

Comprendre le score de propension

Alexandra Rouquette

PUPH Santé Publique – Biostatistique

Université Paris-Saclay, Assistance Publique Hôpitaux de Paris

Insem UMR1018 – Centre de recherche en Epidémiologie et Santé des Populations (CESP)

Déclaration de liens d'intérêt avec les industriels de santé
en rapport avec le thème de la présentation (loi du 04/03/2002) :

L'orateur ne
souhaite
pas répondre

- **Intervenant** : Rouquette Alexandra
- **Titre** : Comprendre le score de propension

- Consultant ou membre d'un conseil scientifique
- Conférencier ou auteur/rédacteur rémunéré d'articles ou documents
- Prise en charge de frais de voyage, d'hébergement ou d'inscription à des congrès ou autres manifestations
- Investigateur principal d'une recherche ou d'une étude clinique

OUI



NON

OUI



NON

OUI



NON

OUI



NON

Modèle de relation causale



Seul schéma d'étude permettant de conclure à la causalité :

essai randomisé en double aveugle bien conduit

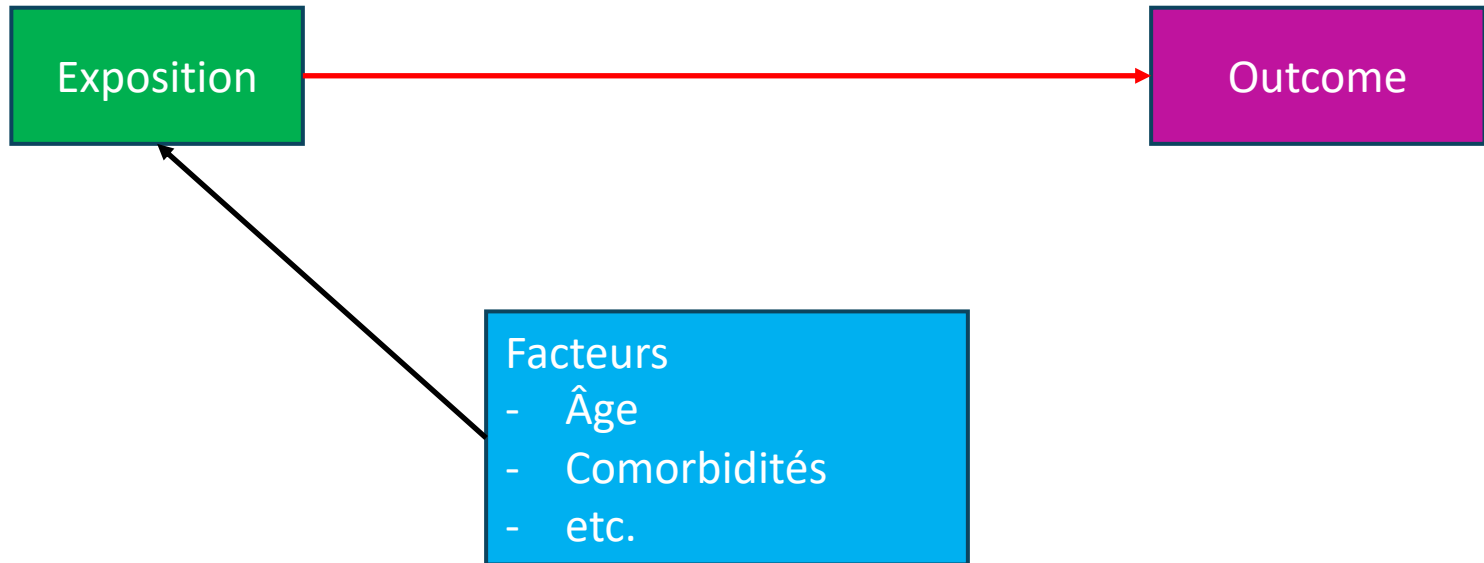
(comparaison de groupes dont la seule caractéristique qui les distingue est l'exposition)

- La **randomisation** permet d'assurer la comparabilité de ces groupes au temps initial
- Le **double aveugle** permet d'assurer le maintien de cette comparabilité jusqu'à la fin du suivi

Modèle de relation causale

Absence de randomisation : non éthique, difficilement réalisable, conditions réelles, etc.

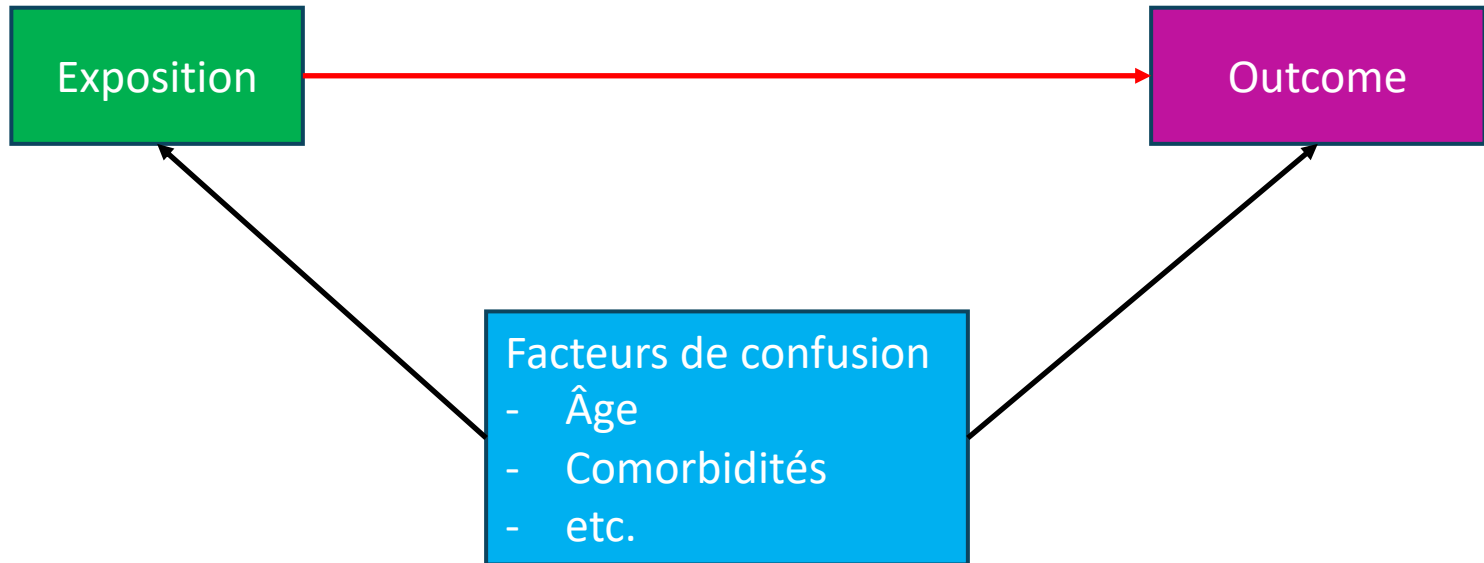
→ Données observationnelles



Modèle de relation causale

Absence de randomisation : non éthique, difficilement réalisable, conditions réelles, etc.

