

# Identification de sous-populations répondant à un traitement : effet individualisé du traitement ou essai adaptatif

Raphaël Porcher

Université Paris Cité, Inserm, Centre de Recherche Épidémiologie et Statistiques (CRESS-UMR1153)

Journées SMAC, Toulouse, 17 mars 2023

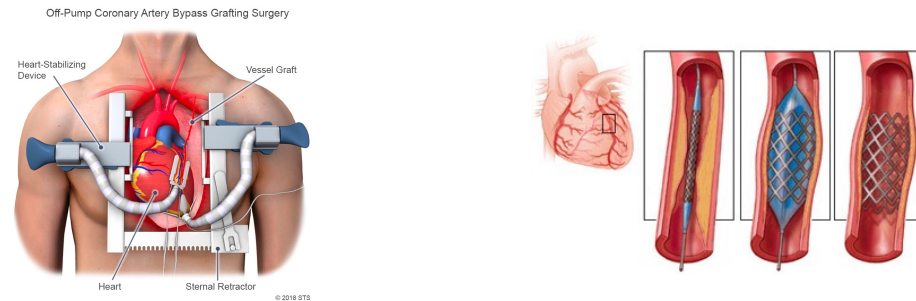


# Médecine personnalisée

- Donner le bon traitement au bon patient au bon moment
- Associer un traitement à des caractéristiques des individus
- De nombreux points de vue différents
  - Thérapie ciblée → adaptée à la connaissance de la maladie → médecine de précision
  - Thérapie ciblée → sous-groupe de patients dont la tumeur exprime la mutation → médecine stratifiée
  - CAR T-cells → adapter le traitement au patient → traitement individualisé
  - Prédire l'effet individualisé du traitement → médecine individualisée ? De précision (modèle)? Stratifiée?
- Une bonne partie des approches de la médecine personnalisée repose sur l'identification d'une sous-population qui « répond » à un traitement

# Un exemple en dehors de la cancérologie

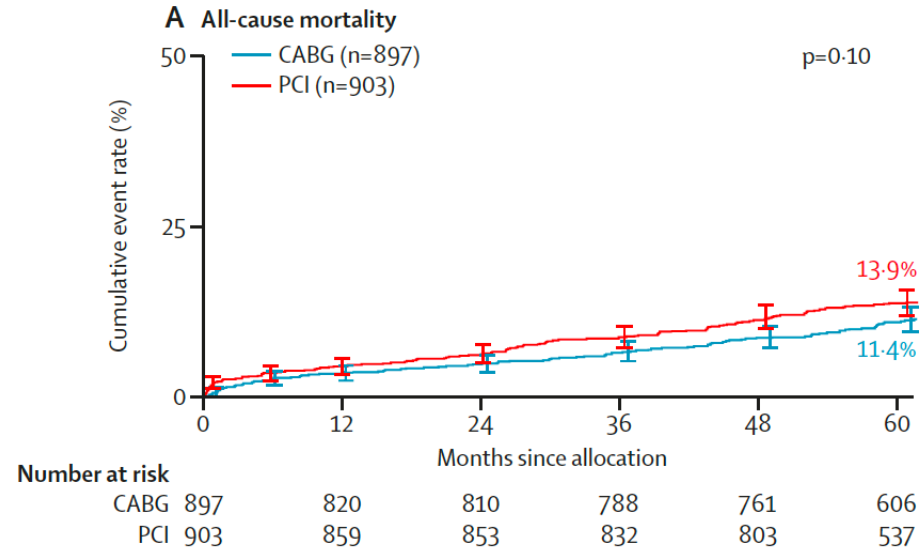
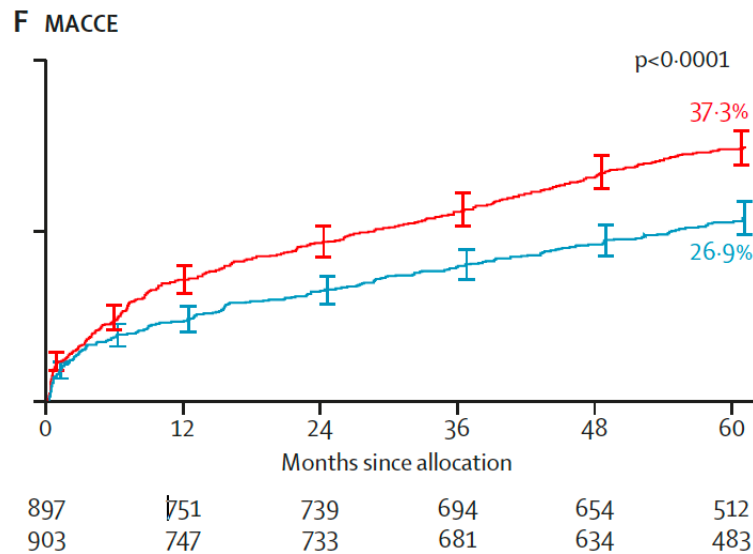
- Etude SYNTAX : essai contrôlé randomisé comparant l'angioplastie avec stent actif au pontage coronarien pour des patients avec une coronaropathie (lésions tri tronculaires et/ou de troncs communs gauches)



- Critère de jugement principal: événement cardiaque ou cérébrovasculaire majeur (MACCE): décès toute cause, AVC, infarctus du myocarde, ou nouvelle revascularisation

# Résultats de SYNTAX

- Plus de MACCE à 12 mois avec angioplastie
- Confirmé par l'analyse du suivi actualisé



- Doit-on arrêter de proposer des stents pour tous ?
- Ou l'angioplastie est-elle plus efficace que le pontage pour certains ?

