

Société Française d'Économie de la Santé
« Modéliser pour Décider »
Lyon, le 19 Avril 2012

*Enjeux et Méthodologie
de L'évaluation Comparative des Politiques
Sanitaires Publiques*

Pr Robert Launois

ReesFrance

28, rue d'Assas

75006 Paris – France

Tel . 01 44 39 16 90 – Fax 01 44 39 16 92

E-mail : reesfrance@wanadoo.fr - Web : www.rees-france.com

Plan de l'intervention

- ✦ Etayer une position par des preuves....un défi
- ✦ La chasse aux Biais
- ✦ Apprendre à gérer les interférences parasites
- ✦ Faire parler les banques médico-administratives
- ✦ Le choix du schéma d'étude, un processus itératif

Les Enjeux

- ✦ La **politique mise en place a-t-elle eu les effets attendus** sur les individus ou la population ciblée?
- ✦ Comment **isoler les effets produits par un programme, des effets produits par d'autres facteurs concomitants ?**
- ✦ **Le repérage de ces relations causales entre intervention et résultats** est au cœur de l'évaluation
- ✦ Il convient de fournir aux décideurs politiques et aux acteurs sociaux **des arguments qui reposent sur des études empiriques**
- ✦ Ce qui **exige des données et un modèle statistique** capable de fournir un résumé : par exemple une différence de moyennes. Ce sont ces résumés que l'on interprète comme des résultats

ETAYER UNE POSITION PAR DES PREUVES ...UN DEFI

L'Imputabilité Individuelle : un Idéal Inaccessible

- ✦ **Objectif** : évaluer pour chaque individu i , le lien de cause à effet
 - Entre l'intervention T_i (**programme**, traitement) et,
 - Le résultat Y_i (indice thérapeutique d'efficacité)
- ✦ **l'effet causal** (ΔE_i) du traitement sur le résultat pour **un** bénéficiaire du programme doté est égale à

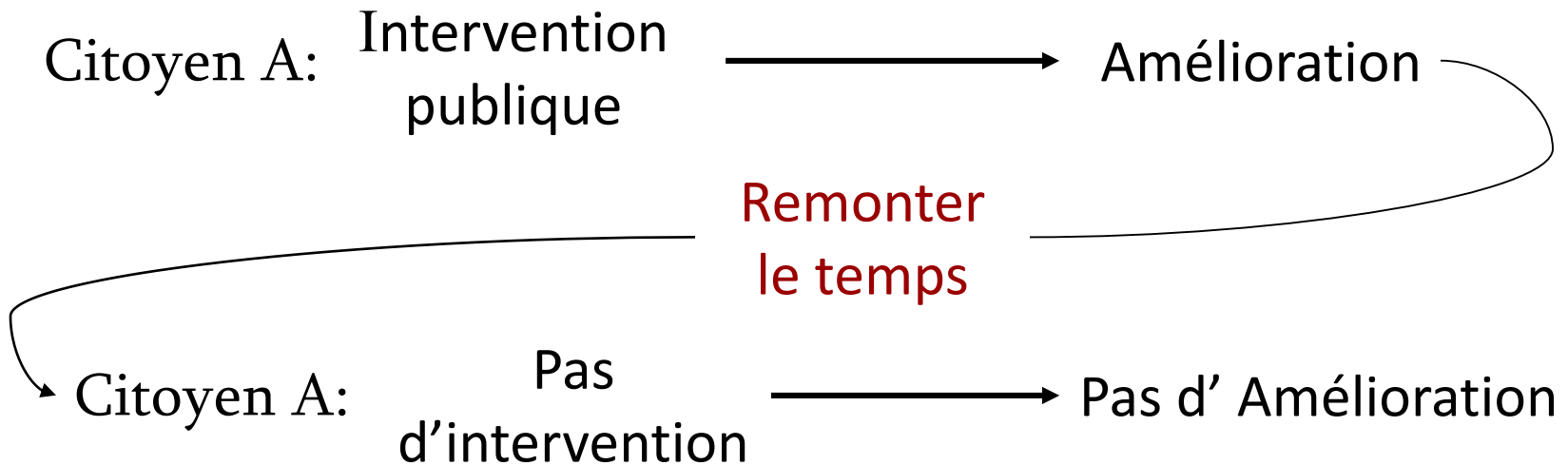
$$\Delta E_i = (Y_{i1} | X_i, T_i = 1) - (Y_{i0} | X_i, T_i = 0)$$

- Y_{i1} correspond au bénéfice que le sujet i a retiré de la mise en place du programme, ce que l'on indique par la variable $T=1$: $(Y_{i1} | X_i, T_i = 1)$
- Y_{i0} correspond au résultat qui aurait été obtenu si le programme n'avait pas été mis en oeuvre, ce que l'on indique par la variable $T=0$: $(Y_{i0} | X_i, T_i = 0)$
- X_i symbolise les caractéristiques observables des bénéficiaires

- ✦ Y_{i0} est dit **contrefactuel** parce que hypothétique. L'estimation de l'effet causal individuel est impossible : un seul des deux résultats potentiels est observé

le même sujet, ne peut pas au même instant, être bénéficiaire et non bénéficiaire d'un traitement ou d'une politique publique

Principe du Raisonnement *Contrefactuel*



Effet Causal Moyen Populationnel

- ✦ Une situation contrefactuelle ne peut être estimée qu'en calculant la différence ΔE^{TT} entre le niveau de santé moyen $E(Y_{i1})$ des participants ($T=1$) et celui $E(Y_{i0})$ des non participants au programme ($T=0$)

$\Delta E^{TT} = E(Y_{i1} | X_i, T_i = 1) - E(Y_{i0} | X_i, T_i = 0)$ Avec le retrait et l'ajout de $(.)$, il vient

$$\Delta E^{TT} = E(Y_{i1} | X_i, T_i = 1) - E(Y_{i0} | X_i, T_i = 1) + E(Y_{i0} | X_i, T_i = 1) - E(Y_{i0} | X_i, T_i = 0)$$

$$\Delta E^{TT} = \underbrace{E(Y_{i1} | X_i, T_i = 1) - E(Y_{i0} | X_i, T_i = 1)}_{\Delta^{TT}} - \underbrace{E(Y_{i0} | X_i, T_i = 1) - E(Y_{i0} | X_i, T_i = 0)}_{B^{TT}}$$

- ✦ Cette différence (ΔE^{TT}) entre participants et non participants a deux explications
 - **l'effet spécifique du programme** ou du traitement : (Δ^{TT})
 - **l'effet population (B^{TT}) qui résulte des différences préexistantes entre les niveaux de santé initiaux (Y_{i0})** de ceux qui ont participé à l'expérimentation ($T=1$), et de ceux qui n'y ont pas participé ($T=0$), Les deux populations ne sont pas identiques : (B^{TT}) mesure le biais de sélection
- ✦ Ceux qui oublient le biais de sélection, attribuent à tort, l'intégralité de l'effet total (ΔE^{TT}) à la politique mise en œuvre (Δ^{TT}) alors qu'une partie ne lui est aucunement imputable. Si on oublie de soustraire le biais de sélection (B^{TT}), on surestime « la quantité d'effet »

Quantité d'Effet vs Association

- ✦ Les populations des groupes ainsi constitués sont elles comparables ?
- ✦ OUI par randomisation → Quantité d'effet mesurée par La réduction des risques absolus (RRA), La réduction risques relatifs (RRR), odds ratio (OR) etc...
- ✦ NON si le groupe de test et le groupe de contrôle ne sont pas au départ des substituts parfaits. **La situation moyenne des individus qui ont reçu le traitement n'aurait pas été la même en l'absence de traitement que celle des individus n'ayant pas reçu le traitement → association**

