

Score de propension

Sibilia Quilici
sanofi pasteur MSD

Séminaire JEM-SFES
31 Janvier 2013

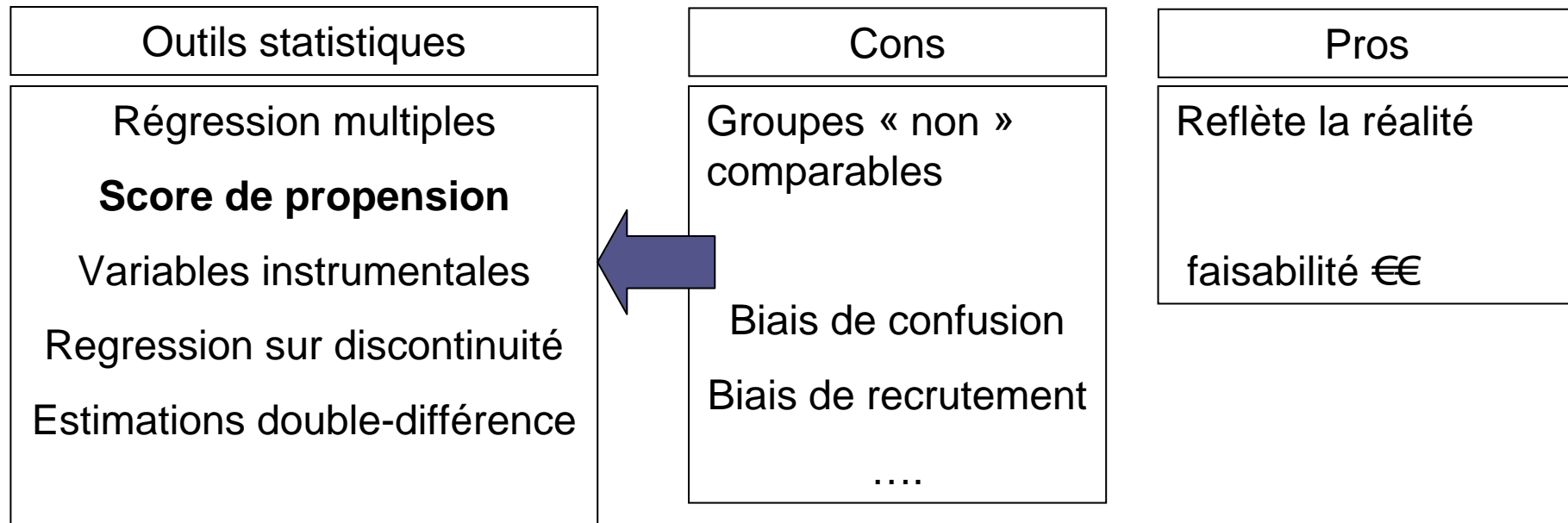
Etudes expérimentales vs. Non-expérimentales