# Deux patients fictifs



Sarah

- 69 ans, non fumeuse
- Diabétique, traitée par insuline
- FEVG 55%
- Clairance de la créatinine 72.5 mL/min
- Lésions tri tronculaires
- Pas de BPCO, pas d'artériopathie périphérique
- Score anatomique SYNTAX de 50



Donald

- 74 ans, fumeur
- Sans diabète, non traitée par insuline
- FEVG 40%
- Clairance de la créatinine 38.6 mL/min
- Lésion des troncs communs gauches
- Pas de BPCO, pas d'artériopathie périphérique
- Score anatomique SYNTAX de 11

Pontage, angioplastie?

# Formalisation de la question

- Modèle contrefactuel
  - Traitement  $A = 0$  ou  $1$ , on veut l'effet de  $A$  sur un critère  $Y$
  - On dispose d'un ensemble de variables (caractéristiques)  $X$
  - On définit des devenirs potentiels  $Y(0)$  et  $Y(1)$  selon que l'on fixe  $A = 0$  ou  $1$  et on suppose que si  $A = a$ , on observe  $Y = Y(a)$
- Effet individuel du traitement  $\tau = Y(1) - Y(0)$ 
  - Jamais observé
- Effet individualisé du traitement  $\tau(x) = E[Y(1) - Y(0) \mid X = x]$
- Si  $A$  est randomisé ou si on suppose que  $X$  comprend tous les facteurs de confusion
  - $E[Y \mid A = a, X = x] = E[Y(a) \mid A = a, X = x] = E[Y(a) \mid X = x]$
  - Et  $\tau(x) = E[Y \mid A = 1, X = x] - E[Y \mid A = 0, X = x]$
- Règle individualisée de traitement  $r$  : associe à chaque valeur de  $x$ ,  $r(x) \in \{0,1\}$ , le traitement (valeur de  $A$ ) recommandé

# Comment identifier une sous-population de répondeurs?

- A partir des données d'un essai (comme SYNTAX), de plusieurs essais (IPD-MA), ou de données observationnelles
- De **très nombreuses** méthodes
- Le plus simple (pour l'instant ?) est d'estimer  $\tau(x)$
- Si  $\tau(x)$  était connu, il est simple de montrer que la règle  $r(x) = I[\tau(x) > 0]$  est optimale
  - Au sens où elle maximise l'espérance de  $Y$  sous cette règle (si valeurs de  $Y$  élevées favorables)
- A partir d'une estimation  $\hat{\tau}(x)$  on peut naturellement définir une règle  $I[\hat{\tau}(x) > 0]$ 
  - Bonne propriétés si l'estimateur a de bonnes propriétés
- Autres stratégies : chercher directement une règle optimale, reformuler le problème comme un problème de classification pondérée, d'autres familles d'algorithmes

# Meta-learners

- T-learner
  - Estimer séparément deux fonctions  $\mu_a(x) = E[Y(a) | X = x]$  en utilisant les données de chaque groupe
  - Modèle de régression ou n'importe quel algorithme de ML
  - En déduire  $\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)$
- S-learner
  - Estimer un seul modèle de réponse en considérant le traitement comme une covariable
  - $\mu(x, a) = E[Y | A = a, X = x]$
  - En déduire  $\hat{\tau}(x) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0)$
- Autre meta-learners : X-learner (plutôt si groupes déséquilibrés), R(orthogonal)-learner (utilise le score de propension), ...
- Si utilisation de ML, risque important de sur-apprentissage → cross-fit



# Méthodes de construction de règles optimales

- Règles optimales maximisant un estimateur pondéré/double-robuste de la valeur<sup>1</sup>
  - Considérer une classe de règles  $r_\eta(x) = r(x, \eta)$
  - Maximiser la « valeur » de la règle sur la valeur de  $\eta$
  - Différents estimateurs possibles de la valeur
  - Utilisent un équivalent du score de propension (probabilité de traitement en accord avec la règle sachant les covariables)
    - IPWE ou AIPWE (double robuste)
- Outcome weighted learning<sup>2</sup>, A-learning<sup>3</sup>, ...