Force des Evaluations Expérimentales

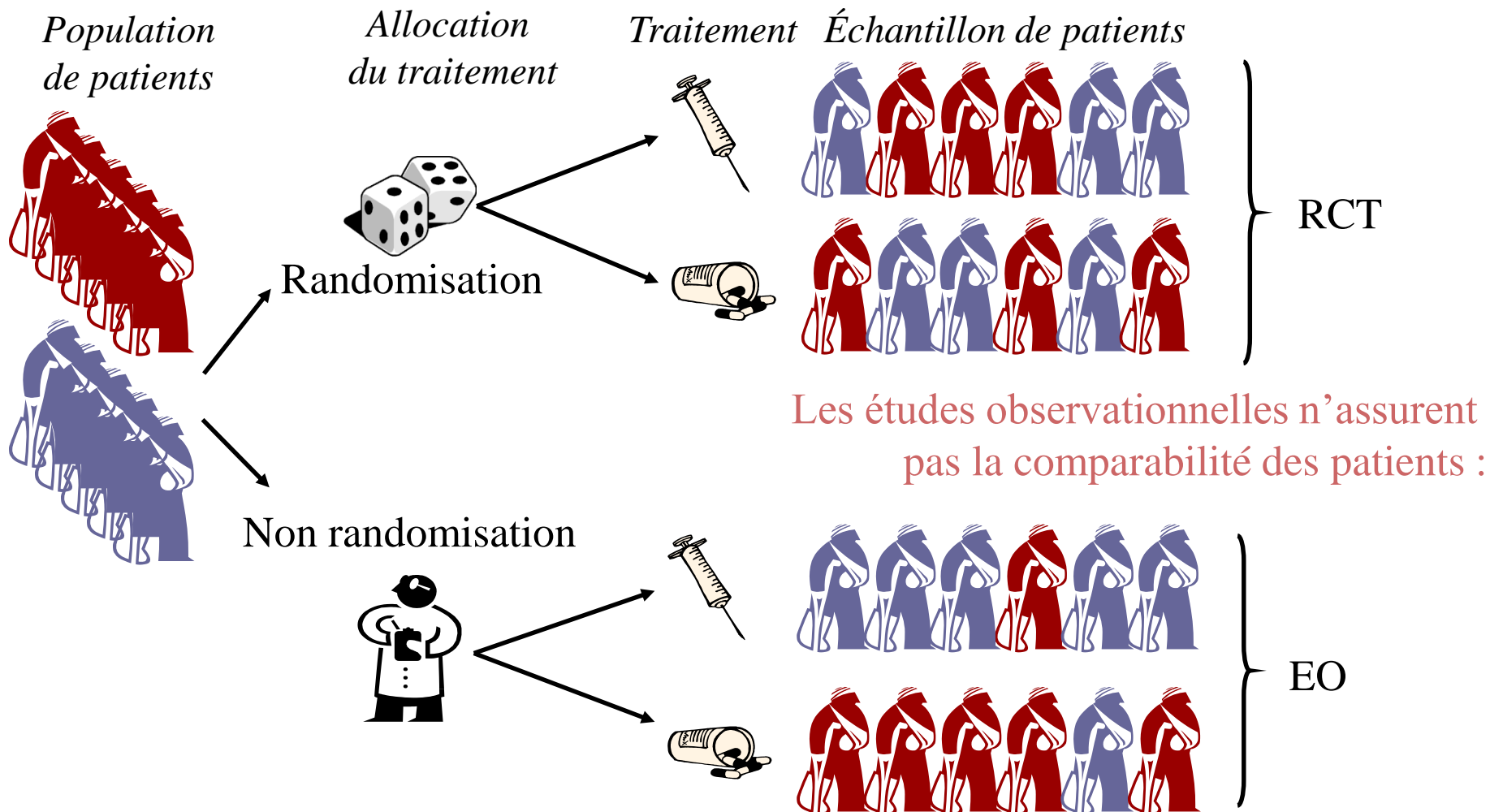
- ✦ Dans les essais randomisés en double insu et avec groupe contrôle, trois conditions permettent d'établir une relation causale:
 - **L'affectation aléatoire garantit la comparabilité des populations** entre les deux groupes, le groupe verum et le groupe placebo, en neutralisant les caractéristiques personnelles observables et non observables des patients inclus dans chacun d'entre eux,
 - **Le placebo garantit la comparabilité des effets** des traitements puisque chacun d'entre eux est confronté à la même situation de référence .
 - **le double insu garantit la comparabilité des comportements** : les effets blouse blanche, les effets patient, les effets structure sont neutralisés.
- ✦ Dans ces conditions les différences qui apparaissent entre les deux groupes, « *ceteris paribus* », sont exclusivement dues au traitement. **Le rapport de cotes ne mesure pas une association mais exprime réellement une quantité d'effet**, le lien de cause à effet est démontré.

LES BIAIS : UN RISQUE INHÉRENT À LA NATURE MÊME DES ETUDES OBSERVATIONNELLES

Evaluations Non expérimentales, Biais et Facteurs de confusion

- ✦ Un biais, c'est une *erreur systématique* qui introduit un écart entre la valeur estimée des variables d'intérêt et leur vraie valeur, ce qui entraîne une interprétation erronée du résultat.
- ✦ Un facteur de confusion est un *mélange d'effets* (facteur tiers et facteur d'exposition) qui biaise l'interprétation des résultats de l'étude parce que *l'association simultanée du facteur tiers*
 - avec l'exposition (i.e le programme, le traitement)
 - et avec la survenue de l'événementfait croire faussement que le facteur d'exposition est le fait générateur de la survenue de l'événement. Ex cheveux gris /infarctus /age

Le Biais de Recrutement



Biais de recrutement : les patients sont sélectionnés selon certaines caractéristiques corrélées à leur devenir.

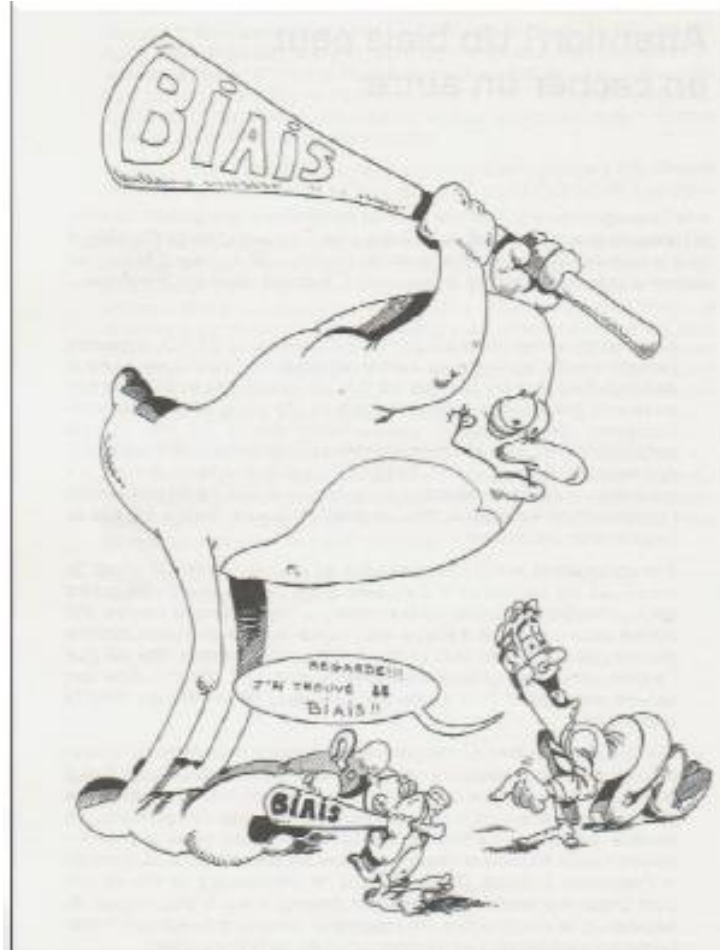
Le Biais d'Indication

- ✦ Les différences observées dans la survenue des événements peuvent ne pas être liées au choix thérapeutique mais aux raisons pour lesquelles ce choix a été fait
- ✦ Les traitements ne sont pas prescrits au hasard
 - l'indication du médecin est posée en raison des caractéristiques **observables** cliniques, fonctionnelles ou comportementales des patients. Or ces paramètres sont autant de facteurs pronostics de l'évolutivité de la maladie
 - l'indication est sélectionnée par le médecin sur la base de facteurs de risque subtils (par ex le désir d'apporter une réponse à une souffrance psychologique ou sociale) qui sont autant de facteurs parasites **inobservables**
- Biais d'indication: il convient de **distinguer et d'isoler l'effet des caractéristiques observables ou inobservables** des sujets de l'effet du traitement sur le critère de jugement

Un Impératif : Travailler sur des Populations Comparables

- ✦ On peut légitimement imaginer dans le cadre d'une étude observationnelle que ceux qui bénéficient d'une prise en charge spécifique
 - présentent des caractéristiques particulières
 - qui sont autant de facteurs de risque,
 - Et que ces caractéristiques auraient joué en leur défaveur,
 - même s'ils n'avaient pas reçu le traitement.
- ✦ Si ce traitement s'avère néfaste, on doit s'attendre à ce qu'ils subissent de son fait un préjudice **encore plus grand** que celui qui eut été le leur si ce traitement ne leur avait pas été prescrit → Le **VRAI** préjudice **de santé est surestimé**.
- ✦ La tâche de l'évaluateur est donc de reconstituer un groupe contrôle à partir du groupe socialement constitué, en **sélectionnant des individus ressemblants autant que possible** à ceux du groupe auxquels le traitement a été administré.