Absence de randomization

Population en vie réelle

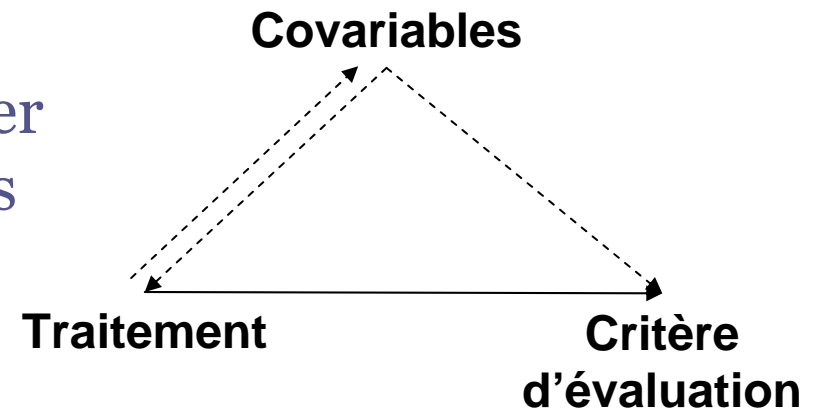
Pratique médicale en conditions réelles



Impact de la non-randomisation sur l'estimation de l'effet traitement

▫ Biais de sélection

- Sujets traités peuvent différer systématiquement des sujets non-traités



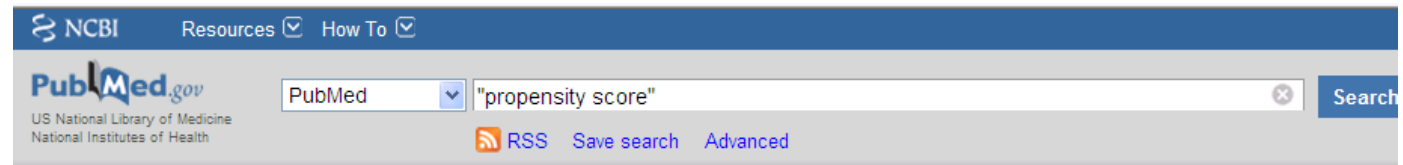
▫ Biais de causalité

- L'attribution du traitement peut être dépendante des caractéristiques individuelles

Contrôle de ces biais → Régression multivariée

Score de propension

Méthode initialement proposée par Rosembaum et Rubin en 1983



En quoi cela consiste?

- Population à partir d'une base de données observationnelles
- Sélection d'une population sur la base de ses caractéristiques
- Probabilité de recevoir un traitement selon ces caractéristiques => score de propension
- Création de groupes comparables en terme de caractéristiques individuelles => méthodes d'ajustement
- Évaluation de l'effet traitement

Définition

- Probabilité pour un individu de recevoir une intervention (traitement) conditionnellement à un ensemble de ses caractéristiques
- $\Pr (Tr = 1 / X_1, X_2, \dots X_n)$
- Dans le cadre d'un schéma d'étude expérimentale (RCT) le score de propension est connu et égal à 0,5
- Dans le cadre d'une étude observationnelle, il varie d'un patient à l'autre => il doit être estimé à partir des données

L'étude doit être exempte de "biais cachés"

- L'attribution du traitement est liée uniquement aux caractéristiques observées des individus
- Les sujets ayant le même score de propension sont alors « échangeables »
- Conditionner l'effet causal du traitement sur le score de propension permet une estimation de cet effet non biaisée
- Hypothèse forte d'indépendance entre appartenance au groupe de traitement et effet de traitement conditionnellement aux co-variables considérées dans le modèle.

Utilisation du score de propension

- Remplacement de toutes les covariables par une variable unique qui est une fonction de ces covariables (condensation de l'information)
- Utilisation comme une covariable de confusion unique

Estimation du score de propension par la régression logistique



$$\text{Log} \frac{\Pr(\text{Tr}_i=1 \mid x_i)}{1 - \Pr(\text{Tr}_i=1 \mid x_i)} = \text{Log} \frac{e^{x_i}}{1 - e^{x_i}} = \alpha_0 + \alpha_1 x_{i,1} + \dots + \alpha_n x_{i,n}$$

$\text{Tr}_i = 1$: Intervention
 $\text{Tr}_i = 0$: Comparateur

X_i : caractéristiques initiales patients

Possibles autres méthodes d'estimation:

- Modèle probit, analyse discriminante, arbre de classification / de régression, réseau de neurones



Choix des variables explicatifs

- Toutes les variables initiales
 - Variables cliniquement pertinentes
 - Facteurs de confusion potentiels
 - Variables liées au traitement
 - Variables liées au critère d'évaluation
-
- Stratégie de sélection des variables est fonction de l'effectif et du type d'ajustement effectué secondairement

Adéquation du modèle

- Critère classique d'adéquation du modèle de régression logistique
 - Test de Hosmer-Lemeshow
 - Aire sous la courbe ROC
- Ces mesures standards de performance ne sont pas recommandées pour évaluer l'adéquation du modèle de score de propension