1. Zhang et al., Biometrics 2012; 68:1010-1018.

2. Zhao et al., JASA 2012; 107:1106-1118.

3. Chen et al., Biometrics 2017; 73:1199-1209.

# Autres approches (liste non-exhaustive)

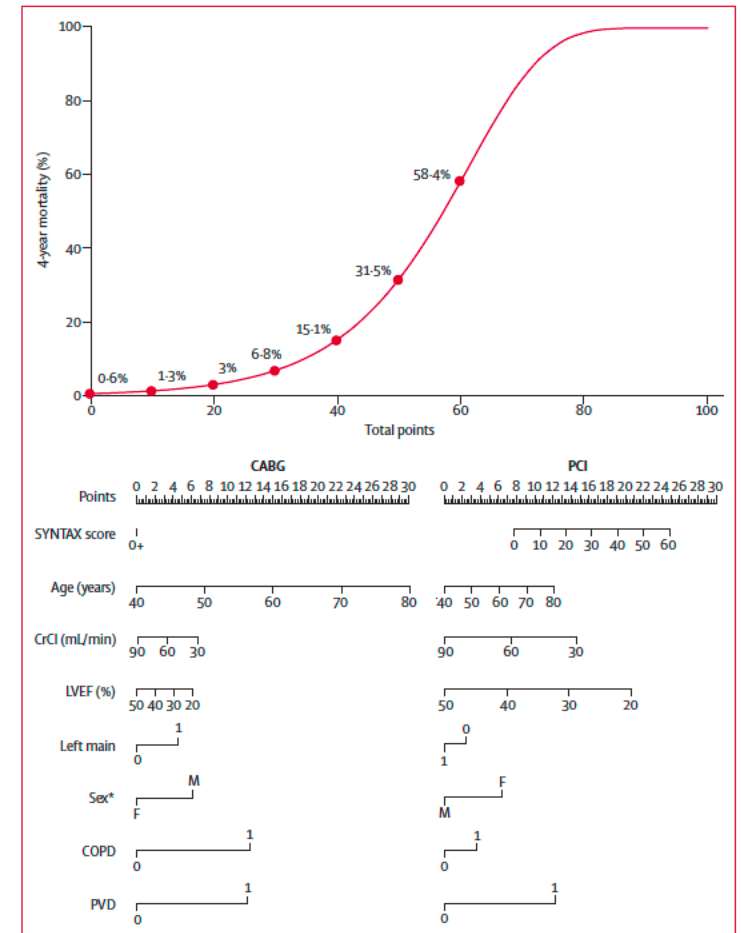
- Jumeaux virtuels (virtual twins VT)<sup>1</sup>
  - Plusieurs étapes
  - 1<sup>ère</sup> étape: estimer  $\tau(x)$  avec des forêts aléatoires (originellement type S-learner), puis en déduire les effets individualisés  $d_i$
  - 2<sup>e</sup> étape: soit construire un arbre de régression pour  $d_i = \hat{\tau}(X_i)$  et choisir les nœuds avec une prédiction  $>$  seuil  $\delta$ , soit construire un arbre de classification pour  $u_i = I(d_i > \delta)$  et prendre les nœuds prédisant  $u_i = 1 \rightarrow$  sous-ensemble  $S$  "bénéficiant" du traitement
  - Estimation du bénéfice dans  $S$  après correction par bootstrap
- Algorithmes type SIDES<sup>2</sup>
  - Identifie des sous-groupes avec un effet plus important
  - A partir d'arbres type CART
  - Algorithme récursif avec les nœuds et les coupes sélectionnés à partir des différences moyennes entre les groupes (plus dans 40 min)

1. Foster, Taylor Ruberg. Stat Med 2011; 30:2867-2880.

2. Lipkovich et al., Stat Med 2011; 30:2601-2621.

# Retour à SYNTAX

- Score SYNTAX-II: un premier modèle<sup>1</sup>, puis un second après actualisation du suivi et simplification (du modèle)<sup>2</sup>
- Premier modèle:
  - Développé dans les données de SYNTAX, validé dans celles du registre DELTA
  - Prédit la mortalité à 4 ans
  - S-learner (modèle de Cox avec covariables et interactions)
  - Variables: le score anatomique SYNTAX, âge, fonction rénale, fonction ventriculaire gauche, lésions du tronc commun coronaire gauche, sexe, BPCO et artériopathie périphérique
  - Estimation de l'effet du traitement individualisé
  - Règle de traitement: choisir le traitement avec la mortalité la plus faible