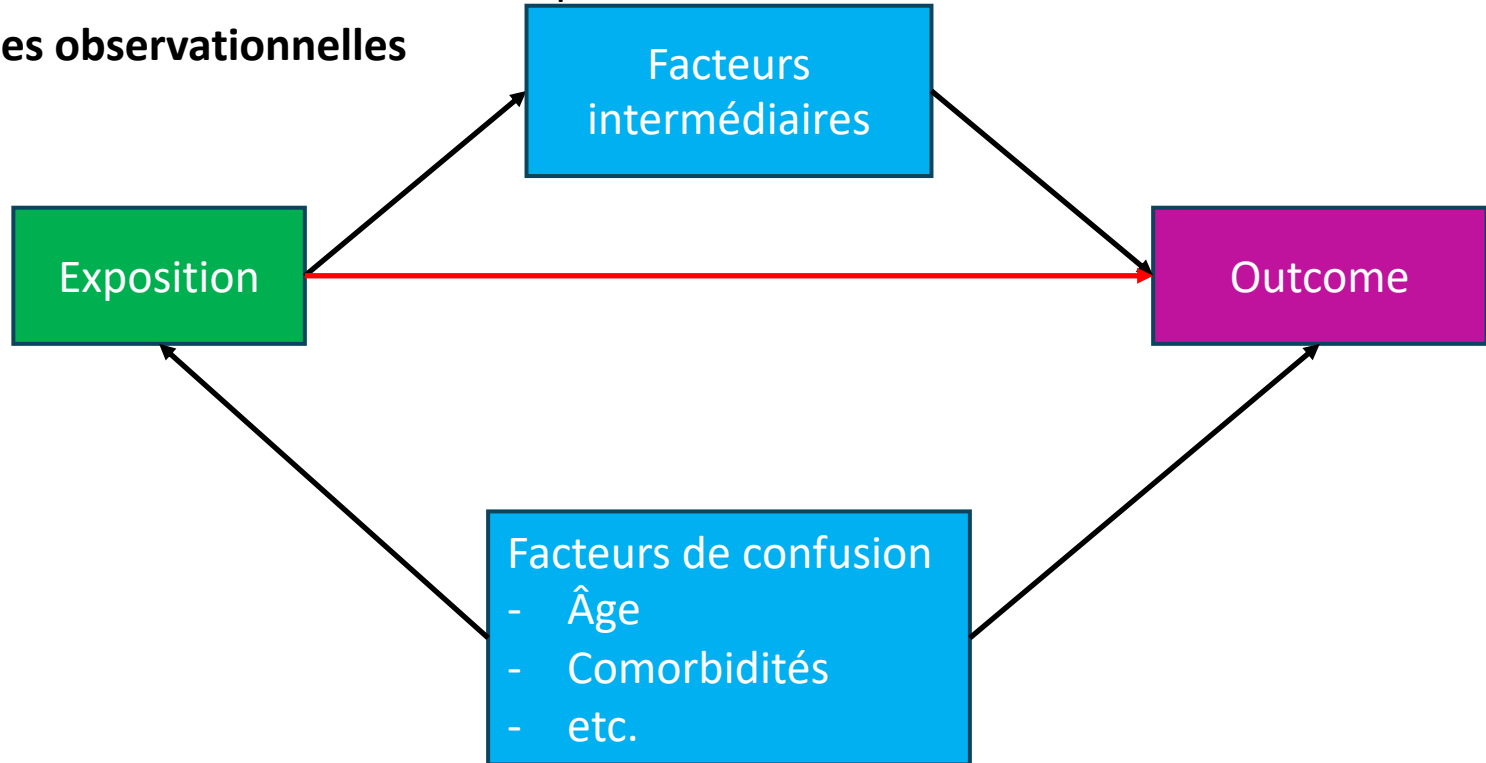
→ Données observationnelles



Modèle de relation causale

Absence de randomisation : non éthique, difficilement réalisable, conditions réelles, etc.

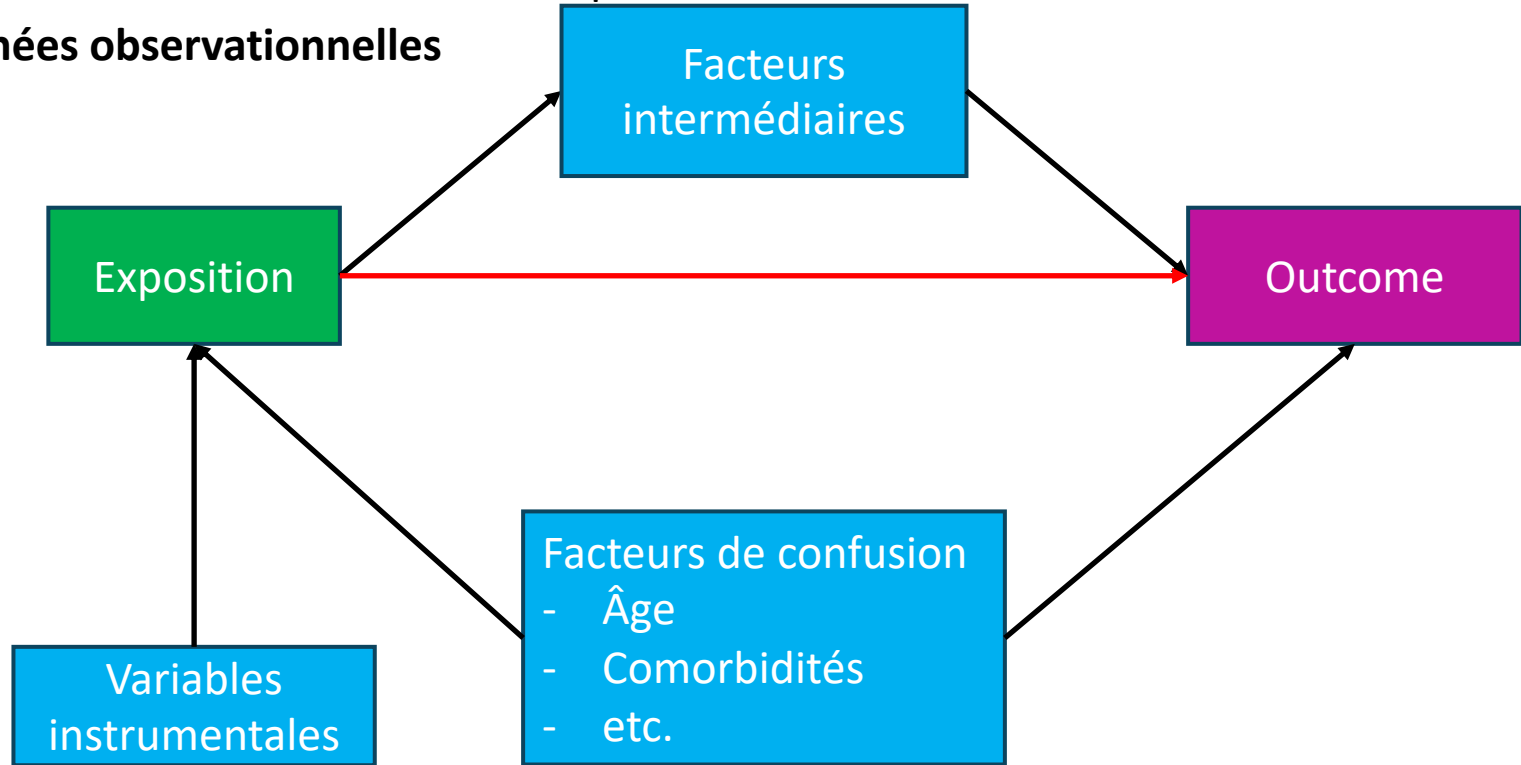
→ Données observationnelles



Modèle de relation causale

Absence de randomisation : non éthique, difficilement réalisable, conditions réelles, etc.