Adéquation du modèle

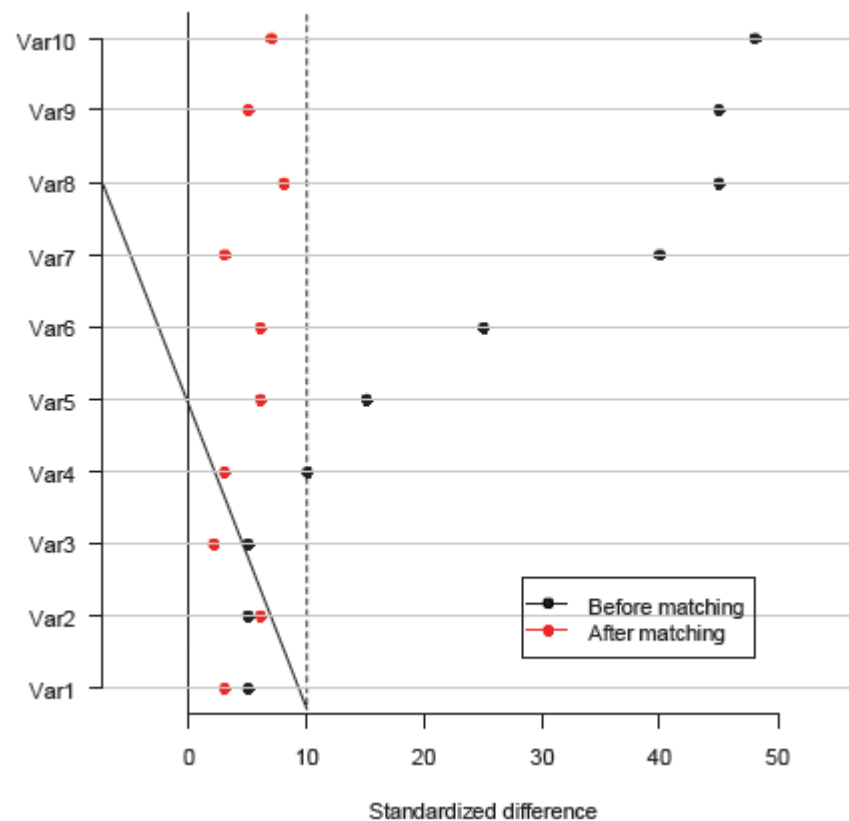
- Il faut utiliser des méthodes permettant d'évaluer l'équilibre en termes de covariables initiales dans la base "matchée"
 - Ne doit pas être influencées par la taille de l'échantillon
 - Doivent être des propriétés de l'échantillon
- Mesures de l'équilibre entre des facteurs pronostics entre les groupes
 - Différences standardisées
 - Quantiles-quantile plots
 - Density plots

Différences standardisées

- Interprétées comme différences de moyenne en unité d'écart-type
- Indépendante de la taille de l'échantillon, et de l'unité de mesure (peuvent être comparées)
- Pas de consensus sur « différence acceptable »
- Peuvent être calculées sur échantillon initial ou apparié

$$d = \frac{100(\bar{x}_{\text{treatment}} - \bar{x}_{\text{control}})}{\sqrt{\frac{s_{\text{treatment}}^2 + s_{\text{control}}^2}{2}}} \quad \text{for continuous variables}$$
$$d = \frac{100(\hat{p}_T - \hat{p}_C)}{\sqrt{\frac{\hat{p}_T(1 - \hat{p}_T) + \hat{p}_C(1 - \hat{p}_C)}{2}}} \quad \text{for dichotomous variables}$$

Exemple de déséquilibre utilisant les différences standardisées



Le score de propension est utilisé pour estimer l'effet traitement

Méthode d'utilisation

Ajustement (Modèle de régression)

Stratification

Appariement

Pondération inverse

1- Modèle de regression avec ajustement par le score de propension

Effet traitement = traitement + score

Avantages


- Identiques au modèles de régression standard pour ajustement
- Utilisation de modèles familiers: Logistique, Cox ...