La Chasse aux Biais

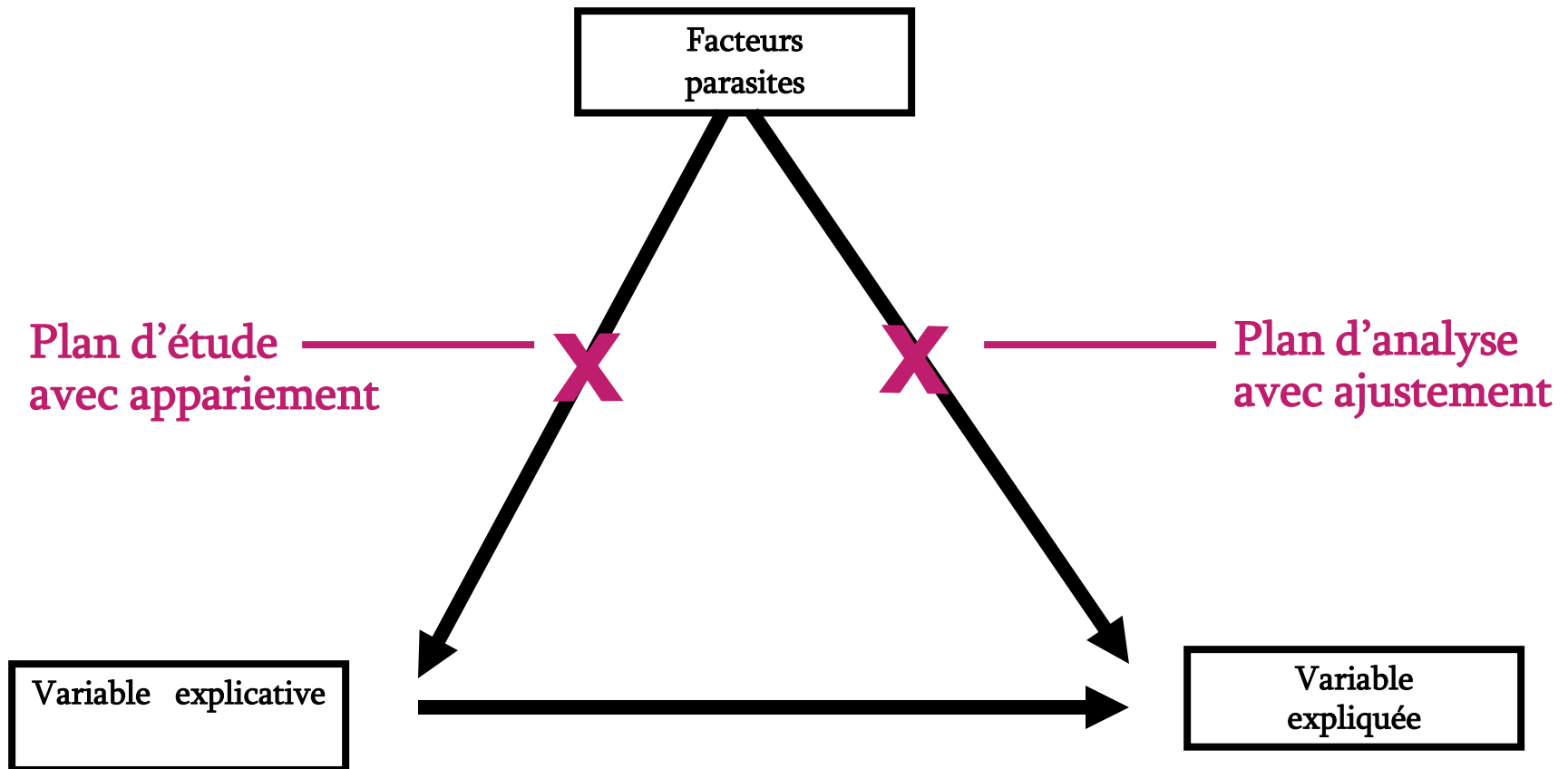


APPRENDRE A GERER LES INTERFERENCES PARASITAIRES

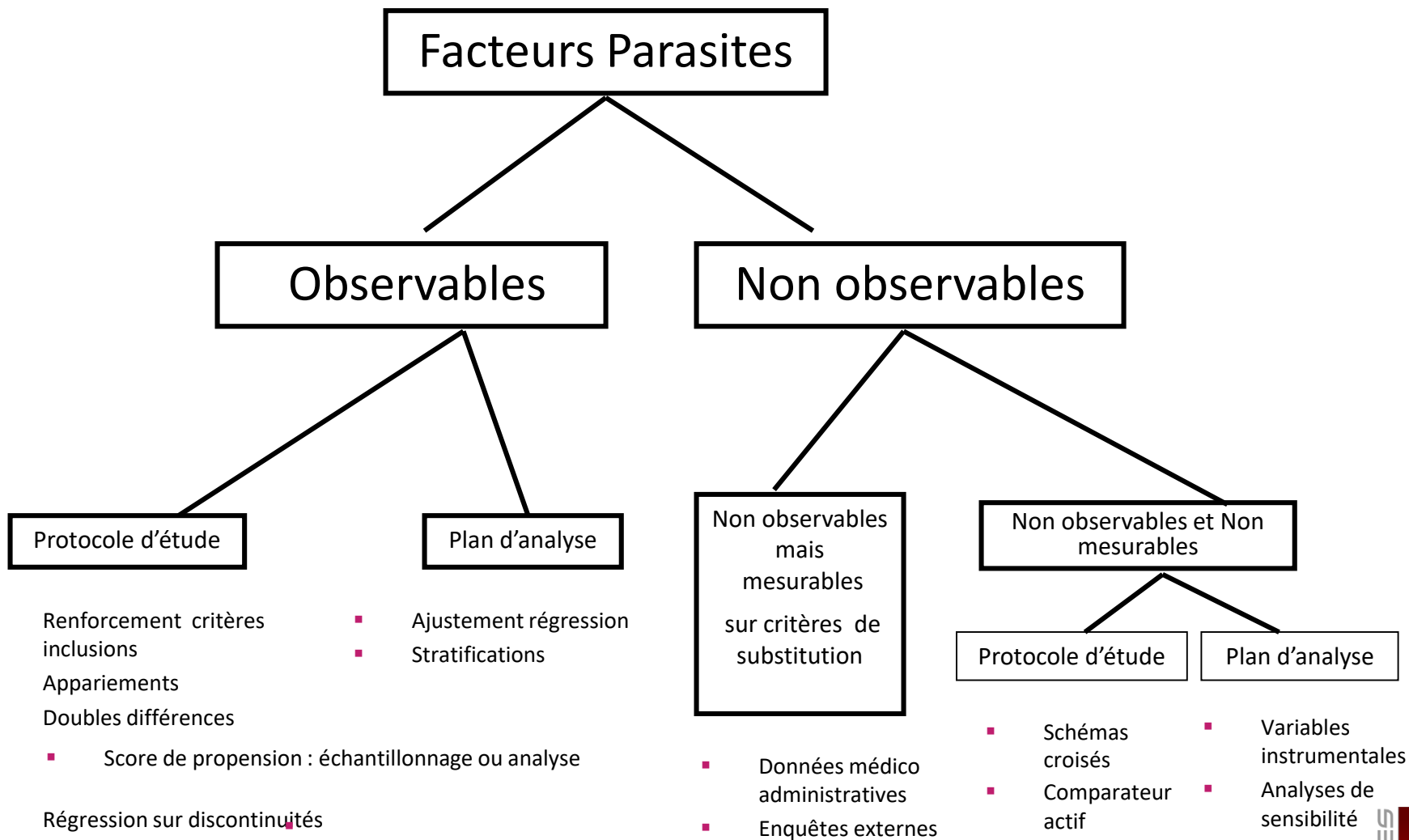
Deux Approches Pour Circonscrire et Exclure « les Autres Causes Possibles du Dommage »(les Biais)

- ✦ Modéliser *a priori* la relation entre les caractéristiques observables du patient et l'exposition à un traitement (score de propension)
- ✦ Modéliser *a posteriori* la relation entre les caractéristiques observables du patient et la survenue observée des événements (analyse multivariée)

Appariement *a priori* ou Ajustement *a posteriori*



Une Grande Diversité d'Outils



L'Appariement Exhaustif Sur Les Caractéristiques Observables Est Impossible

- ✦ Pour les caractéristiques de chaque patient traité,
- ✦ Il faudrait trouver un patient non traité avec des caractéristiques identiques
- ✦ Si l'on suppose :
 - 10 tranches d'âge,
 - Sexe : masculin/féminin,
 - Antécédents diagnostiques: 2 modalités (oui/non), 5 antécédents,
 - Antécédents thérapeutiques: 2 modalités (oui/non), 5 antécédents
 - 5 niveaux de coût
- $10 * 2 * 2^5 * 2^5 * 5 = 102\ 400$ appariements possibles

Appariement sur la Base d'un Résumé Unidimensionnel des Caractéristiques : le Score de Propension

■ Définition :

- SP : Probabilité pour un sujet de bénéficier du traitement étudié, quel que soit le groupe dont il relève, en fonction des caractéristiques initiales qu'il présente = $P(\text{traitement} \mid \text{caractéristiques du patient})$
- Si le score $SP_{\text{Traitement}} = \text{le score } SP_{\text{Contrôle}} \rightarrow$ Les sujets traités et les contrôles auront en moyenne les mêmes caractéristiques initiales

■ Usage :

- **Protocole d'étude** avec **appariement** sur le SP
 - Contrôle du biais de recrutement
 - Moins de patients si SP trop dissemblables \rightarrow perte de puissance
 - **Analyse** avec **ajustement** sur le SP
 - Moindre impact sur la taille de l'échantillon
 - Une seule variable d'ajustement
 - **Analyse** avec **stratification** sur le SP
- Le SP **simule un essai randomisé** en assurant la comparabilité des patients sur les variables **observées**.