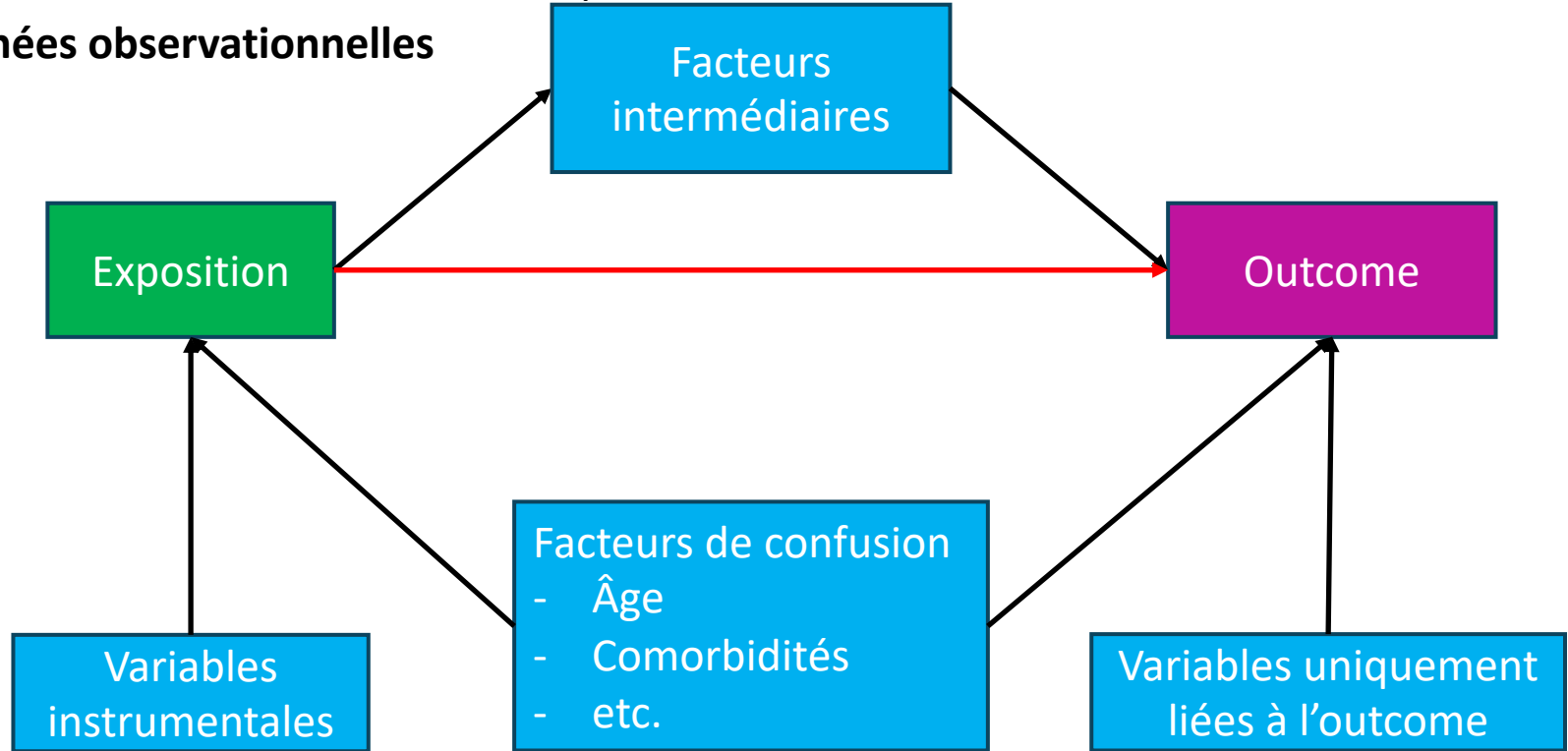
→ Données observationnelles



Modèle de relation causale

Absence de randomisation : non éthique, difficilement réalisable, conditions réelles, etc.

→ Données observationnelles

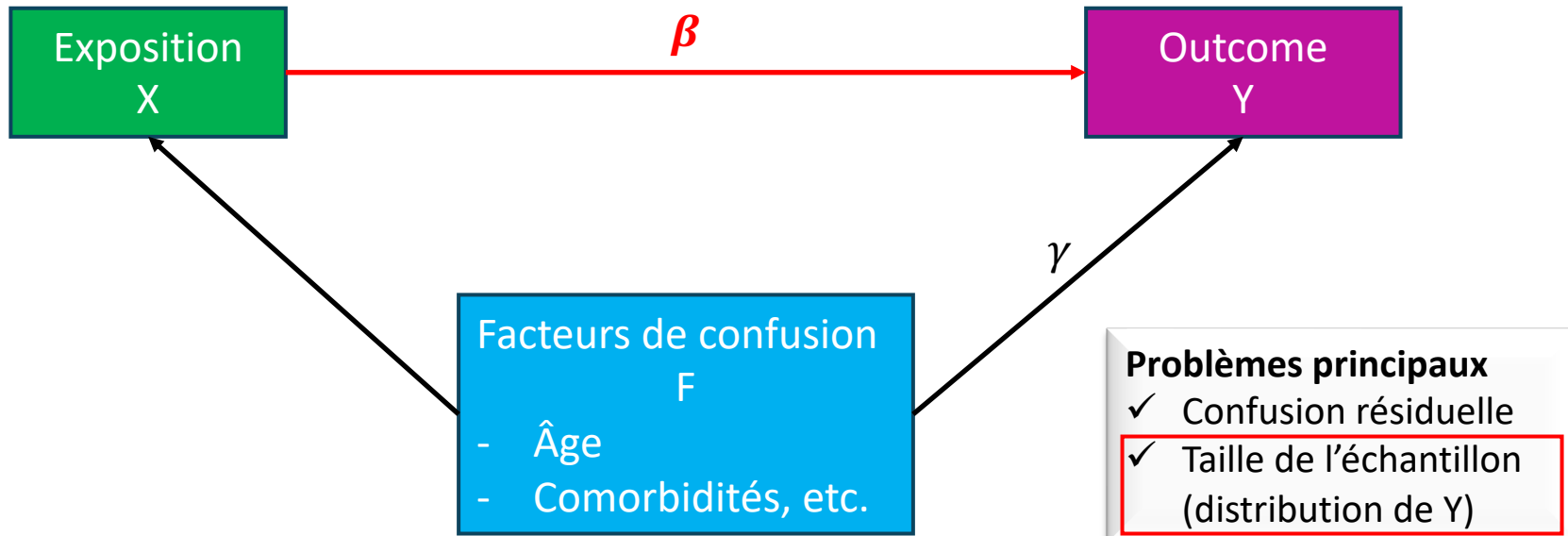


Classiquement

→ Données observationnelles

Modèle de régression avec AJUSTEMENT sur les facteurs de confusion

$$Y = \alpha + \beta X + \gamma_1 F_1 + \gamma_2 F_2 + \dots$$



Le score de propension (SdP)