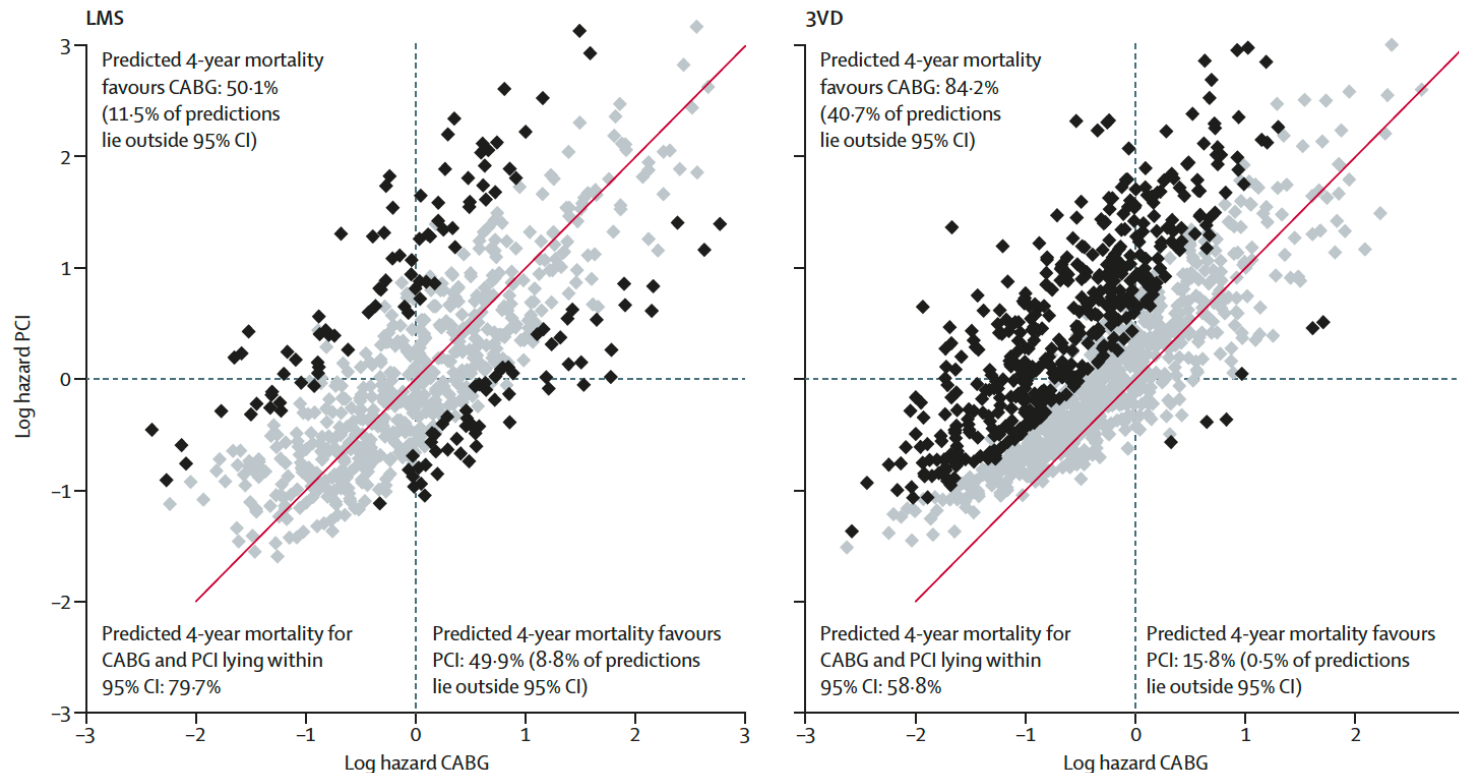


1. Farooq et al., *Lancet* 2013; 381:639-650.

2. Takahashi et al., *Lancet* 2020; 396:1399-1412.

# SYNTAX-II: effets individualisés

- Comparaison des prédictions individuelles sous chacun des traitements sur l'échelle des log-hazards
- « Zone grise » d'incertitude statistique



