

I Astuces pour l'analyse critique d'article scientifique

Comprendre simplement l'intérêt des outils statistiques permettant d'évaluer la valeur additionnelle d'un nouveau marqueur dans la prédiction du risque d'événements ?



T. PEZEL
Service de Cardiologie,
CHU Lariboisière, PARIS ;
Unité Inserm-UMR 942, PARIS.

Ce titre un peu barbare, je vous l'accorde, ne vous a donc pas découragé à lire cet article, et c'est tant mieux ! En effet, nous allons aborder de façon très simple une gamme d'outils statistiques de plus en plus présents dans les études observationnelles.

Le rationnel de ces outils est simple. Pendant longtemps, nous avons réalisé des études évaluant simplement l'association entre un nouveau marqueur (le score calcique coronaire, par exemple) et la survenue des événements cardiovasculaires. Ainsi, il est intéressant de montrer que plus le score calcique est élevé, plus le nombre d'événements cardiovasculaires est important. Cependant, la vraie question clinique que l'on a envie de se poser serait de savoir si cette association persiste même lorsque l'on prend en compte les facteurs de risque cardiovasculaire traditionnels (âge, diabète, hypertension...). Car, en réalité, si ce nouveau marqueur est associé à la survenue d'événements cardiovasculaires, mais que cette association est "annulée" par la prise en compte des facteurs de risque cardiovasculaire traditionnels (ajustement), alors ce nouveau marqueur n'a absolument aucun intérêt en routine clinique !

Rationnel de l'incrémentat de la valeur prédictive du modèle

Comme cela est évoqué dans l'introduction, de nombreuses études épidémiologiques permettent de décrire l'existence d'une association statistique entre un nouveau marqueur tel que le score calcique coronaire et la survenue d'un événement tel que la mort. Ainsi, il est intéressant de montrer que plus la valeur du score calcique coronaire est élevée, plus le risque de mortalité

augmente. On parle ici de "simple association statistique".

Pour aller plus loin, les épidémiologistes ont ensuite réalisé des **analyses multivariées** permettant d'ajuster l'association mesurée. En d'autres termes plus simples, ce que l'on appelle "ajustement" correspond au fait de supprimer l'effet d'autres facteurs comme, par exemple, les facteurs de risque cardiovasculaire traditionnels (âge, sexe, diabète, hypertension, tabac...) sur la

survenue de la mort. On aboutit alors à l'obtention d'un **“modèle”** capable de prédire la survenue de la mort en incluant différents paramètres dont notre nouveau marqueur.

Cependant, même si cet outil est très intéressant, il ne permet pas vraiment d'évaluer à quel point l'utilisation du nouveau marqueur est pertinente pour améliorer la prédiction de risque de ce modèle au-delà de l'utilisation des facteurs de risque traditionnels. Le fait de mesurer cette amélioration apportée dans la prédiction du risque par le nouveau marqueur s'appelle **“l'incrémenta-tion de la valeur prédictive du modèle”**.

Pour évaluer précisément cette incrémenta-tion de la valeur prédictive du modèle pour prédire la survenue d'événements, nous devons évaluer deux paramètres :

- la **discrimination** : capacité du nouveau marqueur à discriminer les patients qui vont faire un événement de ceux qui ne vont pas en faire ;
- la **reclassification** : capacité du nouveau marqueur à mieux reclasser un patient dans la bonne catégorie de risque de survenue de l'événement.

Ensuite, il existe différents outils statistiques permettant d'évaluer ces deux paramètres mais nous verrons dans cet article les trois outils suivants :

- pour évaluer la discrimination : Harrell C index (C-statistic) et *Integrative discrimination index* (IDI).
- pour évaluer la reclassification : *Net reclassification improvement* (NRI).

Quel est le meilleur outil pour évaluer cette incrémenta-tion de la valeur prédictive du modèle ?