Limites

- Suppose que le modèle de regression est correctement spécifié
- Suppose qu'il y a un "overlap" suffisant des scores entre traités et non traités
- Perte de la capacité à mimer une étude randomisée

2 - Stratification sur le score de propension

Stratification des sujets à partir du score estimé
(utilisation des quintiles)



Estimation de l'effet traitement dans chaque strate
(traité vs. non-traité)



Calcul de l'effet traitement global: "Pool" des estimations spécifiques de chaque strate

2 -Stratification sur le score de propension

Avantages

- La stratification par quintiles du score de propension permet d'éliminer environ 90% des biais liés aux variables de confusion mesurées si on estime un effet linéaire

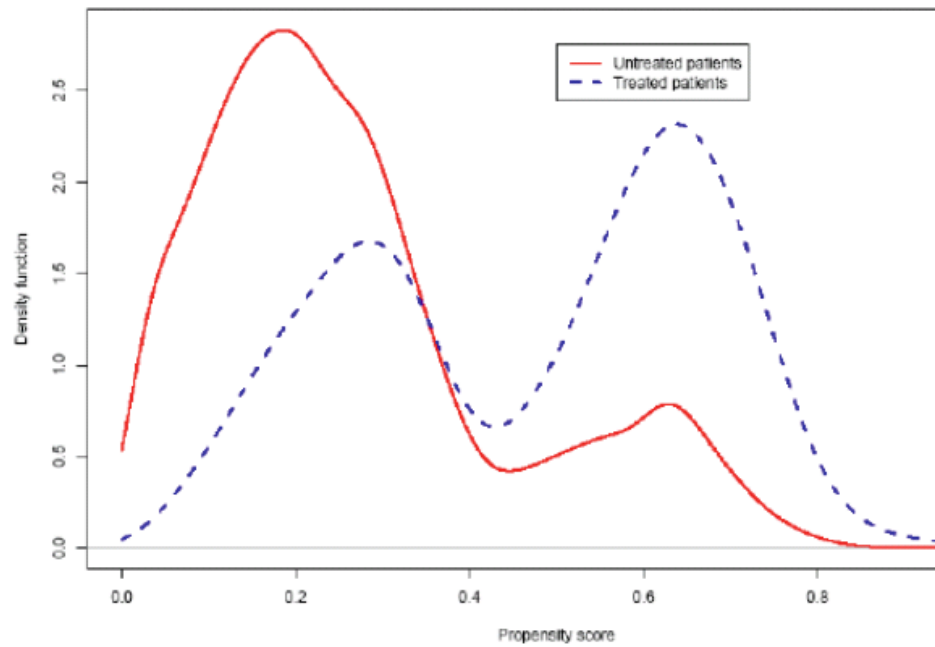
Limites

- Les strates en quintiles n'assurent pas l'homogénéité en terme de score de propension
- => il existe un déséquilibre résiduel entre traités et non-traités au sein d'une même strate

3 - Appariement sur le score de propension

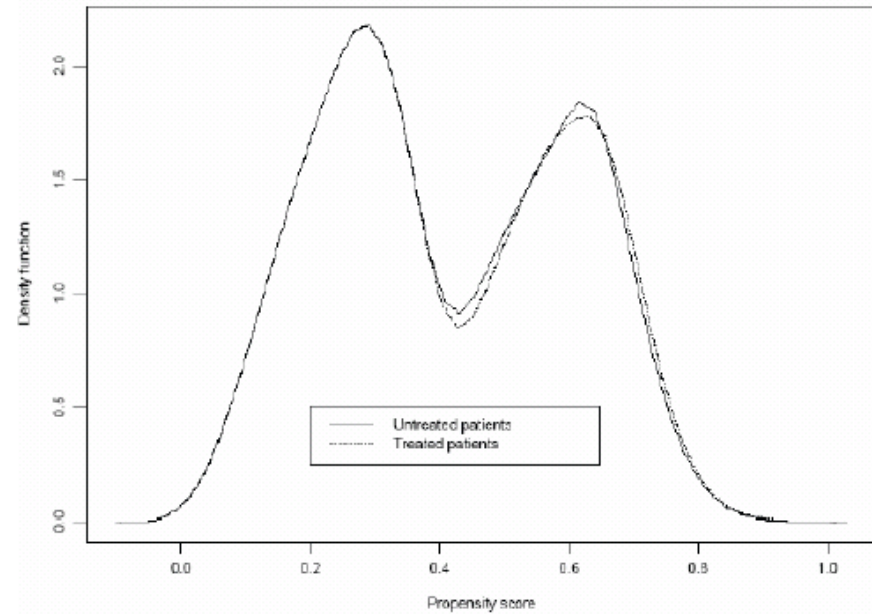
- Création de paires de sujets: traité - non traité qui ont une « même » valeur de score
- L'effet traitement est estimé sur la base appariée
- Pas de consensus sur ce que signifie “même score”
 - Il existe différentes méthodes d'appariement
 - De nombreuses méthodes sont basées sur le plus proche voisin (e.g. “Caliper matching”, “nearest neighbor matching”, “Mahalanobis metric matching” etc...)