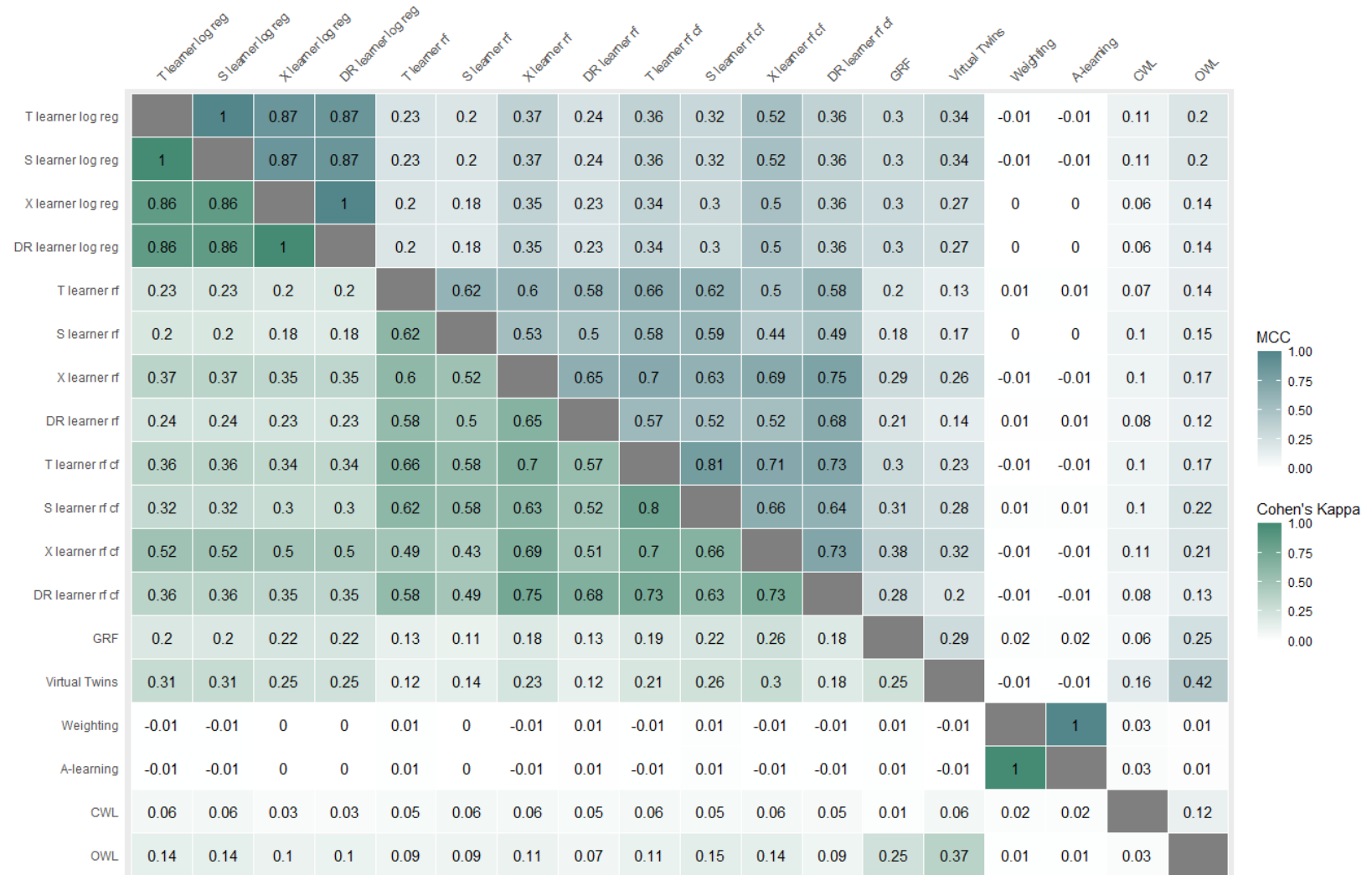
# Re-développement et validation

- Suivi actualisé
- Risque de surestimation de l'hétérogénéité de l'effet du traitement avec des interactions dans S-learner (encore pire avec T-learner)
- Redévelopper le modèle en développant un score pronostique sans le traitement, et inclure deux modificateurs de l'effet du traitement prédéfinis (type d'atteinte et score anatomique SYNTAX)
- Validation sur les données de trois essais (FREEDOM, BEST, PRECOMBAT)
- Calibration des modèles de prédiction, mais aussi de l'effet individualisé du traitement

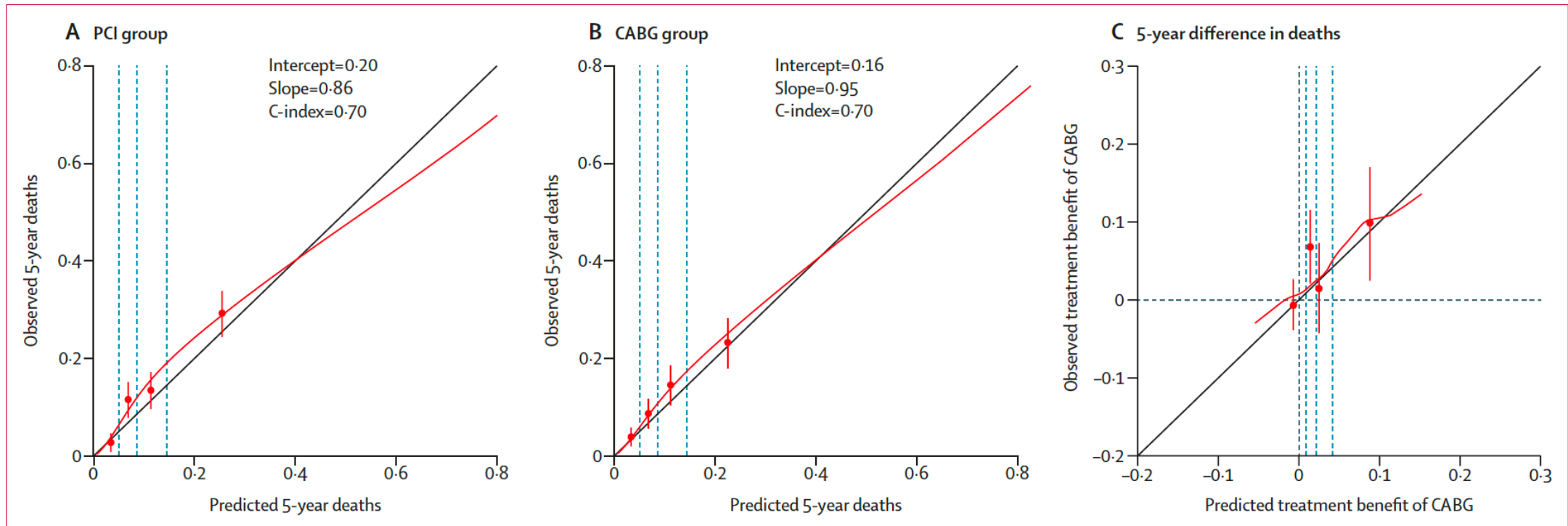
	Coefficient	HR (95% CI)	p value
Prognostic index	0.74	2.10 (1.84–2.39)	<0.0001
CABG × three-vessel disease	-0.48	0.62 (0.46–0.83)	0.0012
CABG × left main coronary artery disease	-0.10	0.91 (0.67–1.22)	0.5117
PCI × left main coronary artery disease	-0.23	0.80 (0.59–1.08)	0.1397
PCI × (SYNTAX score – 29) / 10	0.19	1.21 (1.07–1.37)	0.0021

# Aparté: risque de surestimation de l'hétérogénéité?

Concordance 2 à 2 des règles de traitement de 18 algorithmes utilisées sur les données de l'essai IST