Nous allons définir ensemble les principaux outils statistiques permettant d'évaluer ces deux paramètres que sont la discrimination et la reclassification. Mais ce qu'il est important

de comprendre, c'est qu'aucun de ces paramètres ne permettra jamais à lui seul de trancher sur cette incrémenta-tion de prédiction du risque ! En effet, il existe une littérature riche décrivant toutes les limites de chacun de ces outils. C'est d'ailleurs pour cela que les experts (*reviewers*) en statistiques du journal *Circulation* recommandent fréquemment d'associer les trois outils évoqués précédemment : C-statistic, NRI et IDI.

Effectivement, le rationnel étant que, bien qu'isolément très imparfaits, **si ces tests montrent tous ensemble que le nouveau marqueur a une incrémenta-tion dans la prédiction du risque, alors cela renforce fortement la fiabilité de ce résultat**.

Définitions

>>> Le **C-statistic** (ou C-index) correspond à la probabilité qu'un patient sélectionné au hasard et qui a présenté l'événement ait un score de risque plus

élevé qu'un patient qui ne présentera pas l'événement. Il est égal à l'aire sous la courbe ROC et varie de 0,5 à 1. Plus le C-statistic est proche de 1, plus le modèle a une bonne discrimination pour prédire la survenue de l'événement.

>>> L'**IDI** correspond à la différence moyenne du risque prédit pour chaque patient entre le modèle de référence et le modèle incluant le marqueur étudié. Ainsi, plus l'IDI est élevé, plus le nouveau marqueur permet d'améliorer la prédiction du modèle par rapport au modèle de référence.

>>> Le **NRI** correspond à une valeur qui additionne :

- la proportion de sujets sains qui ne réaliseront pas l'événement correctement car reclassés vers une catégorie de risque inférieur ;
- et la proportion de patients qui réaliseront l'événement correctement reclassés vers une catégorie de risque supérieur.

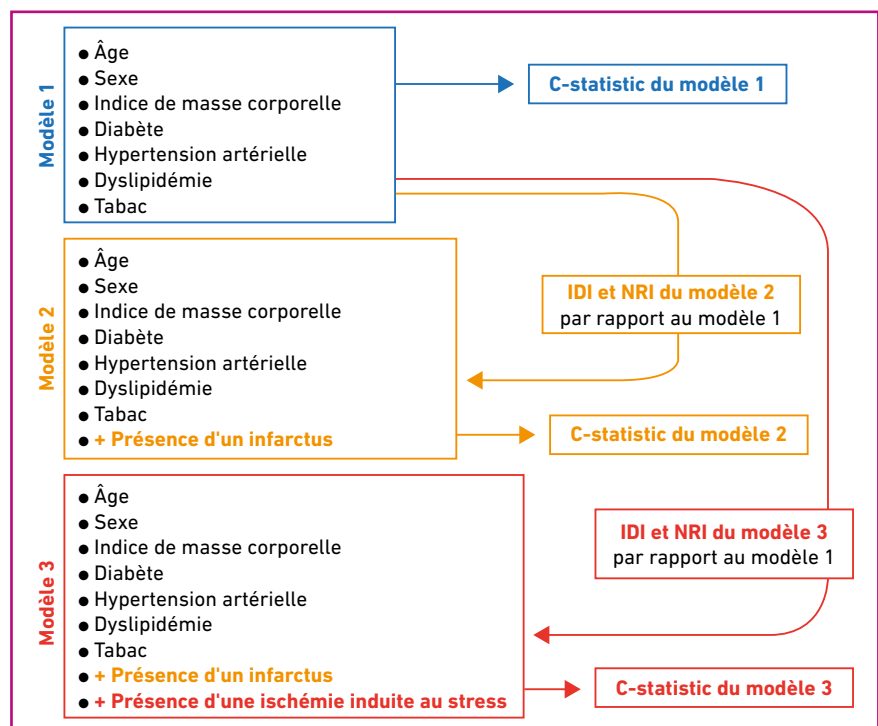


Fig. 1 : Principe des “modèles multivariés emboîtés” au service de l'évaluation de l'incrémenta-tion de la prédiction de risque.

Astuces pour l'analyse critique d'article scientifique

Enfin, la valeur globale du NRI peut être interprétée sous la forme d'un pourcentage de reclassification correcte des patients. Par exemple, dans l'**encadré 1**, le NRI du marqueur "AVA/height" égal à 0,33 signifie que l'utilisation de ce marqueur permet de bien reclasser 33 % des patients par rapport au modèle de référence.