❖ Rosenbaum P. et Rubin D. (1983)

- **Probabilité** d'être exposé conditionnellement aux variables observées



❖ $P(X = 1|Z)$ où

- X est une indicatrice pour identifier l'exposition
- Z est une série de covariables observées

❖ **Calculé** classiquement par régression logistique

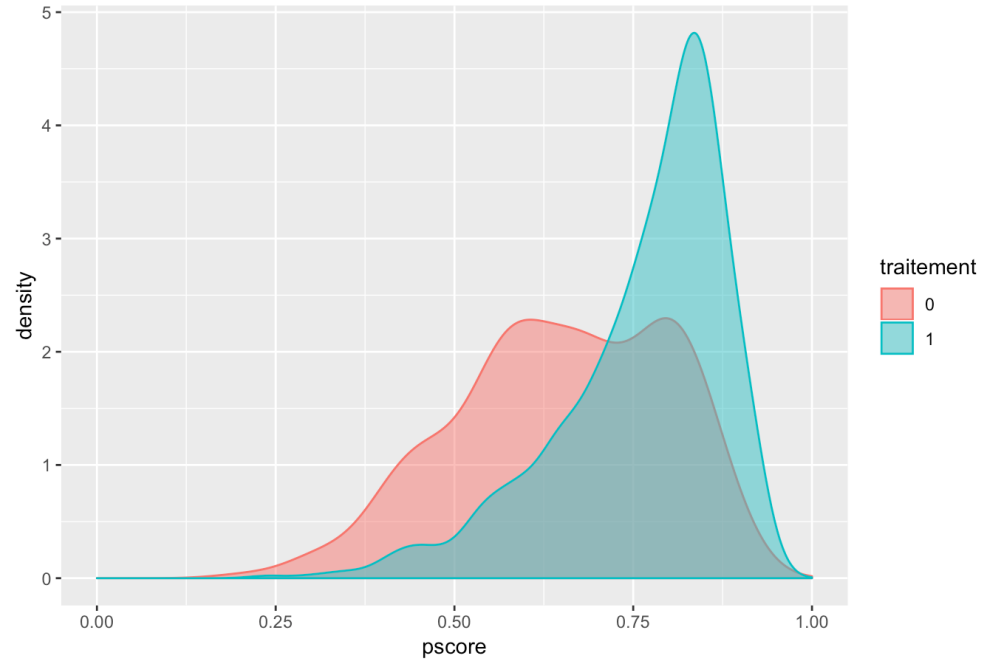
$$\text{Logit}(P(X = 1|Z_1, Z_2, \dots)) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \dots$$

• **Hypothèses**

- **d'indépendance conditionnelle** à ensemble de covariables observées
- **de positivité** : tous les sujets ont une probabilité non nulle d'être exposés
- **du support commun** : il y a suffisamment d'individus exposés (E) et non-exposés (NE) avec le même profil de covariables

Le SdP

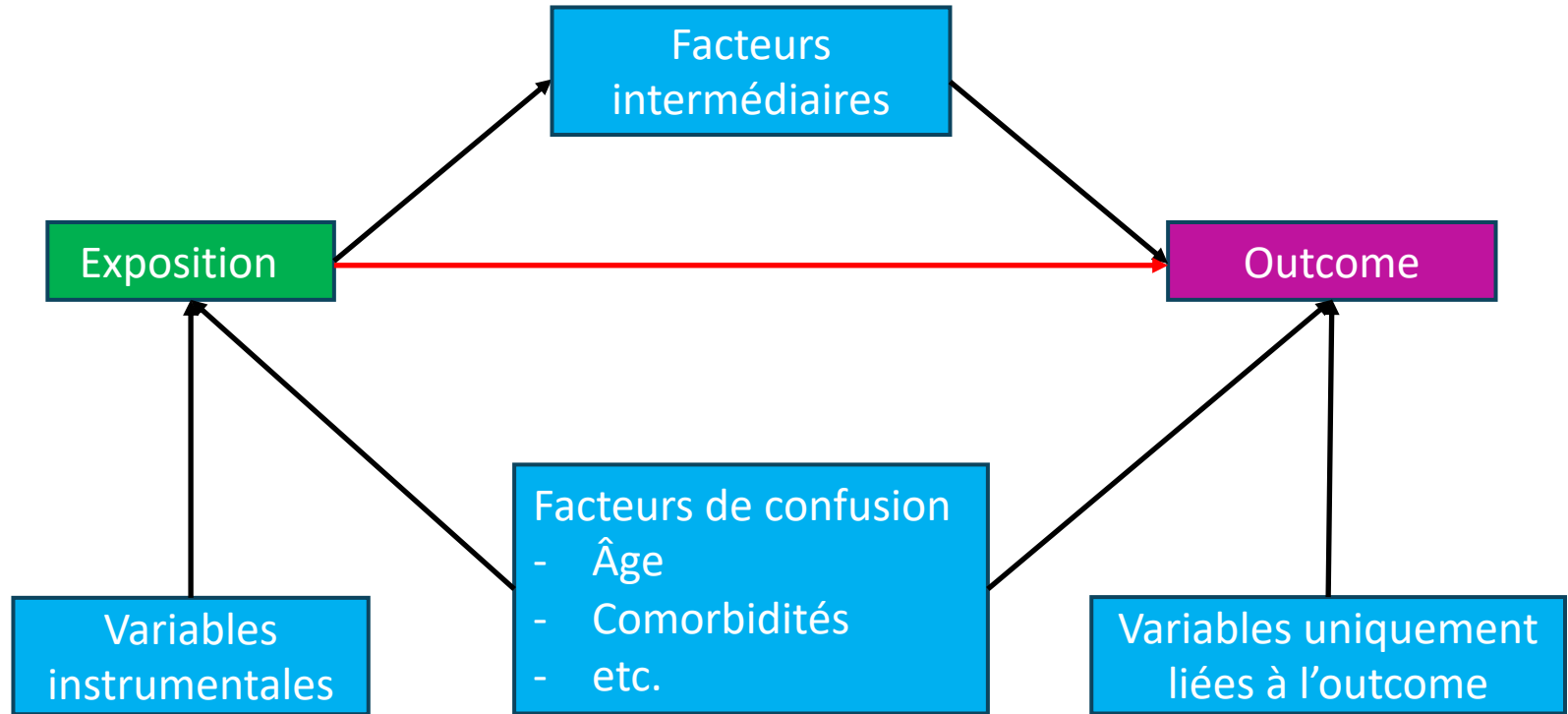
❖ Support commun (overlap)