Opérationnalisation de la Méthode SP

1. Rassembler le plus grand nombre d'informations possibles sur les caractéristiques cliniques initiales des patients;
2. Modéliser la probabilité que les malades soient amenés à recevoir le traitement en fonction des caractéristiques initiales qu'ils présentent à l'aide du Score de Propension (SP);

– SP = P(Traitement | Caractéristiques initiales)

– A l'aide d'une fonction logistique :

$$\ln\left(\frac{PS}{1-PS}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

3. Les groupes de patients dont les scores de propension sont similaires sont considérés comparables.
- ➔ Le SP **simule un essai randomisé** en assurant la comparabilité des patients sur les variables **observées**.

Procédure d'Appariement



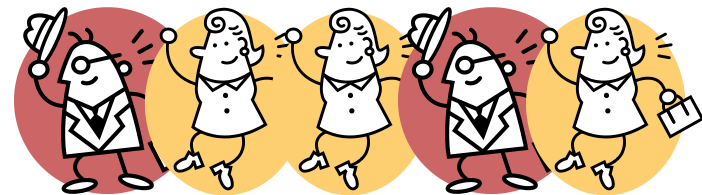
$$SP_H = P(T|Homme) = 4/6 = 0.66$$

$$SP_F = P(T|Femme) = 1/4 = 0.25$$

	T	C	
H	4	2	6
F	1	3	4

Groupe test : T

Groupe contrôle : C



SP = 0.25

SP = 0.66

SP = 0.66

Appariement sur le SP



Mesurer la “Quantité” de Biais Avant et Après Appariement par le SP

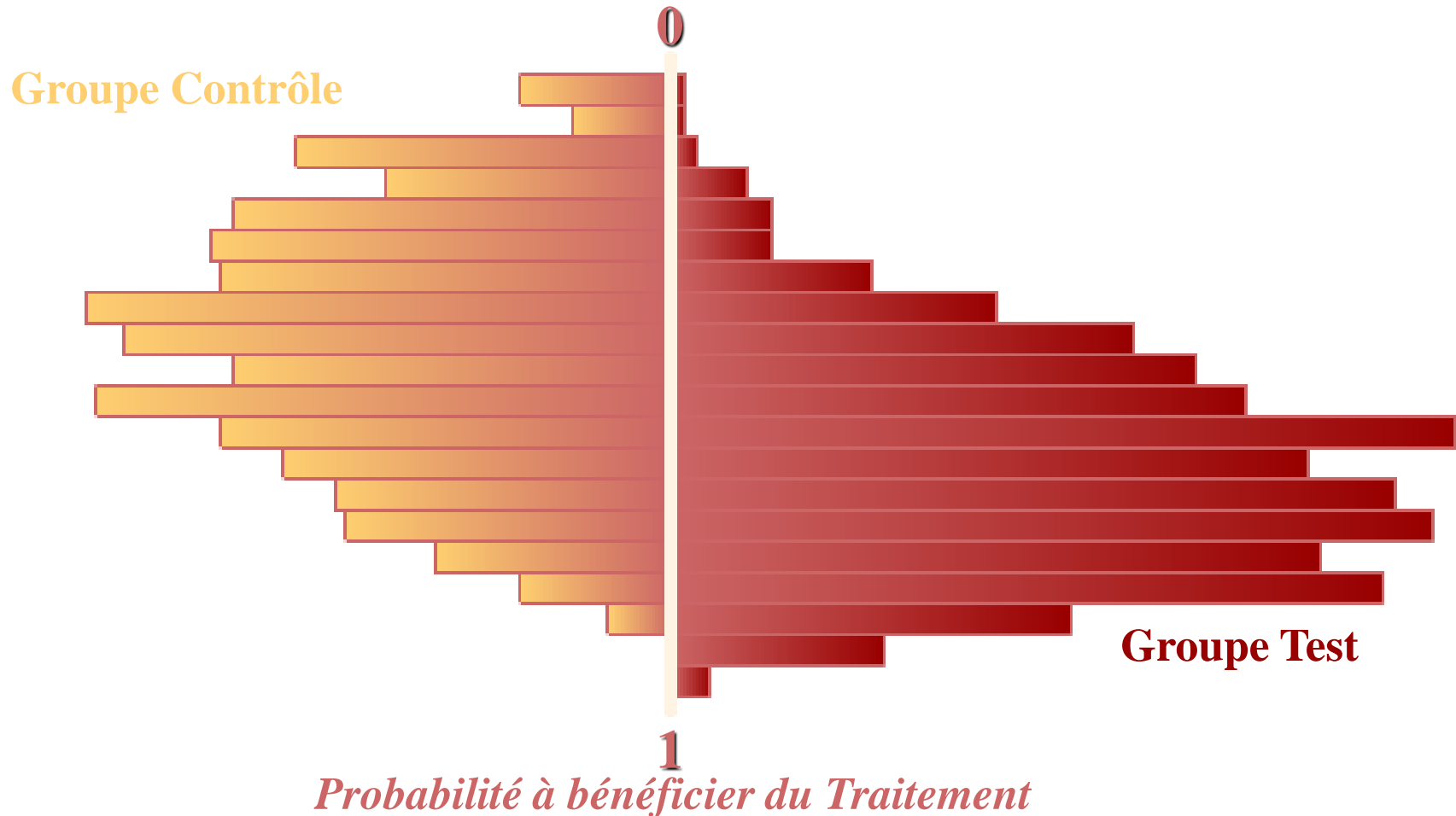
- ✦ Inconvénients de la p-value χ tests (pour des variables dichotomiques) où t tests pour les variables continues :
 - Majore l’importance du déséquilibre en fonction de la taille de l’échantillon. La diff. peut être statistiquement significative avec grd échantillon sans l’être cliniquement
- ✦ → Différences standardisées :

$$d = \frac{(x_{\text{traitement}} - x_{\text{contrôle}})}{\sqrt{\frac{S_{\text{traitement}}^2 + S_{\text{contrôle}}^2}{2}}}$$