Principe des modèles multivariés emboîtés

Afin de pouvoir utiliser ces outils diagnostiques qui montrent à quel point l'ajout d'un nouveau marqueur améliore la prédiction du modèle, nous allons

devoir utiliser ce que l'on appelle des "**modèles multivariés emboîtés**".

Le principe est simple :

- tout d'abord, on construit un premier modèle de variables correspondant à des variables connues comme des facteurs de risque traditionnels de survenue de l'événement que vous étudiez (exemple : âge, sexe, diabète...). On appellera ce modèle "**le modèle de référence**";
- puis, on construit un second modèle en reprenant exactement les mêmes variables du modèle de référence mais en ajoutant en plus le marqueur à tester;
- en outre, si l'on veut évaluer comment différents marqueurs se complètent dans la prédiction du risque en partant du

modèle de référence, alors on peut répéter cela plusieurs fois en créant encore un nouveau modèle emboîté (en ajoutant encore un nouveau marqueur au modèle précédent), et ainsi de suite...

Au final, chaque modèle emboîté va pouvoir être comparé au modèle de référence permettant d'évaluer à chaque étape à quel point le nouveau marqueur ajouté permet d'améliorer la prédiction de risque du modèle. Pour ce faire, nous aurons pour chaque modèle étudié les trois outils précédents cités : C-statistic, NRI et IDI (**fig. 1**).

Exemple d'une étude évaluant la valeur pronostique de l'IRM cardiaque de stress

Table. Discrimination and Reclassification Associated With AVA Normalized to Different Indexes of Body Size

Model	AIC	Harrell C Index	P Value	Net Reclassification Improvement	P Value	Integrated Discrimination Improvement	P Value
Multivariable model	1760	0.56 (0.52–0.61)		Reference		Reference	
Multivariable model* with AVA/BSA	1742	0.65 (0.60–0.69)	0.009	0.28 (0.15–0.43)	0.00018	0.08 (0.03–0.15)	0.007
Multivariable model* with AVA/height	1737	0.67 (0.62–0.71)	0.002	0.33 (0.17–0.45)	<0.00001	0.10 (0.05–0.16)	0.001
Multivariable model* with AVA/weight	1745	0.63 (0.59–0.68)	0.04	0.22 (0.06–0.34)	0.0006	0.06 (0.02–0.12)	0.02
Multivariable model* with AVA/BMI	1747	0.64 (0.60–0.68)	0.03	0.27 (0.10–0.40)	0.0002	0.07 (0.03–0.13)	0.01

Discrimination and reclassification was based on Harrell C statistic, net reclassification improvement, and integrated discrimination improvement. Results are for 2-year follow-up and 95% confidence intervals are presented. AVA/BSA, AVA/height, AVA/weight, and AVA/BMI are used in the models as continuous variables. AVA, aortic valve area; BMI, body mass index; and BSA, body surface area.

*The multivariable model included age, sex, Charlson comorbidity index, hypertension, coronary artery disease, atrial fibrillation, and left ventricular ejection fraction. AIC indicates Akaike Information Criterion.

Comprendre la présentation de ce tableau typique dans ce type d'étude :

- Chaque ligne représente le résultat pour un modèle multivarié constitué de plusieurs variables :
 - la première ligne "**Multivariable model**" correspond au modèle de base incluant les facteurs de risque traditionnels (âge, sexe, index de Charlson, hypertension, maladie coronaire, FA et FEVG);
 - puis, chaque ligne en dessous correspond à un modèle multivarié emboîté, pour lequel on a pris le "**Multivariable model**" de base afin de lui ajouter chacun des marqueurs à tester (chaque cadre bleu sur le tableau ci-dessus);
 - ainsi, chaque ligne permet de facilement visualiser comment chaque marqueur incrémente la valeur prédictive du modèle, permettant alors de comparer les marqueurs entre eux.
- Chaque colonne avec le cadre rouge présente les 3 outils statistiques (accompagnés de leur intervalle de confiance et p-value) permettant d'évaluer discrimination et reclassification de chaque modèle.