# Calibration

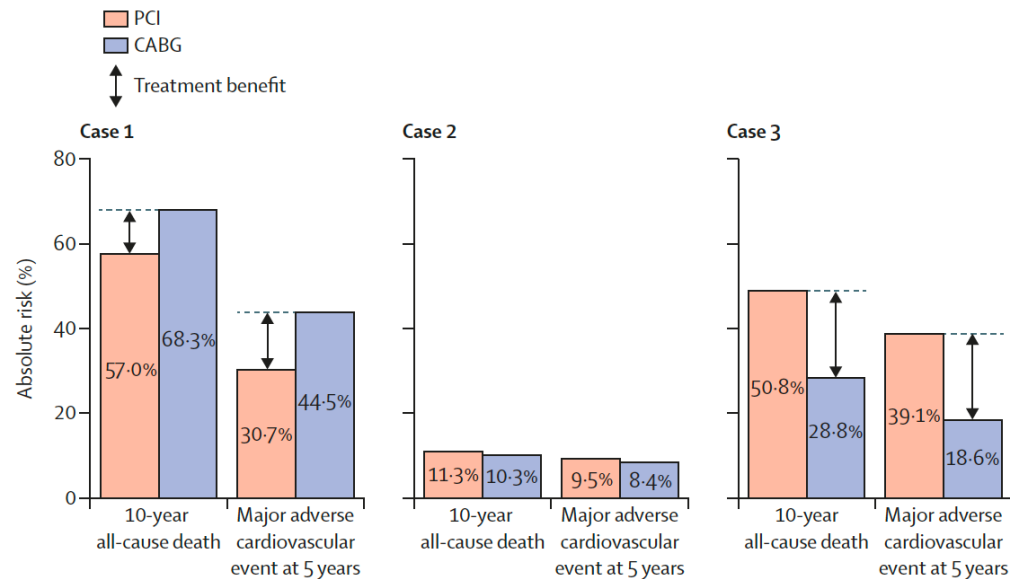


**Figure 2: External validation of the 10-year death model of the SYNTAX score II 2020 for predicting 5-year all-cause death in patients with multivessel disease or left main coronary artery disease in the FREEDOM, BEST, and PRECOMBAT trials (n=3380)**

Calibration plots showing the observed versus predicted 5-year deaths according to the SYNTAX score II 2020 in the PCI group (A) and CABG group (B). Vertical dashed lines represent quartiles of predicted 5-year deaths. (C) Calibration plot showing the observed versus predicted treatment benefit of CABG, according to the SYNTAX score II 2020. Vertical dashed lines represent quartiles of predicted treatment benefit of CABG. PCI=percutaneous coronary intervention. C-index=concordance index. CABG=coronary artery bypass grafting.

# Nos patients fictifs

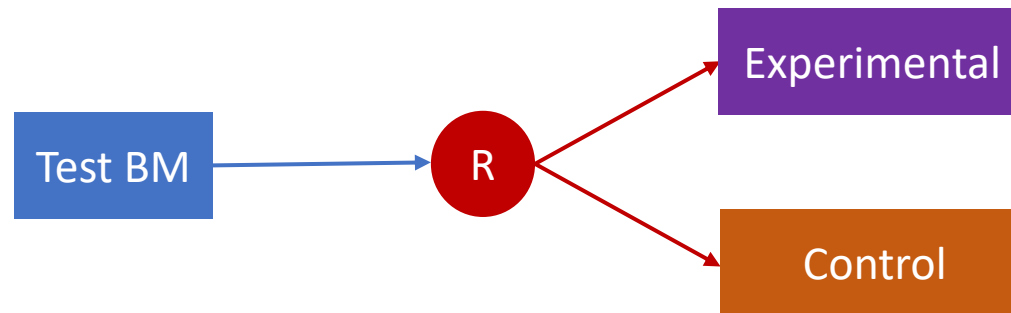
	Donald	Sarah
	Case 1	Case 2
Age, years	74	59
Diabetes	No	Yes
Receiving insulin	No	No
Creatinine clearance, mL/min per 1.73 m <sup>2</sup>	38.6	67.6
Left ventricular ejection fraction	40%	67%
Chronic obstructive pulmonary disease	No	No
Peripheral vascular disease	No	No
Current smoker	Yes	No
Three-vessel disease or left main coronary artery disease	Left main coronary artery disease	Three-vessel disease
Anatomical SYNTAX score	11	10





# Et les essais ?

- Problématique un peu différente de celle des *master protocols* (Stefan et Etienne, hier) (vraiment?)
- Un essai stratifié sur un biomarqueur permet de tester si l'effet du traitement diffère selon le sous-groupe
  - La stratification n'est même pas une condition formellement nécessaire



- Mais pas l'objectif principal de l'essai (contrôle du risque d'erreur de 1<sup>ère</sup> espèce)
- Et le sous-groupe (biomarqueur) doit être défini à l'avance
  - Ou on revient sur les analyses précédentes
- Autres designs adaptatifs?

# Essais adaptatifs d'enrichissement

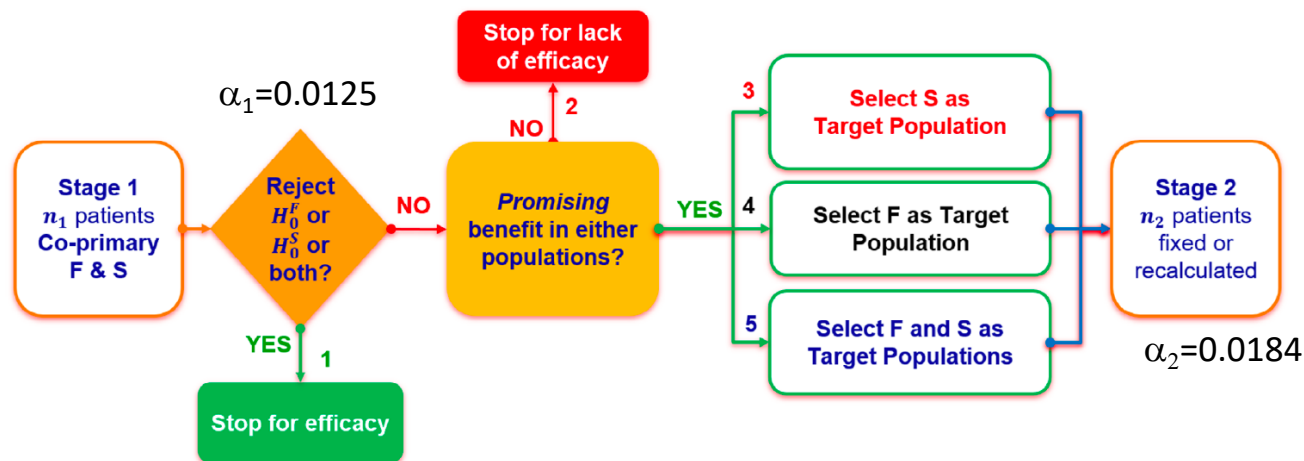
- Sélectionner un sous-groupe de façon adaptative au cours de l'essai
  - Lors d'une analyse intermédiaire planifiée
  - Avec un contrôle du risque de première espèce au sens fort
  - (sinon c'est très facile !)
  - Sous-groupe : pourquoi pas un sous-groupe clinique, mais plus fréquemment une sous-population génomique
- Plusieurs objectifs possibles dans le détail, mais l'idée est d'arrêter précocement dans la sous-population de non-répondeurs

