UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE APPLIQUÉES

PAR
Faustin Kagabo

INFÉRENCE CAUSALE AVEC LA MÉTHODE PEACE ET L'APPRENTISSAGE AUTOMATIQUE

Août 2025

REMERCIEMENTS

Avant tout, je tiens à exprimer ma plus profonde gratitude envers le Seigneur Tout-Puissant, dont la force et la guidance m'ont soutenu tout au long de ce parcours.

Je suis profondément reconnaissant envers mon directeur de recherche, Professeur Usef Faghihi, pour son encadrement précieux, son soutien indéfectible et ses conseils éclairés, qui ont été déterminants dans la réalisation de ce mémoire et de l'ensemble de mon parcours de maîtrise. Je tiens également à remercier chaleureusement le Dr Amir Saki pour sa disponibilité, la richesse de nos échanges, ainsi que pour ses contributions précieuses, qui ont nourri et approfondi ma réflexion tout au long de ce travail.

Je tiens également à adresser mes remerciements les plus sincères à Alioune Badara Diao, dont les retours attentifs et le soutien collégial ont été grandement appréciés. Nos séances de relecture mutuelle et nos échanges de suggestions ont grandement contribué à l'amélioration de ce travail.

Enfin, je remercie chaleureusement ma famille et mes amis proches pour leur soutien constant, leur patience et leurs encouragements, sans lesquels cette réalisation n'aurait pas été possible.

Cette aventure a été une expérience enrichissante, qui m'a permis d'aiguiser non seulement mes compétences académiques, mais aussi ma résilience, mon esprit critique et mon sens de l'engagement.

DÉDICACE

À mon père, dont le parcours académique m'a toujours inspiré.

Pendant qu'il préparait son doctorat en agriculture et en climat, il profitait de ses congés pour planter des carottes, du céleri et des pommes de terre dans notre jardin. Il comptait les jours et suivait leur croissance dans le cadre de ses recherches. Je l'aidais à arroser ces plantes, sans me rendre compte que ces moments simples allaient profondément me marquer.

Ils m'ont appris la patience, la curiosité et la valeur de l'apprentissage par la pratique. Son dévouement et son soutien constant m'ont accompagné tout au long de mon propre parcours.

Ce travail lui est dédié, avec tout mon amour et ma profonde gratitude.

TABLE DES MATIÈRES

LISTE DES FIGURES

LISTE DES TABLEAUX

RÉSUMÉ

L'inférence causale a pour objectif de mieux comprendre les relations de cause à effet à partir de données observationnelles complexes.

Ce mémoire étudie la capacité de la méthode PEACE (Probabilistic Easy Variational Causal Effect) à identifier des effets causaux directs potentiellement non linéaires ou spécifiques à certaines sous-populations dans des données de santé complexes. Contrairement aux approches traditionnelles telles que l'estimation de l'ATE/CATE ou les modèles causaux structurels (SCM), qui reposent sur des hypothèses structurelles fortes, ainsi qu'aux méthodes neuronales comme TARNet ou DragonNet, qui offrent une interprétabilité limitée au niveau des variables, la méthode PEACE estime les effets causaux directs à l'aide de la variation totale. Sa version normalisée, Mean PEACE, affine cette estimation en pondérant les effets proportionnellement à leur vraisemblance selon la densité conditionnelle $f_{X|Z}$, permettant ainsi de prendre en compte à la fois les cas rares et fréquents. Appliquée à des données de santé portant sur des maladies chroniques, PEACE révèle des contributions causales directionnelles et interprétables tout en restant robuste dans des environnements hétérogènes et de grande dimension.

Mots-clés : Inférence causale, effets causaux directs, PEACE, interprétabilité.

CHAPITRE 1
INTRODUCTION

L'inférence causale joue un rôle fondamental dans un large éventail de disciplines, allant de la philosophie et l'épidémiologie à l'économie, aux sciences sociales et à l'intelligence artificielle. Elle permet aux chercheurs non seulement d'identifier des régularités dans les données, mais aussi de poser des questions contrefactuelles plus profondes telles que « Que se serait-il passé si une autre action avait été entreprise ? » essentielles pour l'explication, la prédiction et la prise de décision. Contrairement à la modélisation statistique classique, qui se concentre sur les associations, l'inférence causale cherche à dévoiler les mécanismes sous-jacents qui génèrent les résultats observés. Cette démarche repose toutefois sur des hypothèses comme l'absence de confusion latente qui sont souvent difficiles à vérifier, en particulier dans les contextes observationnels ou à haute dimension.