- Avant appariement

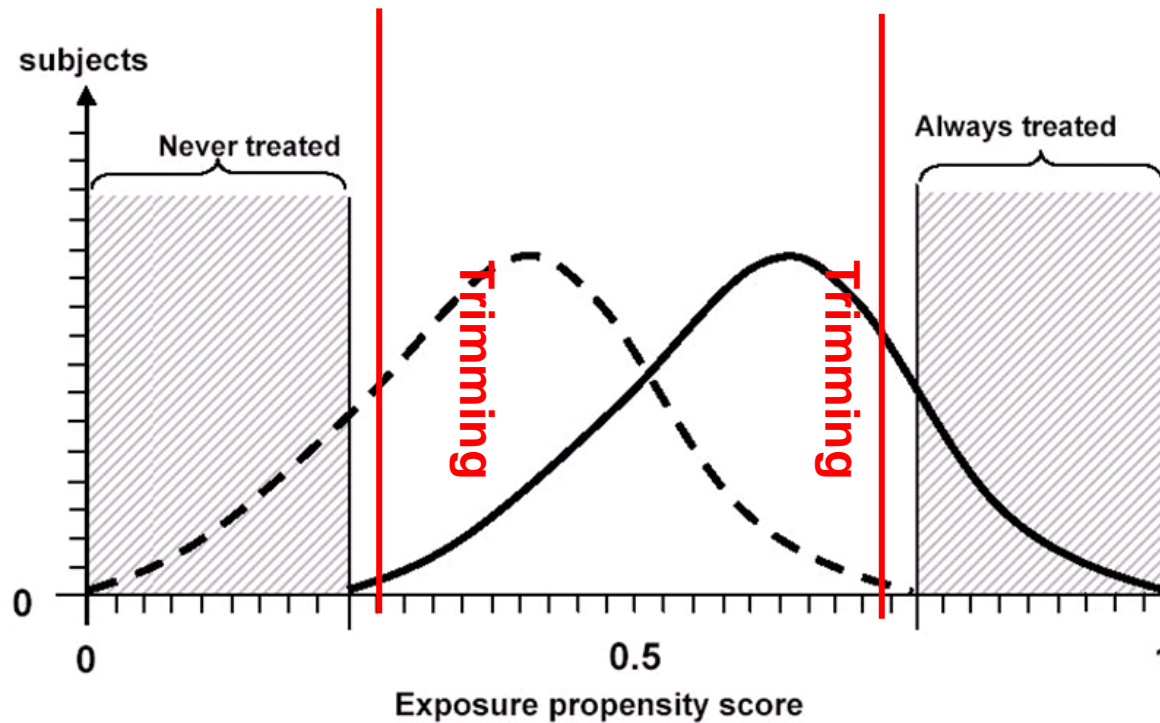


- Après appariement

Figure 4. Distribution of propensity score in treated/untreated patients (matched analysis)



Score de propension: Concept



- = Treated subjects
- - - = Untreated subjects

the non-overlap of the exposure propensity score distribution among treated and untreated study subjects

in this example subjects with very low propensity score are never treated while subjects with very high propensity score are all treated.