# Différents types d'essais adaptatifs d'enrichissement

- Pré-spécification de la sous-population
  1. Complètement spécifiée lors de la planification
  2. Complètement spécifiée à la fin de la première étape de l'essai
- Sélection sur un critère de substitution ou pas
- Test de ( $H_0$ ,  $H_{0+}$ ): hypothèse nulle "dual-composite"
  - On n'étudie pas spécifiquement la sous-population BM-
  - Identifier l'ensemble de patients "responsable" de l'effet traitement
  - Conclure dans la population globale si possible
  - Se limiter à la sous-population BM+ sinon
- Ou tester  $H_{0+}$  et  $H_{0-}$  (hypothèse nulle "dual-independent")
  - Plusieurs approches:
    - Tester une interaction traitement x sous-population à l'analyse intermédiaire
    - Possibilité de séparer les tests en supériorité/futilité dans les sous-groupes
  - Questions de l'ajustement pour multiplicité débattue

# Exemple: IMpassion031

- Cancer du sein triple-négatif précoce
- Chimiothérapie + atezolizumab 840 mg i.v. ou placebo
- Co-objectifs principaux: réponse complète anapath dans l'ensemble de la population (F) ou dans le sous-groupe PD-L1+ (S)
- *Adaptive enrichment design*, hypothèses nulle dual-composite

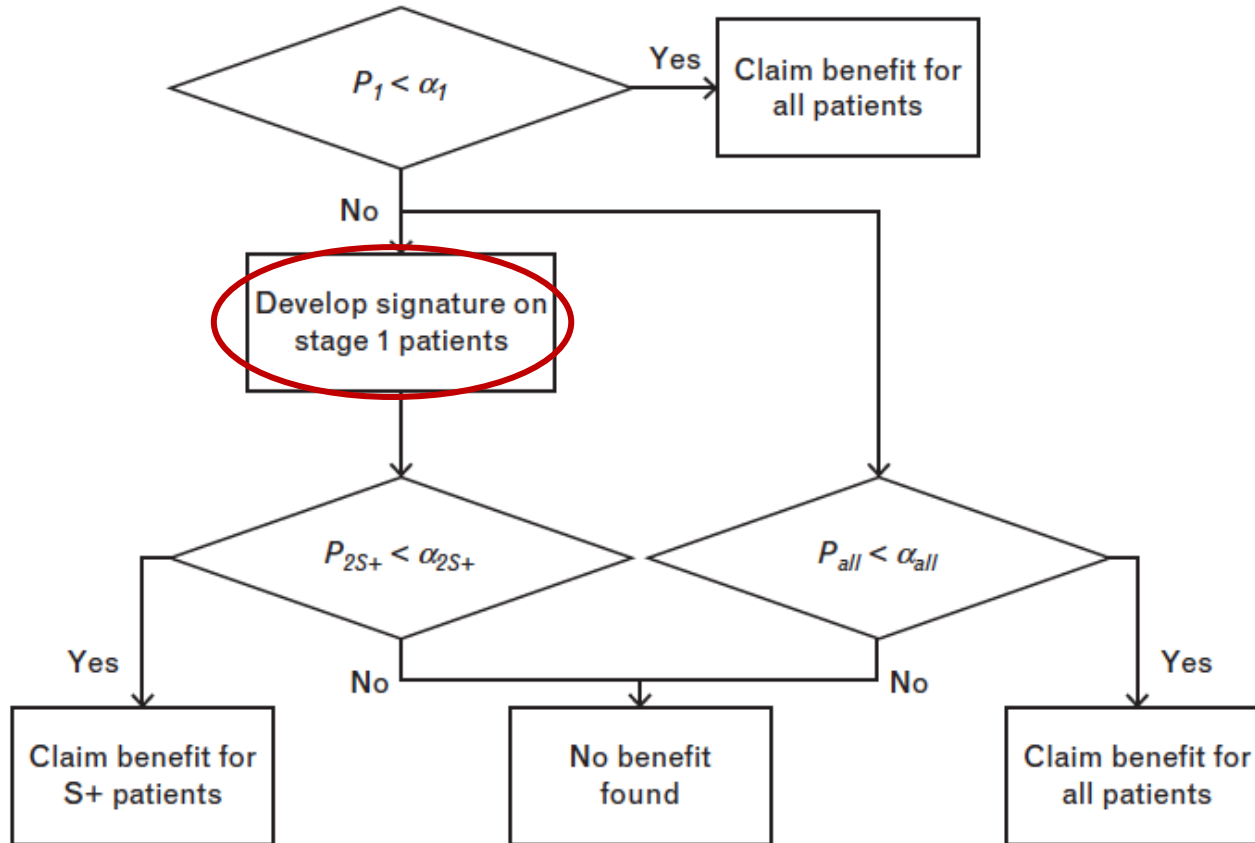


$\alpha_1$  et  $\alpha_2$  fixés pour contrôler le risqué d'erreur de 1<sup>ère</sup> espèce global à 0.025 en utilisant un test de combinaison