Interprétation

- On peut alors aisément comparer chaque modèle (chaque ligne) et donc mettre en évidence ici que c'est le marqueur "AVA/height" qui semble le plus pertinent car il donne :
 - la meilleure discrimination (C-statistic le plus élevé à 0,67 et IDI le plus élevé à 0,10);
 - et la meilleure reclassification (NRI le plus élevé à 0,33).
- Au final, le principal message de ce tableau est que le marqueur "AVA/height", autrement dit le fait de mesurer la surface aortique rapportée à la taille du patient, était le marqueur (test diagnostique) le plus performant pour améliorer le modèle de prédiction au-delà des facteurs de risque traditionnels connus.

Encadré 1 : Étude évaluant différents marqueurs de sévérité en échocardiographie d'un rétrécissement aortique sévère chez un patient asymptomatique dans la prédiction de survenue d'un remplacement valvulaire aortique ou de la mort (étude de Tribouilloy C et al. *Circulation Cardiovascular Imaging*, 2016).

permettant d'identifier la présence d'un infarctus et d'une ischémie induite au stress :

- le modèle 1 correspond au modèle de référence qui comprend uniquement les facteurs de risque traditionnels ;
- le modèle 2 inclut les variables du modèle 1 auquel on ajoute la variable “présence d'un infarctus” ;
- le modèle 3 inclut les variables du modèle 3 auquel on ajoute la variable “présence d'une ischémie induite au stress”.

La valeur prédictive de chaque modèle est alors évaluée par trois variables :

- C-statistic de chaque modèle ;
- NRI et IDI des modèles 2 et 3 prenant à chaque fois le modèle 1 comme référence.

Il est important de noter ici que le modèle de référence n'a pas de NRI ou IDI ! En effet, NRI et IDI se calculent toujours par rapport à un modèle référence afin de rendre compte d'une amélioration.

Utiliser ces méthodes pour comparer différents marqueurs entre eux

Nous venons d'aborder dans l'exemple de la **figure 1** l'utilisation des modèles

multivariés emboîtés afin d'évaluer comment plusieurs marqueurs se complètent entre eux, pour améliorer la prédiction du risque d'événement. Mais, ces outils peuvent aussi permettre de comparer différents marqueurs entre eux, par rapport à leur capacité respective d'améliorer la prédiction du modèle de référence. Pour cela, il suffit de comparer les C-statistic, NRI et IDI de chaque modèle (**encadré 1**).

Conclusion

Dans les études épidémiologiques, au-delà de la classique mesure d'association statistique entre un marqueur étudié et la survenue d'un événement, il est possible de mesurer à quel point **ce marqueur permet d'améliorer notre capacité à prédire l'événement par rapport aux facteurs de risque connus**. C'est la notion d'incrémentation du modèle de risque qui donne la possibilité d'aller encore plus loin, pour évaluer la pertinence clinique d'un marqueur ou d'un test diagnostique.

Cependant, attention, **ces outils peuvent présenter des limites importantes et il sera toujours essentiel de les interpréter**

avec grande précaution, et ne pas hésiter à les associer afin de renforcer la validité des résultats étudiés.

POUR EN SAVOIR PLUS

- TRIBOUILLOY C, BOHBOT Y, MARÉCHAUX S *et al.* Outcome Implication of Aortic Valve Area Normalized to Body Size in Asymptomatic Aortic Stenosis. *Circ Cardiovas Imaging*, 2016;9:e005121.
- Pencina MJ, D'Agostino RB, D'Agostino RB *et al.* Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*, 2008;27:157-172.
- HLATKY MA, GREENLAND P, ARNETT DK *et al.* Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation*, 2009;119:2408-2416.
- HOSMER DW, LEMESHOW S. Applied Logistic Regression (2nd Edition). New York, NY: John Wiley & Sons; 2000.
- PEZEL T. *Réussite à la lecture Critique d'article scientifique*. Paris, ed. Estem-Vuibert, Septembre 2020 ; 5^e édition.