3- Appariement sur le score de propension

Avantages

- Mime l'effet traitement dans le cadre d'une randomisation
- Réduction très efficace du biais

Limites

- Réduction de la taille de l'échantillon sur sujet « matchable » uniquement
- => baisse de puissance
- => réduction l'application aux sujets matchés

4 - Pondération inverse sur le score de propension (IPTW)

- Auteurs: Imbens, Hirano (2000, 2001)
- Développé dans le cadre de modèles pour des traitements évoluant au cours du temps (modèles structurels marginaux)
- Création d'une pseudo-population dans laquelle à chaque instant t :
 - La probabilité d'être traité n'est plus liée aux facteurs pronostiques
 - L'effet de l'exposition est la même que dans la population cible

4 - Pondération inverse sur le score de propension

- Auteurs: Imbens, Hirano (2000, 2001)
- Développé dans le cadre de modèles pour des traitements évoluant au cours du temps
- Création d'une pseudo-population dans laquelle à chaque instant t :
 - La probabilité d'être traité n'est plus liée aux facteurs pronostiques
 - L'effet de l'exposition est la même que la population cible

En plus clair...

- Chaque sujet reçoit une pondération inverse de sa probabilité d'avoir reçu son traitement
 - Pour les sujets traités: $w = [PS]^{-1}$
 - Pour les sujets non-traités: $w = [1-PS]^{-1}$

Avantages

- Inclusion de tous les sujets ($w > 0$)
- Ne dépend pas d'échantillonnage aléatoire
=> plus robuste
- Estimation de l'effet traitement via une régression logistique pondérée

Le score de propension est utilisé pour estimer l'effet traitement

Méthodes recommandées pour la minimisation des biais

Ajustement (Modèle de régression)

Stratification

Appariement

Pondération inverse

5 avantages pour utiliser le score de propension - Glynn et al. 2006

- *The value of focus on indications for drug use*
- *Optimal matching strategies from alternative designs*
- *Improved control of confounding with scarce outcomes*
- *Ability to identify interactions between propensity of treatment and drug effects on outcomes*
- *Correction for unobserved confounders via propensity score calibration*

Discussion

- **Préoccupations statistiques**
 - Les méthodes statistiques conventionnelles supposent l'indépendance des observations
 - Les méthodes utilisant le score de propension doivent prendre la nature appariée des données
 - Les sujets d'une paire sont plus proche que des sujets aléatoirement choisis parmi traité – non traité
 - Les sujets d'une même paire ne sont pas indépendant => les groupes exposés - non exposé ne sont pas indépendant
- **Les méthodes de score de propension permettent:**
 - la création de strates de patients
 - De recréer le schéma d'un essai thérapeutique randomisé conditionnellement aux covariables observées
 - De déterminer explicitement le degré avec lequel les variables confondantes mesurées sont bien équilibrées entre groupe traité et non traité

Conclusion

- Objectif de la méthode: obtenir deux groupes équilibrés en terme de co-variables déterminées a-priori et non comme si c'était une analyse ad-hoc
- Importance du choix des co-variables: Inclure toutes les variables dans le modèle de score de propension
- Obtention d'un estimateur non biaisé de l'effet conditionnel du traitement
- Variables de confusion non mesurées: influence limitée si d'autres variables corrélées sont observées.



Conclusion

- L'utilisation du score de propension ne peut corriger les biais inhérents aux variables de confusion non-observées....
- Quid : utilisation variables instrumentales?

References

- Rosenbaum P, Rubin DB; *Biometrika* 1983. The central role of the propensity score in observational studies for causal effect
- Austin PC. *Stat Med* 2007 A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003.
- Schneeweiss S, *Pharmacology & therapeutics* 2007 Development in post marketing comparative effectiveness research
- Schneeweiss S, *Pharmacoepidemiology and drug safety* 2006. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics
- Glynn et al. *Basic Clin Pharmacol Toxicol.* 2006 . Indications for Propensity Scores and Review of Their Use in Pharmacoepidemiology
- Johnson et al. *Value in health* 2009. Task force ISPOR - Good Research Practices for Comparative Effectiveness Research: Analytic Methods to Improve Causal Inference from Nonrandomized Studies of Treatment Effects Using Secondary Data Sources: The ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report—Part III

Merci

Sibilia Quilici