# *Adaptive signature designs*

- Assez proche du précédent
- Différence principale: développer une « signature » pour prédire la réponse à partir des données des patients de la première étape
- Signature: plutôt données -omiques
- Applique la signature prospectivement aux patients de la 2<sup>e</sup> étape pour identifier des « répondeurs prédits »
- Analyser cette sous-population et la population globale (hypothèse nulle dual-composite)
- Extension avec une approche de validation croisée pour optimiser le développement et la validation de la signature → Cross-validated(CV) ASD

# Adaptive signature designs: schéma



En fixant  $\alpha_1$ ,  $\alpha_{2S+}$  et  $\alpha_{all}$  pour contrôler le risque d'erreur de 1<sup>ère</sup> espèce sur l'ensemble

CVASD permet de limiter le risque de faux positifs qui sinon, peut être élevé<sup>1</sup>

# Conclusion

- Deux approches différentes ... pas complètement incompatibles
- Les objectifs ne sont pas superposables : quel est le meilleur traitement en fonction d'un ensemble de caractéristiques vs. (dé)montrer un effet traitement dans un groupe particulier
- Pléthore de méthodes existantes (surtout pour les règles individualisées de traitement)
- Titre de la session : Comment arriver à un enregistrement dans une sous-population ?

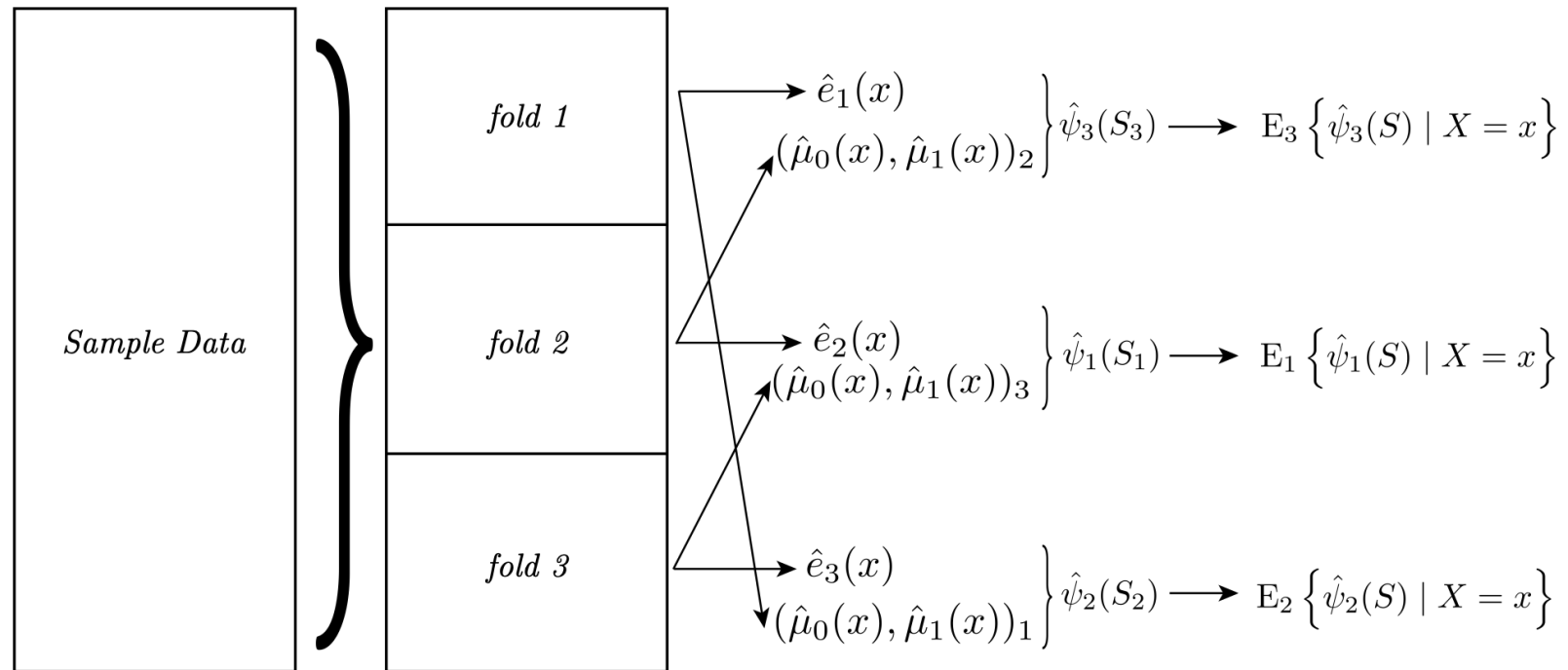
Merci de votre attention

et à François Grolleau, François Petit, Alan Balendran,  
Céline Béji, Florie Brion-Bouvier, Etienne Peyrot,  
Corentin Ségalas, Cecilia Superchi



# Cross-fitting

- De nombreuses variantes ...



**Figure 2.2:** Double sample splitting procedure and cross-fitting.