Le SdP : 1 – Choix des variables

1 – Pas sur le chemin causal

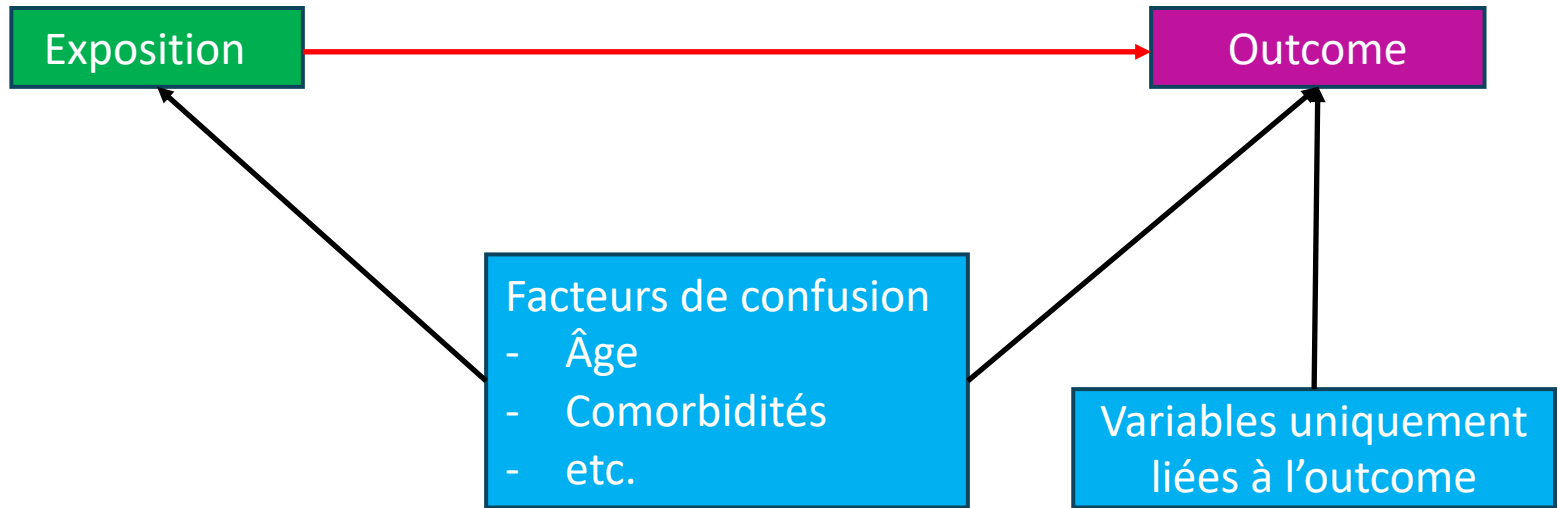
2 – Pas instrumentales (diminue la précision)



Le SdP : 1 – Choix des variables

1 – Pas sur le chemin causal

2 – Pas instrumentales (diminue la précision)



Le SdP : 2 – Principe d'utilisation

❖ Même score de propension

- même probabilité d'être exposé
- caractéristiques individuelles différentes

❖ Mais dans un échantillon de paires « exposé (E), non-exposé (NE) »

- En moyenne, toutes les variables prises en compte pour le calcul du SdP sont équilibrées entre les groupes E et NE
- ~ étude randomisée
- **MAIS variables prises en compte dans le calcul du SdP sont OBSERVEES**
=> confusion résiduelle

Le SdP : 2 – Principe d'utilisation

- ❖ Appariement (Propensity Score Matching – PSM)
 - Paires E – NE
 - => exclusion des sujets non-appariés
 - ❖ Pondération (Inverse Probability of Treatment Weighting – IPTW)
 - Pondération par
 - L'inverse de la probabilité d'être exposé ($1/SdP$) pour les E
 - L'inverse de la probabilité de ne pas être exposé ($1/(1-SdP)$) pour les NE
 - => conserve l'ensemble des sujets
 - ❖ Stratification
 - ❖ Ajustement
- } Confusion résiduelle plus importante

Le SdP : 3 – Vérification de la qualité

- ❖ Equilibre des variables utilisées pour le calcul du SdP dans les groupes E - NE
 - Pas de test statistique
 - **Différence moyenne standardisée <10% (voire 20%)**
 - ~différence de moyenne μ en unité d'écart-type s
 - Indépendante de la taille de l'échantillon et de l'unité de mesure

Variables continues :

$$d = \frac{(\bar{\mu}_E - \bar{\mu}_{NE})}{\sqrt{\frac{s_E^2 + s_{NE}^2}{2}}}$$

Variables binaires :

$$d = \frac{(p_E - p_{NE})}{\sqrt{\frac{p_E(1 - p_E) + p_{NE}(1 - p_{NE})}{2}}}$$



Original Investigation | Psychiatry

Comparison of Depression and Anxiety Following Self-reported COVID-19-Like Symptoms vs SARS-CoV-2 Seropositivity in France

Alexandra Rouquette, MD, PhD; Arthur Descarpentry, MSc; Fallou Dione, MSc; Bruno Falissard, MD, PhD; Stéphane Legleye, PhD; Cécile Vuillermoz, PhD; Anne Pastorello, MSc; Laurence Meyer, MD, PhD; Josiane Warszawski, MD, PhD; Camille Davisse-Paturet, PhD; Maria Melchior, PhD; for the EpiCoV Study Group
JAMA Network 2023; 6(5): e2312892

Conséquences à long terme de la covid-19 sur la santé mentale

Bases médico-administratives aux USA et UK

- ✓ Faibles associations significatives avec les troubles dépressifs et anxieux à 1 an
- ✓ Mais
 - Facteurs socioéconomiques et de mode de vie peu disponibles → Confusion ?
 - Infection SARS-CoV-2 pauci ou asymptomatique → Statut sérologique ?

Illustration - Objectif

❖ Etudier les associations entre

- Symptômes évocateurs de covid-19
- Statut sérologique SARS-CoV-2

} 2 expositions

❖ Et la présence 8 mois plus tard

- d'une symptomatologie dépressive élevée
- ~~d'une symptomatologie anxieuse élevée~~

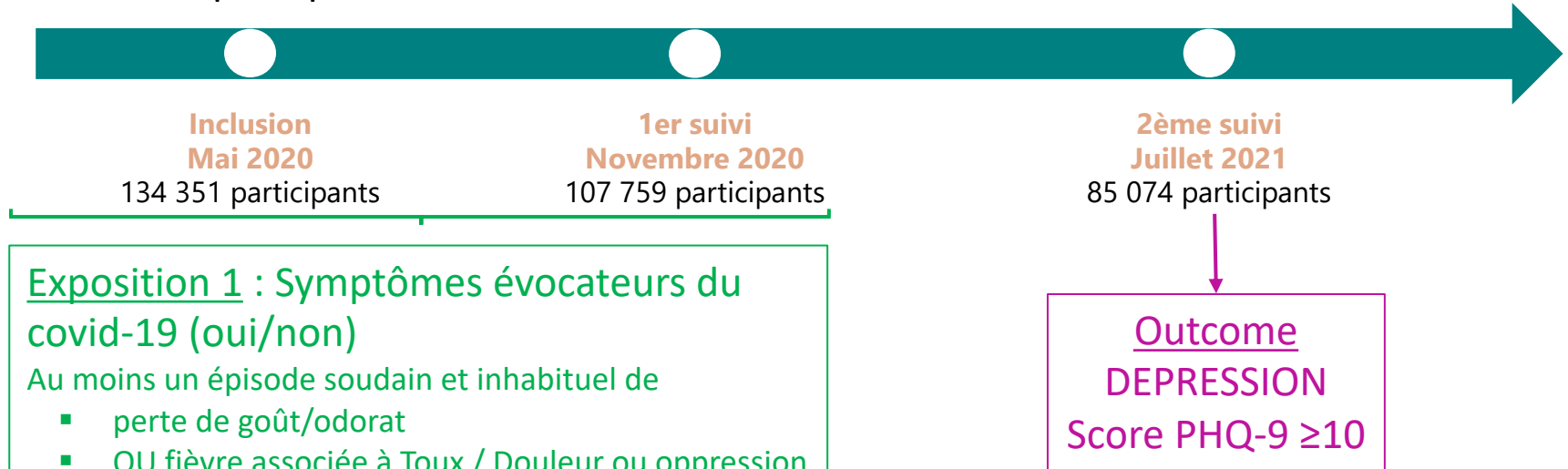
} ~~2 outcomes~~

❖ En population générale

➔ Cohorte EpiCoV (Epidémiologie et Conditions de Vie)

