinical benefits: meta-analysis including recent trials. *Clin Ther* 2007;29:778–94 doi:10.1016/j.clinthera.2007.05.012; PMID:17697899;
- 47 Holme I. Relationship between total mortality and cholesterol reduction as found by meta-regression analysis of randomized cholesterol-lowering trials. *Controlled Clinical Trials* 1996;17:13–22 doi:10.1016/0197-2456(95)00083-6; PMID:8721798;
- 48 Johnson KR, Freemantle N, Anthony DM, et al. LDL-cholesterol differences predicted survival benefit in statin trials by the surrogate threshold effect (STE). *Journal of Clinical Epidemiology* 2009;62:328–36 doi:10.1016/j.jclinepi.2008.06.004; PMID:18834708;

- 49 Gould AL, Rossouw JE, Santanello NC, et al. Cholesterol reduction yields clinical benefit: impact of statin trials. *Circulation* 1998;97:946–52 doi:10.1161/01.cir.97.10.946; PMID:9529261;
- 50 van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med* 2002;21:589–624 doi:10.1002/sim.1040; PMID:11836738;
- 51 Thompson SG, Higgins JPT. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559–73 doi:10.1002/sim.1187; PMID:12111920;
- 52 HTA CG. Guidance on outcomes for joint clinical assessments.
- 53 Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 1982;72:1336–44 doi:10.2105/ajph.72.12.1336; PMID:7137430;
- 54 Geissbühler M, Hincapié CA, Aghlmandi S, et al. Most published meta-regression analyses based on aggregate data suffer from methodological pitfalls: a meta-epidemiological study. *BMC Med Res Methodol* 2021;21:123 doi:10.1186/s12874-021-01310-0; PMID:34130658;
- 55 Berlin JA, Santanna J, Schmid CH, et al. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med* 2002;21:371–87 doi:10.1002/sim.1023; PMID:11813224;
- 56 Ridker PM, Danielson E, Fonseca FA, et al. Reduction in C-reactive protein and LDL cholesterol and cardiovascular event rates after initiation of rosuvastatin: a prospective study of the JUPITER trial. *The Lancet* 2009;373:1175–82 doi:10.1016/S0140-6736(09)60447-5; PMID:19329177;
- 57 Naci H, Davis C. Inappropriate use of progression-free survival in cancer drug approvals. *BMJ-BRITISH MEDICAL JOURNAL* 2020;368:m770 doi:10.1136/bmj.m770; PMID:32156802;
- 58 Kemp R, Prasad V. Surrogate endpoints in oncology: when are they acceptable for regulatory and clinical decisions, and are they currently overused? *BMC Med* 2017;15:134 doi:10.1186/s12916-017-0902-9; PMID:28728605;
- 59 Hilal T, Gonzalez-Velez M, Prasad V. Limitations in Clinical Trials Leading to Anticancer Drug Approvals by the US Food and Drug Administration. *JAMA Internal Medicine* 2020;180:1108–15 doi:10.1001/jamainternmed.2020.2250; PMID:32539071;
- 60 Chen EY, Joshi SK, Tran A, et al. Estimation of Study Time Reduction Using Surrogate End Points Rather Than Overall Survival in Oncology Clinical Trials. *JAMA Internal Medicine* 2019;179:642–47 doi:10.1001/jamainternmed.2018.8351; PMID:30933235;
- 61 Gyawali B, Hey SP, Kesselheim AS. Assessment of the Clinical Benefit of Cancer Drugs Receiving Accelerated Approval. *JAMA Internal Medicine* 2019;179:906–13 doi:10.1001/jamainternmed.2019.0462; PMID:31135808;
- 62 Downing NS, Aminawung JA, Shah ND, et al. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA* 2014;311:368–77 doi:10.1001/jama.2013.282034; PMID:24449315;
- 63 Naci H, Smalley KR, Kesselheim AS. Characteristics of Preapproval and Postapproval Studies for Drugs Granted Accelerated Approval by the US Food and Drug Administration. *JAMA* 2017;318:626–36 doi:10.1001/jama.2017.9415; PMID:28810023;
- 64 Pease AM, Krumholz HM, Downing NS, et al. Postapproval studies of drugs initially approved by the FDA on the basis of limited evidence: systematic review. *BMJ-BRITISH MEDICAL JOURNAL* 2017;357:j1680 doi:10.1136/bmj.j1680; PMID:28468750;
- 65 Prasad V, Kim C, Burotto M, et al. The Strength of Association Between Surrogate End Points and Survival in Oncology: A Systematic Review of Trial-Level Meta-analyses. *JAMA Internal Medicine* 2015;175:1389–98 doi:10.1001/jamainternmed.2015.2829; PMID:26098871;
- 66 Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007;25:5218–24 doi:10.1200/JCO.2007.11.8836; PMID:18024867;

- 67 Ciani O, Buyse M, Garside R, et al. Meta-analyses of randomized controlled trials show suboptimal validity of surrogate outcomes for overall survival in advanced colorectal cancer. *J Clin Epidemiol* 2015;68:833–42 doi:10.1016/j.jclinepi.2015.02.016; PMID:25863582;
- 68 Kim C, Prasad V. Cancer Drugs Approved on the Basis of a Surrogate End Point and Subsequent Overall Survival: An Analysis of 5 Years of US Food and Drug Administration Approvals. *JAMA Internal Medicine* 2015;175:1992–94 doi:10.1001/jamainternmed.2015.5868; PMID:26502403;
- 69 Chen EY, Haslam A, Prasad V. FDA Acceptance of Surrogate End Points for Cancer Drug Approval: 1992–2019. *JAMA Internal Medicine* 2020;180:912–14 doi:10.1001/jamainternmed.2020.1097; PMID:32338703;
- 70 Chen EY, Raghunathan V, Prasad V. An Overview of Cancer Drugs Approved by the US Food and Drug Administration Based on the Surrogate End Point of Response Rate. *JAMA Internal Medicine* 2019;179:915–21 doi:10.1001/jamainternmed.2019.0583; PMID:31135822;
- 71 Wallach JD, Ciani O, Pease AM, et al. Comparison of treatment effect sizes from pivotal and postapproval trials of novel therapeutics approved by the FDA based on surrogate markers of disease: a meta-epidemiological study. *BMC Med* 2018;16:45 doi:10.1186/s12916-018-1023-9; PMID:29562926;
- 72 Ciani O, Manyara AM, Davies P, et al. A framework for the definition and interpretation of the use of surrogate endpoints in interventional trials. *EClinicalMedicine* 2023;65:102283 doi:10.1016/j.eclinm.2023.102283; PMID:37877001;

Index

biais écologique, 25
Buyse, 17
Corrélation des effets, 17
JCA, 24
Join Clinical Assessment, 24

méta-épidémiologie, 27
rétrospectif, 23
surrogate, 8
Surrogate threshold effect, 19