- ✦ $|d| > 10\%$ → déséquilibre entre les deux bras

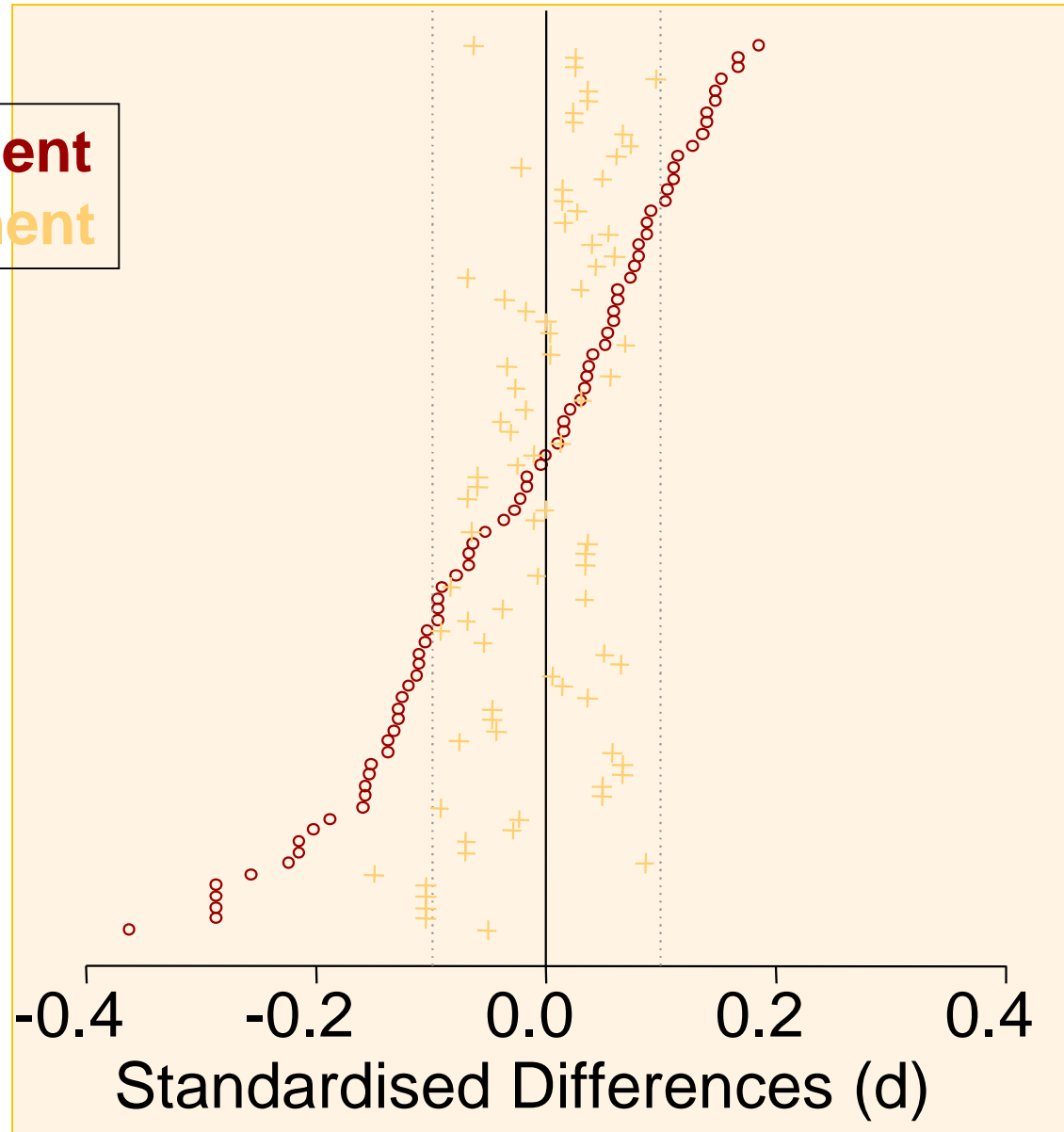
Biais de Recrutement

Distribution du SP dans le groupe contrôle et le groupe test



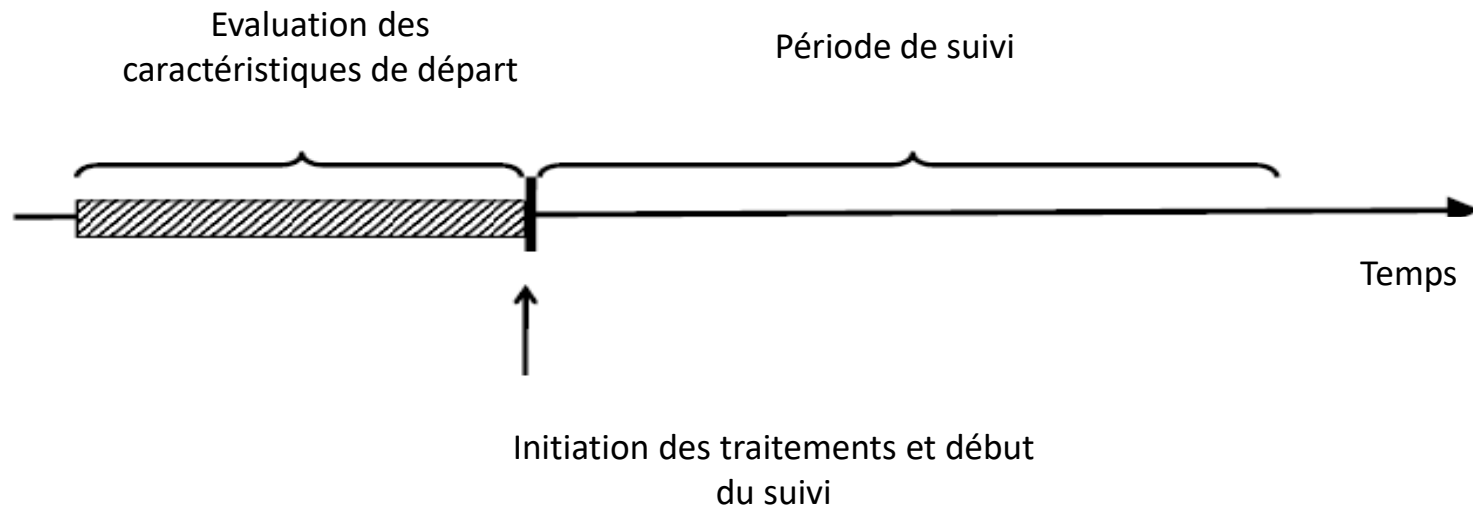
L'Appariement sur le PS Réduit les Biais

o Avant appariement
+ Après appariement



FAIRE PARLER LES BANQUES DE DONNEES MEDICO ADMINISTRATIVES

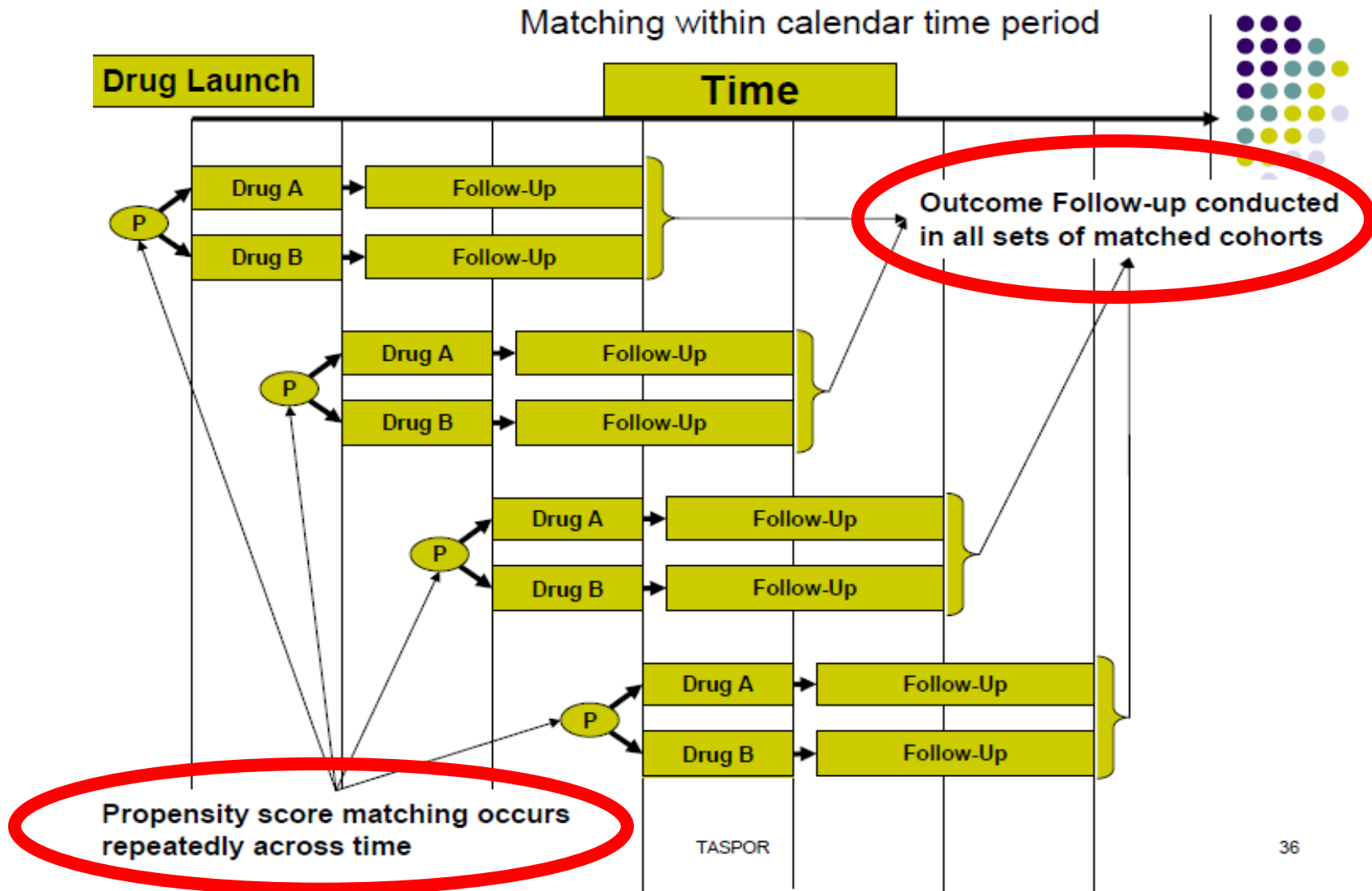
Schéma Simplifié de Cohorte pour l'Analyse Longitudinale d'une BDMA*