Illustration - Méthodes

- ❖ Mai 2020 : sélection aléatoire de 371 000 individus ≥ 15 ans dans la base Fideli
 - Taux de participation : 36,2%



Exposition 1 : Symptômes évocateurs du covid-19 (oui/non)

Au moins un épisode soudain et inhabituel de

- perte de goût/odorat
- OU fièvre associée à Toux / Douleur ou oppression thoracique / Difficultés respiratoires ou essoufflement

Entre Février et Novembre 2020

Illustration - Méthodes

- ❖ Mai 2020 : sélection aléatoire de 371 000 individus ≥ 15 ans dans la base Fideli
 - Taux de participation : 36,2%

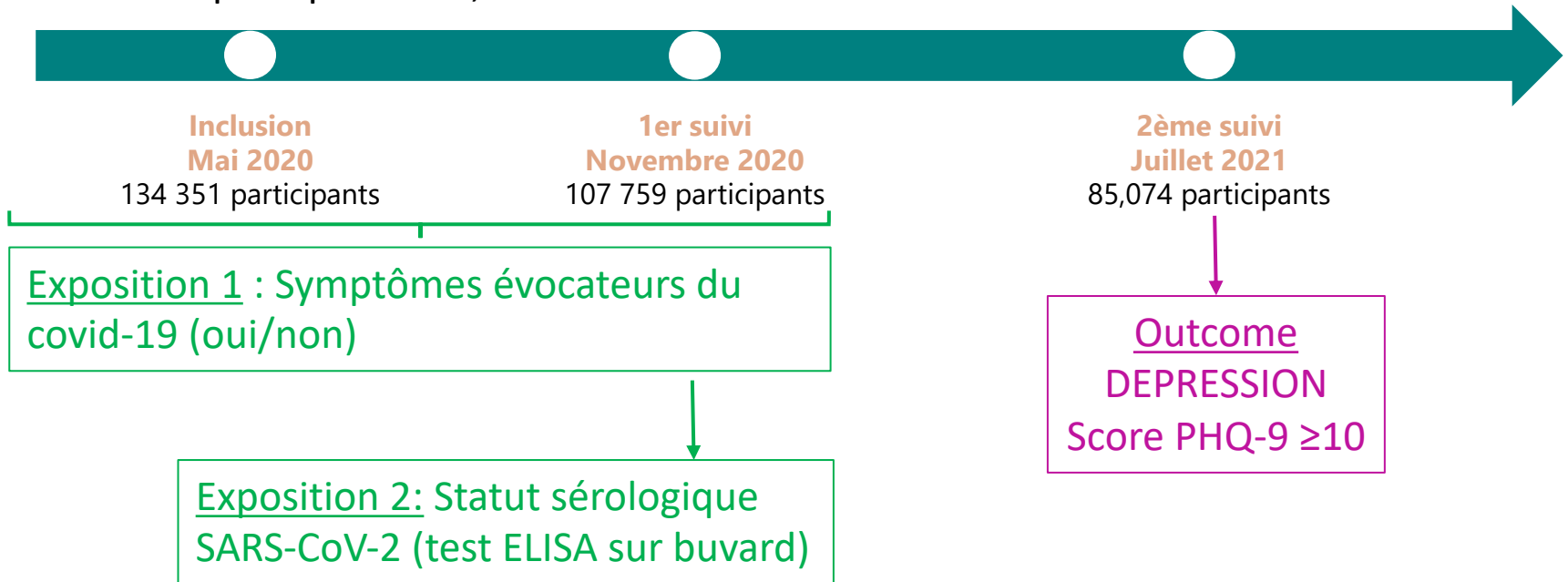


Illustration - Méthodes

- ❖ Mai 2020 : sélection aléatoire de 371 000 individus ≥ 15 ans dans la base Fideli
 - Taux de participation : 36,2%



**Inclusion
Mai 2020**

Caractéristiques sociales

- Age
- Sexe
- Lieu de naissance
- Niveau de diplôme
- Structure du foyer
- Lieu de confinement
- Jardin ou balcon
- Lieu de vie (densité urbaine, quartier prioritaire, taux d'hospitalisation)

**1er suivi
Novembre 2020**

Caractéristiques économiques

- Emploi
- Situation financière perçue
- Revenu du foyer
- Foyer surpeuplé

Caractéristiques liées à la santé

- Santé perçue
- Indice de masse corporelle
- Maladie chronique somatique
- Maladie chronique psychiatrique
- Consommation de tabac
- Consommation d'alcool