32

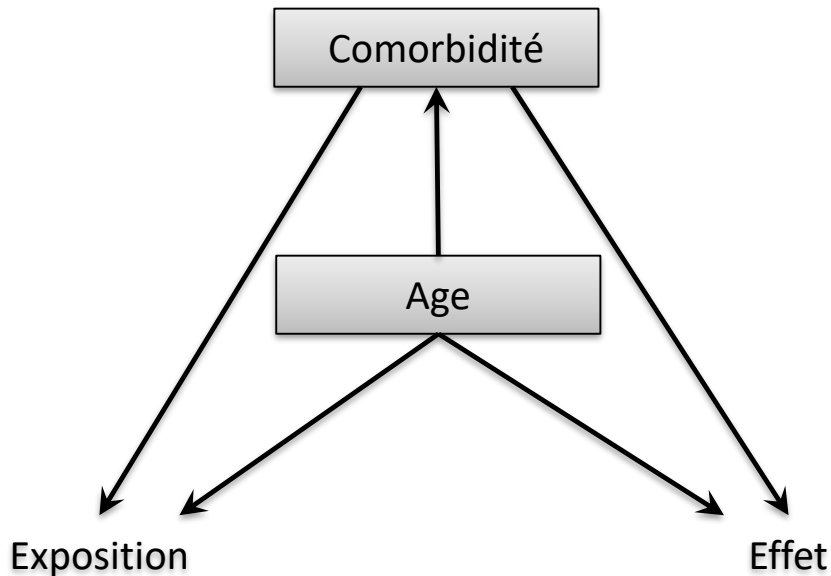
28

Technique Longitudinale d'Appariement



Identification des Critères de Substitution dans les BDMA

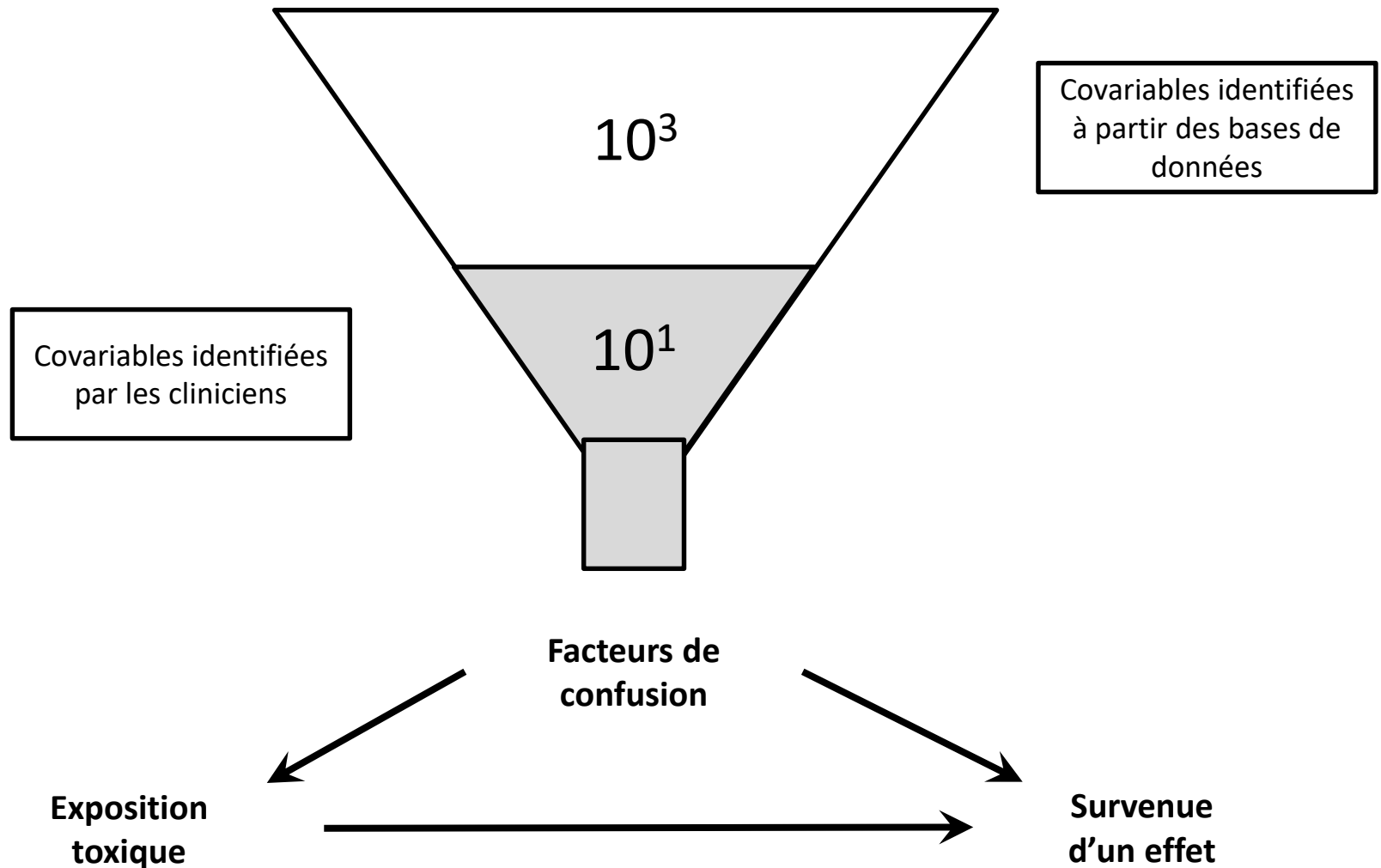
- Les facteurs de confusion identifiés à travers les dépenses reconnues et remboursées sont autant de critères d'évaluation de l'état de santé



Plus il y a de critères de substitution mieux c'est

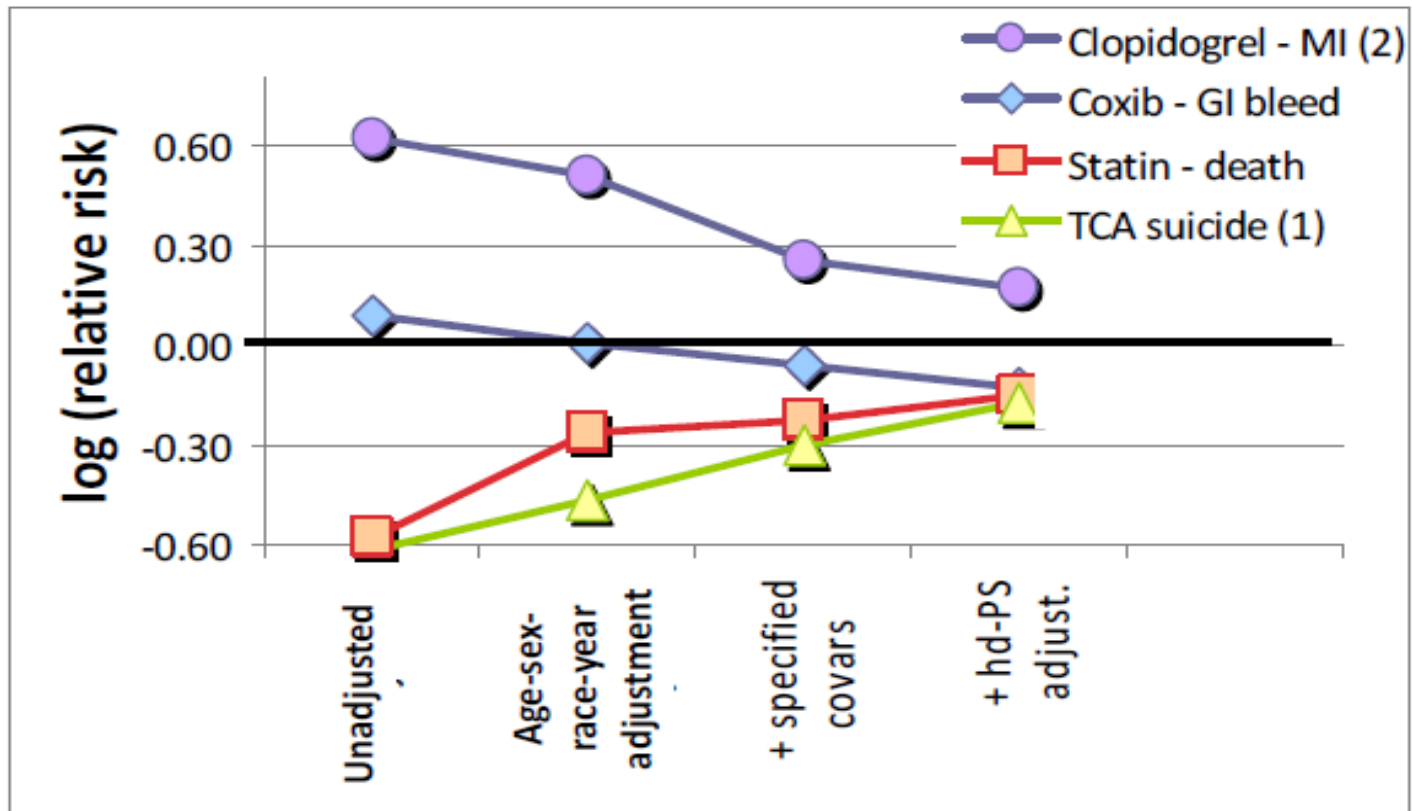
Critère de substitution	Etat de santé
Utilisation de bouteilles d'oxygène	Santé très fragile
Diagnostic d'hypertension artérielle durant un séjour hospitalier	Pas trop mauvaise santé
Check-up annuel et coloscopie	Très observant
1 ^{ère} statine à 70 ans	Plutôt en bonne santé
Polymédication, nombreuses consultations	Plutôt en mauvaise santé

Proportion de l'Origine des Covariables



Techniques d'Ajustement et Estimation de la Quantité d'Effet Vraie

Performance of different adjustment procedures, including hd-PS adjustment



Rassen *et al.* Circulation 2009, 120 :2322-2329

Coronary Heart Disease

Cardiovascular Outcomes and Mortality in Patients Using Clopidogrel With Proton Pump Inhibitors After Percutaneous Coronary Intervention or Acute Coronary Syndrome

Jeremy A. Rassen, ScD; Niteesh K. Choudhry, MD, PhD;
Jerry Avorn, MD; Sebastian Schneeweiss, MD, ScD

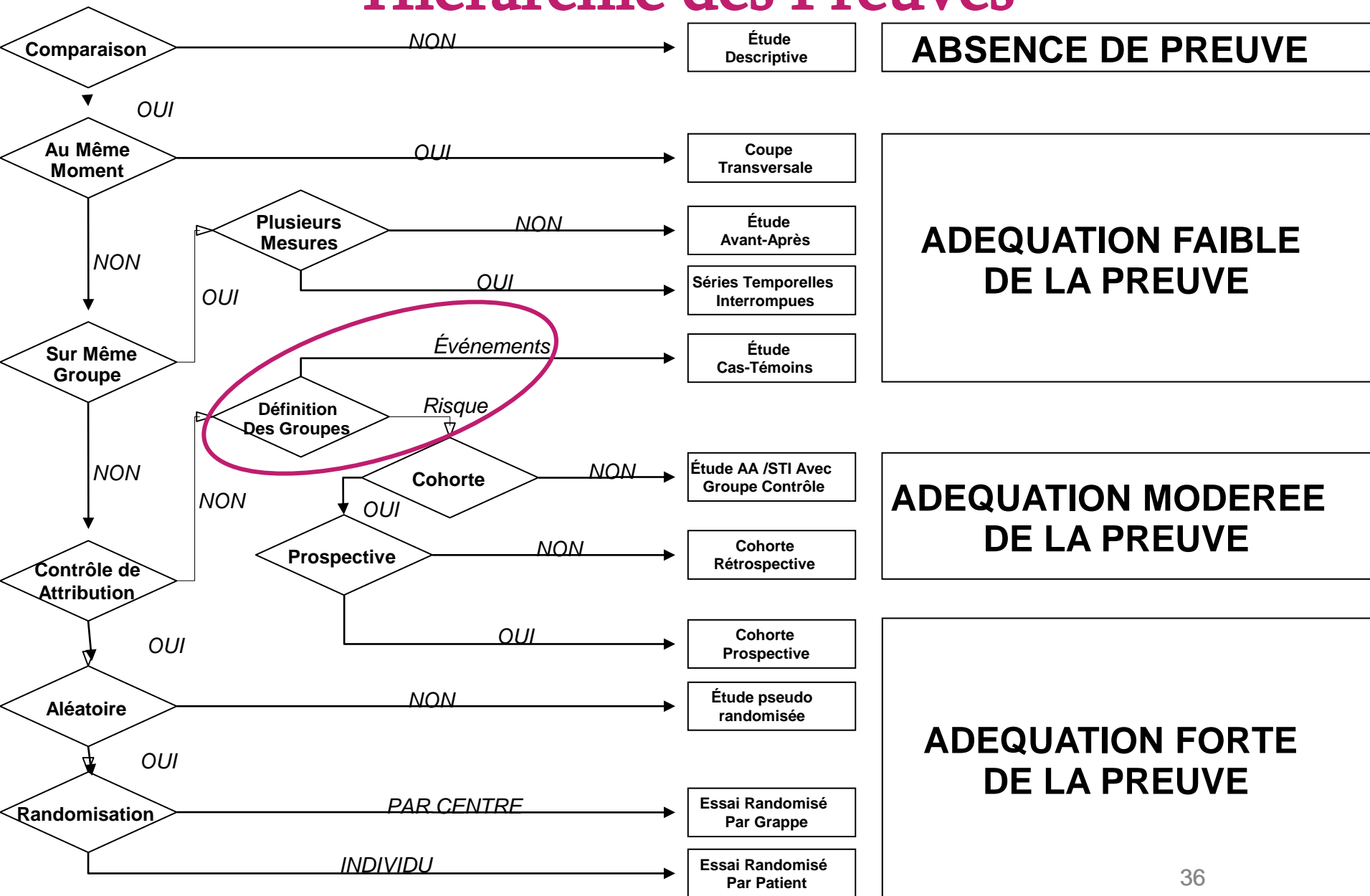
Outcome	British Columbia (n=10 391)	RR (95% CI)			P for Heterogeneity
		Pennsylvania (n=4176)	New Jersey (n=3998)	Pooled* (n=18 565)	
Primary end point					
MI hospitalization or death					
Unadjusted	1.90 (1.47–2.46)	1.96 (1.34–2.88)	1.26 (0.85–1.88)	1.74 (1.44–2.10)	0.19
Age/sex-adjusted	1.69 (1.30–2.20)	2.01 (1.37–2.96)	1.18 (0.79–1.75)	1.62 (1.34–1.96)	0.15
Multivariate-adjusted	1.33 (1.01–1.75)	1.71 (1.14–2.56)	0.97 (0.64–1.47)	1.32 (1.08–1.61)	0.16
PS-adjusted†	1.22 (0.92–1.61)	1.69 (1.09–2.63)	1.08 (0.70–1.68)	1.22 (0.99–1.51)	0.31

LE CHOIX DU BON SCHEMA D'ETUDE, UN PROCESSUS CONTINU

Cause de l'Effet Ou Effet de la Cause ?

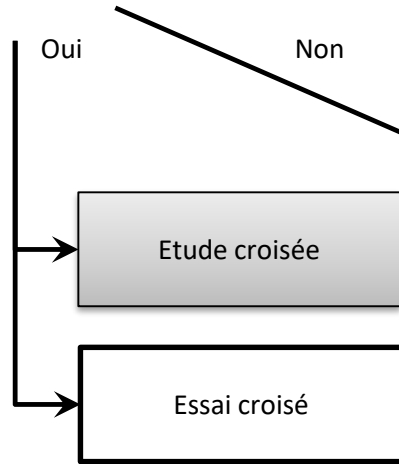
- ✦ **L'évaluation structurelle** « fait un lien précis **ex ante** entre le modèle formel où les comportements fondamentaux des agents sont modélisés et le modèle statistique estimé pour prédire l'impact de programme en projet »
- ✦ **L'évaluation réduite** « étudie **ex post** l'effet de deux variables d'intérêt (« *Les résultats potentiels* ») pour un individu : celle où il bénéficie du TrT et celle où il n'en bénéficie pas, sans détailler les mécanismes intermédiaires, ni estimer les paramètres fondamentaux »