**2ème suivi
Juillet 2021**

Illustration - Analyses

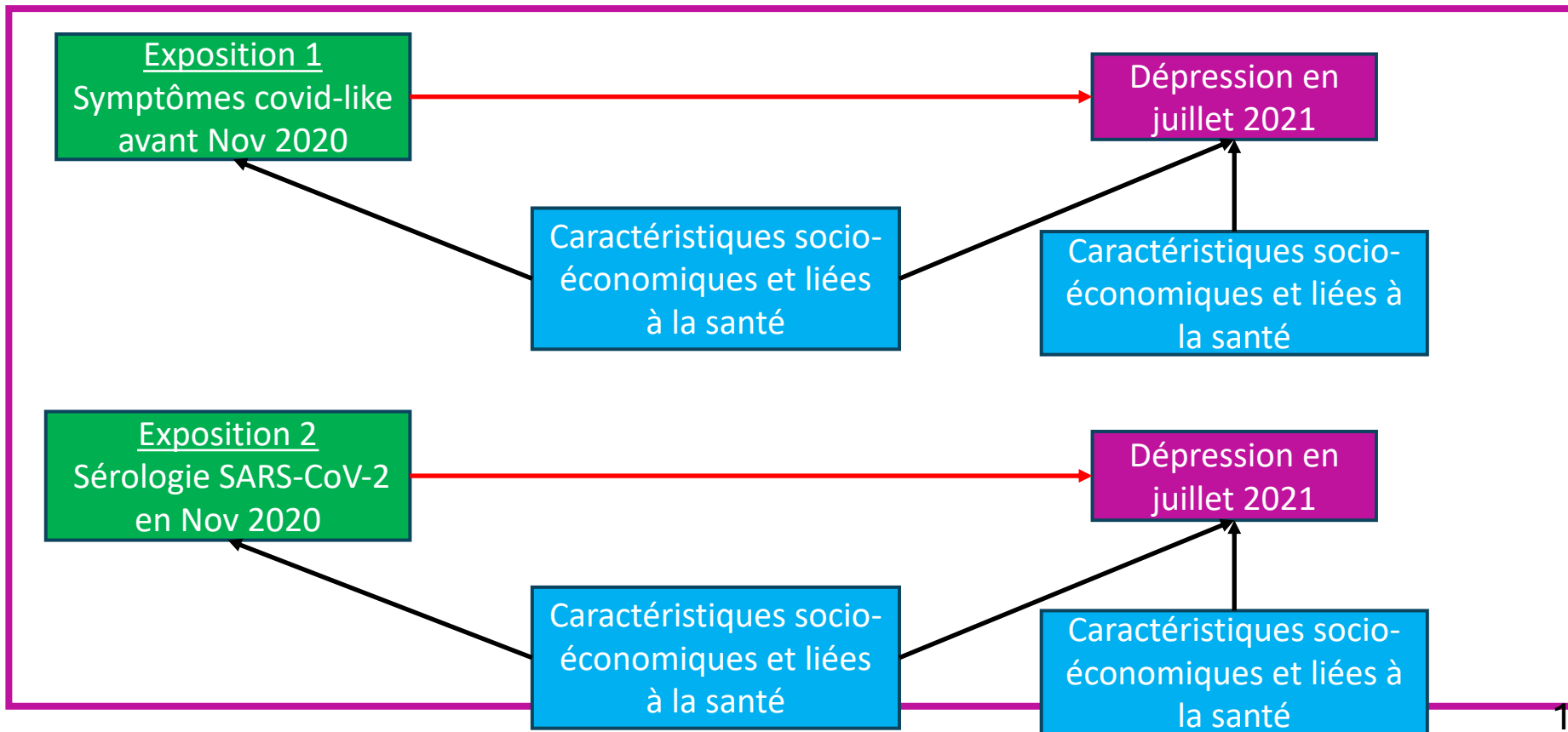


Illustration - Analyses

Exposition 1
Symptômes covid-like
avant Nov 2020

PROPENSITY SCORE MATCHING

Dépression en
juillet 2021

Exposition 2
Sérologie SARS-CoV-2
en Nov 2020

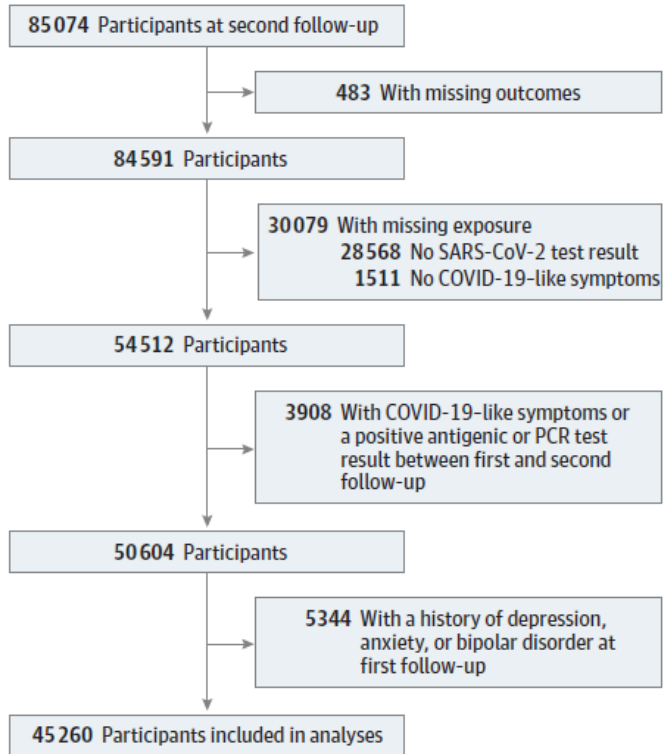
PROPENSITY SCORE MATCHING

Dépression en
juillet 2021

NB :

- **Imputations multiples** pour gérer les données manquantes
- Analyses descriptives **pondérées** pour tenir compte de la non-réponse

Illustration – Résultats : Description



Symptômes Covid-like

	Avec (N= 5534)	Sans (N= 39 726)
Dépression	14.5% (13.2 – 15.7)	7.1% (6.8 – 7.5)
Séropositivité SARS-CoV-2	32.6% (31.0 – 34.1)	6.3% (6.0 – 6.6)

Sérologie SARS-CoV-2

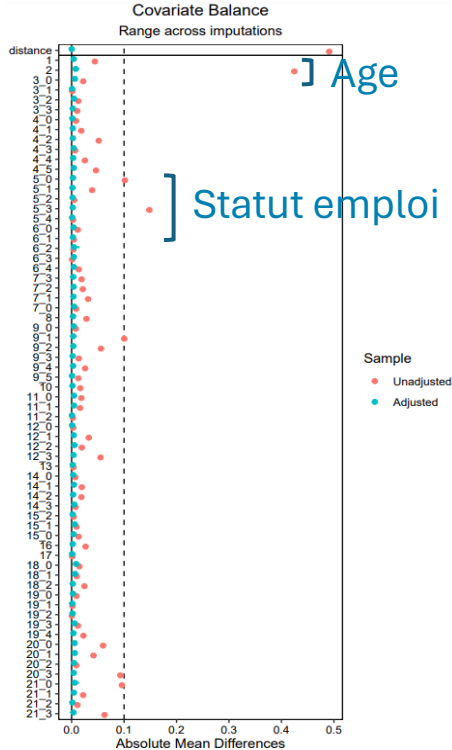
	Positive (N= 4348)	Négative (N= 40 912)
Dépression	9.1% (7.9 – 10.3)	7.9% (7.5 – 8.2)
Symptômes Covid-like	40.9% (39.0 – 42.8)	8.8% (8.5 – 9.2)

Illustration – Résultats : Qualité des SdP

**Equilibre des variables
utilisées pour le calcul du SdP
dans les groupes E – NE**

Différences moyennes
standardisées

Symptômes Covid-like



Sérologie SARS-CoV-2

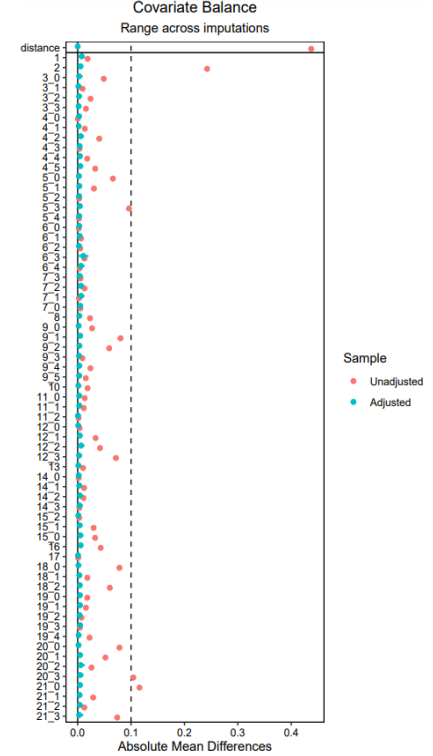


Illustration – Résultats

		Effectif avant appariement	Nombre de paires	Dépression	
				OR _{brut} [IC95%]	OR _{ajusté} [IC95%] ^{ps}
Tous		45 260			
Symptômes Covid-like	Non	39 726	5531	1 (Référence)	1 (Référence)
	Oui	5534		2.16 [1.97-2.35]	1.70 [1.45-1.99]
Sérologie SARS-CoV-2	Négative	40 912	4348	1 (Référence)	1 (Référence)
	Positive	4348		1.19 [1.06-1.33]	1.11 [0.85-1.44]