Choix du Schéma d'Etude en Fonction de la Hiérarchie des Preuves

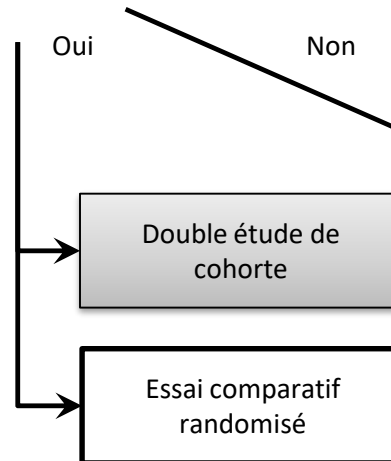


Choix du Schéma d'Etude en Fonction de la Variabilité des Grandeurs Mesurées

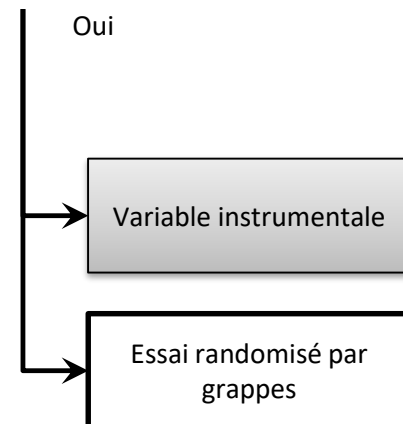
Variabilité Intra-individuelle



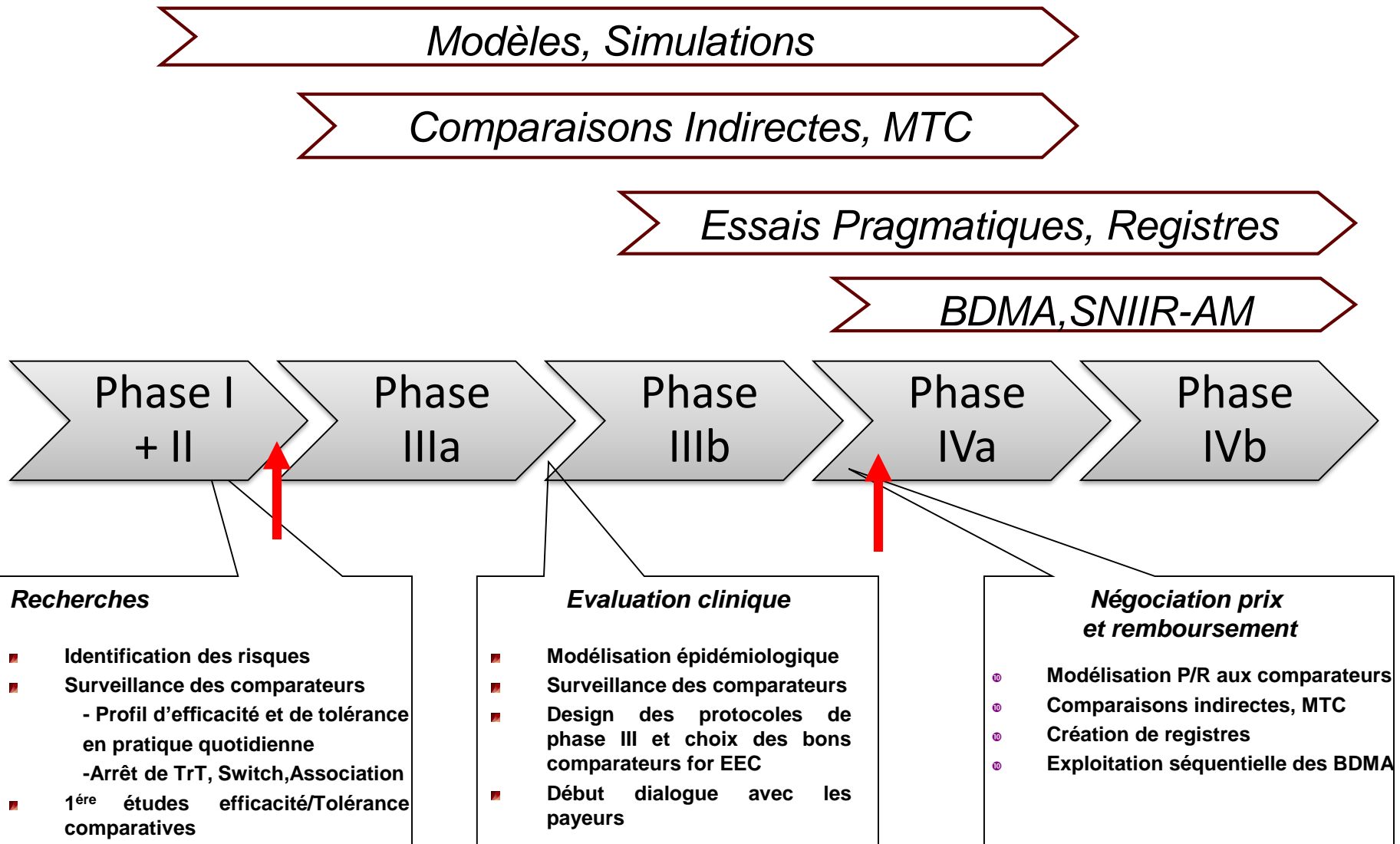
Variabilité Inter-individuelle



Variabilité Inter-structure



Choix du Schéma d'Etude Selon la Phase de Développement dans Laquelle se Trouve le Produit



CONCLUSION: ACCEPTER LE VERDICT DES FAITS

- ✦ L'évaluation des politiques publiques constitue une « valeur montante » en France
 - Création d'indicateurs dans le cadre de la LOFT (2001)
 - Possibilités d'expérimentations inscrite dans la Constitution depuis la loi du 28 mars 2003
 - Lancement des expérimentations-évaluations ex RSA
 - ✦ Ces méthodes d'expérimentation sociale et d'évaluation aléatoire ou non porte une ambition forte:
 - « *Identifier les politiques efficaces et celles qui ne le sont pas* »
- Esther Dufflot*

Sélection Bibliographique

- Austin P. An introduction to propensity-score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*. 2011;46(3):399-424.
- Behaghel L. « Lire l'économétrie » Collection repères. La découverte. Paris 2012
- Brodaty T, Crépon B, Fougère D. Les méthodes micro-économétriques d'évaluation et leurs applications aux politiques actives de l'emploi. *Economie et prévision*. 2007;177(1):93-118.
- Cling J, Razafindrakoto M, Roubaud F. L'évaluation d'impact des politiques publiques : enjeux, méthodes, résultats. Paper presented at: Les Journées de Tam Dao 2008.
- Crépon B, « L'apport des expérimentations dans l'évaluation de l'impact des dispositifs publics », *Informations sociales*, 2008/6 (150) : 56-67.
- Duflo E. L'approche expérimentale en économie du développement. *Revue d'économie politique*. 2009:691-726.
- Fougère D; Expérimenter pour évaluer les politiques d'aide à l'emploi : les exemples anglo-saxons et nord-européens *Revue Française des affaires sociales*;2000 :(1) : 111-144
- Fougère D. Les méthodes économétriques d'évaluation. *Revue Française des affaires sociales*. 2010;1-2:105-128,
- Givord P. Méthodes économétriques pour l'évaluation de politiques publiques. Paris: INSEE;2010.
- Heckman J. Building bridges between structural and program evaluation approaches to evaluating policy. Cambridge: National Bureau of economic research;2010.
- Kurth T, Sonis J. Assessment and control of confounding in trauma research. *J Trauma Stress*. Oct 2007;20(5):807-820.
- Kurth T, Walker AM, Glynn RJ, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol*. Feb 1 2006;163(3):262-270.
- Little RJ, Rubin DB. Causal effect in clinical and epidemiological studies via potential outcomes : concepts and analytical approaches. *Annu. Rev. Public Health*. 2000;21:121-145.
- Mamdani M., Sykora K, Ping Li, T Normand SL., & al. Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding *BMJ* 2005;330:960-2
- Payet S, Riou França L, Launois R, Le Lay K, Vallet B, Dhainaut JF, et le groupe PREMISS « Evaluation coût-efficacité de la drotrecogine alfa comparée à la prise en charge conventionnelle dans le traitement du sepsis sévère en pratique réelle » *Journal d'Economie Médicale* 2007, Vol. 25, n° 2 : 75-91
- Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol*. Dec 2009;62(12):1226-1232.
- Rassen JA, Choudhry NK, Avorn J, Schneeweiss S. Cardiovascular outcomes and mortality in patients using clopidogrel with proton pump inhibitors after percutaneous coronary intervention or acute coronary syndrome. *Circulation*. Dec 8 2009;120(23):2322-2329.
- Rosebaum P. et Rubin D.,. Constructing a Control Group Using Multivariate Matched Sampling Methods. *American Statistician*. 1985 :, 39: 35-39.
- Riou-França L., Payet S., Le Lay K., Launois R. « Use of the propensity score method for recruitment bias reduction in observational studies/ application to the estimation of Drotrecogin Alfa's impact on intensive care units workload » 2006. *ISPOR Connections Vol.12(1)* : 8-10
- Rubin, DB. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66: 688-701.
- Schneeweiss S, Seeger JD, Maclure M, Wang PS, Avorn J, Glynn RJ. Performance of comorbidity scores to control for confounding in epidemiologic studies using claims data. *Am J Epidemiol*. Nov 1 2001;154(9):854-864.
- Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol*. Apr 2005;58(4):323-337.
- Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiol Drug Saf*. May 2006;15(5):291-303.
- Schneeweiss S, Solomon DH, Wang PS, Rassen J, Brookhart MA. Simultaneous assessment of short-term gastrointestinal benefits and cardiovascular risks of selective cyclooxygenase 2 inhibitors and nonselective nonsteroidal antiinflammatory drugs: an instrumental variable analysis. *Arthritis Rheum*. Nov 2006;54(11):3390-3398.
- Schneeweiss S. Developments in post-marketing comparative effectiveness research. *Clin Pharmacol Ther*. Aug 2007;82(2):143-156.
- Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology*. Jul 2009;20(4):512-522.
- Schneeweiss S, Gagne JJ, Glynn RJ, Ruhl M, Rassen JA. Assessing the comparative effectiveness of newly marketed medications: methodological challenges and implications for drug development. *Clin Pharmacol Ther*. Dec 2011;90(6):777-790.
- Seeger JD, Kurth T, Walker AM. Use of propensity score technique to account for exposure-related covariates: an example and lesson. *Med Care*. Oct 2007;45(10 Supl 2):S143-148.
- Stuart EA. Matching methods for causal inference: A review and a look forward. *Stat Sci*. Feb 1 2010;25(1):1-21.
- Stuart EA, Marcus SM, Horvitz-Lennon MV, Gibbons RD, Normand SL. Using Non-experimental Data to Estimate Treatment Effects. *Psychiatr Ann*. Jul 1 2009;39(7):41451.*



BACK UP