Le SdP : Conclusion

- **Mime la randomisation** → **Confusion résiduelle**

- **Taille de l'échantillon** →

Régression logistique peut être préférable

- Faible overlap
- Faible prévalence de l'exposition
- Taille d'échantillon > 10 000
- Etudes des interactions
- Prise en compte des poids de sondage

- **Avantages du SdP**



- Petits échantillons ou outcomes rares
- Choix théorique des variables utilisées pour la construction SdP
- Limite l'inférence là où il y a de l'information (support

Illustration – Remerciements

- ❖ Etude labelisée Priorité Nationale de Recherche par CAPNET.
- ❖ Les investigateurs remercient pour leur financement et leur soutien :
 - ANRS | MIE
 - Ministère de la Santé et de la Prévention
 - Ministère de l'enseignement supérieur et de la recherche
 - INSERM
 - DREES
 - Région Ile de France
 - Programme Européen Horizon 2020 de recherche et d'innovation (grant agreement 101016167 ORCHESTRA)
- ❖ Les investigateurs remercient aussi tous les participants à EpiCoV



Le SdP : Discussion

- **~Randomisation** → Confusion résiduelle
- **Taille de l'échantillon** → Régression logistique peut être préférable
 - Faible overlap
 - Faible prévalence de l'exposition
 - Taille d'échantillon > 10 000
 - Etudes des interactions
 - Prise en compte des poids de sondage
- **Limites**
 - Hypothèses
 - Equilibre E-NE
- **Avantages**
 - Choix théorique des variables utilisées pour la construction SdP
 - Limite l'inférence là où il y a de l'information (support commun)
 - Eviter les extrapolations extrêmes (≠ régression ordinaire)

Le SdP : 3 - Méthodes d'ajustement

❖ Support commun

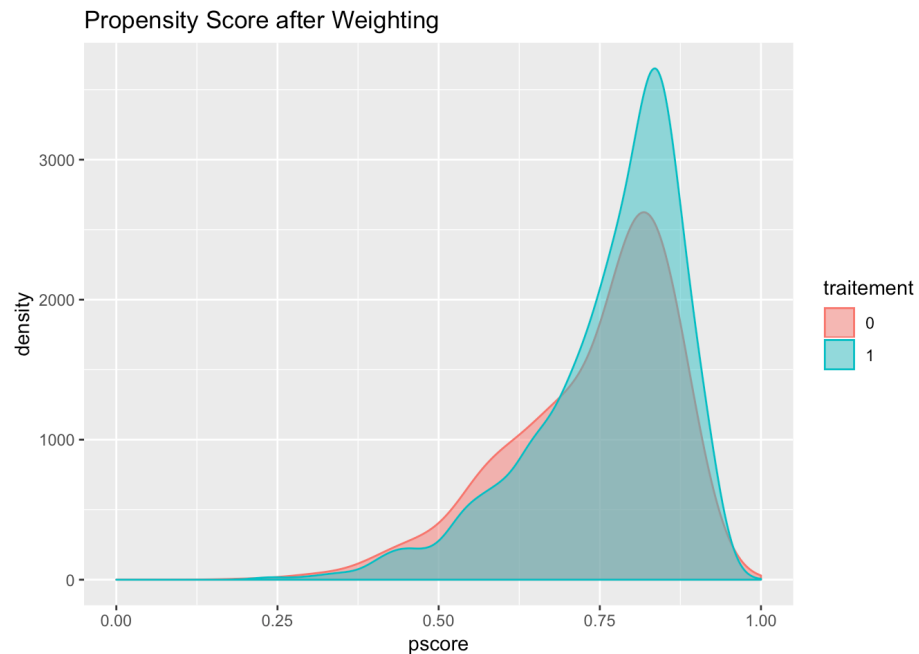
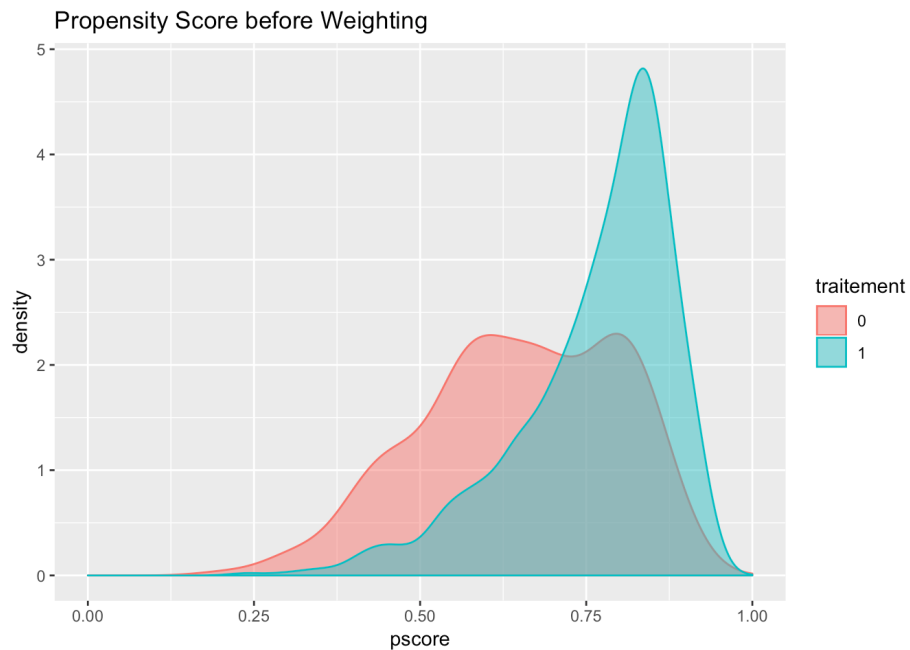


Illustration – Discussion

- ✓ **Pas d'association** observée entre la **séropositivité SARS-CoV-2** en Novembre 2020 et la présence de symptomatologie dépressive ou anxieuse en juillet 2021
 - L'infection par SARS-CoV-2 n'est pas en soi un facteur de risque de dépression ou d'anxiété à long terme ?
- ✓ **Association** entre les **symptômes évocateurs de covid-19** avant Novembre 2020 et la présence de symptomatologie dépressive ou anxieuse en juillet 2021
 - Chez les personnes **séropositives** pour SARS-CoV-2 en Novembre 2020
 - **Différence selon les formes cliniques (sévérité, durée) de covid-19?** (Taquet *et al.* 2021.1, 2021.2, 2022, *Lancet Psychiatry*, Xie *et al.* 2022, *BMJ*)
 - Chez les personnes **séronégatives** pour SARS-CoV-2 en Novembre 2020 (Thompson *et al.* 2022 *Lancet Psychiatry*; Matta *et al.* 2022 *JAMA Intern Med*)

→ Association contextuelle ?

Illustration – Résultats

	Effectif avant appariement	Nombre de paires	Dépression		
			OR _{brut} [IC95%]	OR _{ajusté} [IC95%] ^{ps}	
Tous	45 260				
Symptômes Covid-like	Non	39 726	5531	1 (Référence)	1 (Référence)
	Oui	5534		2.16 [1.97-2.35]	1.70 [1.45-1.99]
Sérologie SARS-CoV-2	Négative	40 912	4348	1 (Référence)	1 (Référence)
	Positive	4348		1.19 [1.06-1.33]	1.11 [0.85-1.44]