Plusieurs cadres théoriques ont façonné le développement de l'inférence causale. Le modèle des *Potential Outcomes* [1] formalise le raisonnement contrefactuel sous des hypothèses telles que l'ignorabilité et la consistance. Le *Structural Causal Model* (SCM) de Pearl [2] utilise des structures graphiques et le do-calculus pour dériver des effets causaux à partir de graphes acycliques dirigés. Plus récemment, Janzing et al. [3] ont introduit une perspective informationnelle, mettant l'accent sur les influences causales à l'échelle macro. Chacun de ces cadres apporte un éclairage spécifique, mais leur applicabilité peut être restreinte face à la complexité ou à l'hétérogénéité des données empiriques, et leur explicabilité reste souvent limitée lorsqu'il s'agit d'en extraire des leviers d'action concrets.

Le cadre *PEACE (Probabilistic Easy Variational Causal Effect)* [4] propose une alternative. Il estime les effets causaux directs à l'aide d'une formulation variationnelle qui combine densités de probabilité conditionnelles et variation totale ; le tout s'inspire de la notion d'intervention introduite par Pearl [2]. PEACE introduit également un principe probabiliste innovant, la « disponibilité naturelle du changement », qui mesure la probabilité de transition entre différents niveaux de traitement au sein des sous-populations observées. Un paramètre de degré ajustable permet de mettre l'accent soit sur les sous-groupes rares, soit sur les tendances dominantes. Grâce à cette flexibilité, PEACE s'adapte aussi bien aux problématiques causales micro qu'aux perspectives macro, quelles que soient la complexité ou l'hétérogénéité des données.

Dans ce mémoire, la méthode PEACE est appliquée à un ensemble de données synthétique conçu pour reproduire la complexité des données de santé réelles, en intégrant des variables comportementales, socio-économiques et cliniques [5]. Le critère principal étudié est la présence de maladies chroniques. Ce cadre offre un environnement à la fois contrôlé et réaliste pour évaluer la capacité de PEACE à détecter des relations causales significatives sans dépendre d'hypothèses structurelles fortes.

Nous examinons dans ce mémoire dans quelle mesure la méthode PEACE [4] permet d'identifier des effets causaux directs, y compris non linéaires ou spécifiques à certaines sous-populations, dans des données de santé complexes, là où les approches classiques telles que l'estimation de l'ATE/CATE [1] ou les modèles causaux structurels (SCM) [2] reposent sur des hypothèses structurelles fortes, et où les méthodes neuronales modernes comme TARNet [6] ou DragonNet [7] offrent une interprétabilité limitée au niveau des variables.

1.1 Fondements historiques et conceptuels de la causalité

La notion de causalité possède des racines philosophiques profondes, remontant à Aristote, qui distinguait plusieurs types de causes : matérielle, formelle, efficiente et finale [8, 9]. Bien plus tard, David Hume a remis en question la possibilité d'observer directement des liens causaux, affirmant que nous ne percevons que des régularités, sans nécessité réelle [10]. Ce scepticisme a motivé des efforts modernes pour formaliser le raisonnement causal à l'aide d'outils mathématiques et probabilistes [11, 2].

Un défi majeur en analyse de données consiste à distinguer la corrélation de la causalité. Alors que la corrélation reflète une association statistique entre des variables, la causalité implique qu'un changement dans une variable induit directement un changement dans une autre [12]. Confondre les deux peut entraîner des interprétations erronées, notamment en attribuant un lien causal à des motifs fallacieux influencés par des facteurs cachés. Un exemple bien connu est la corrélation observée entre les ventes de crème glacée et les noyades : les deux augmentent durant l'été, mais la crème glacée ne cause pas les noyades c'est la température qui constitue la cause commune cachée [2] [1].

Une illustration frappante des limites de l'association est le *paradoxe de Simpson*, où une tendance observée dans plusieurs sous-groupes se renverse lorsque les données sont agrégées [2, 13]. Par exemple, pendant la pandémie de COVID-19, les données agrégées pouvaient laisser penser que les personnes vaccinées étaient plus souvent hospitalisées que les non-vaccinées. Cependant, en stratifiant par âge, on observe que les personnes âgées plus susceptibles d'être vaccinées et aussi plus vulnérables aux formes graves étaient surreprésentées dans le groupe vacciné. Au sein de chaque groupe d'âge, la vaccination réduisait en réalité le risque d'hospitalisation. Ce renversement est dû à des variables de confusion et ne peut être résolu par l'association seule. Les cadres d'inférence causale tels que les modèles de résultats potentiels [14] et les modèles causaux structurels [2] sont conçus pour traiter ce type de paradoxe en modélisant explicitement le processus de génération des données [13].

1.2 Organisation du mémoire

Ce mémoire est organisé en cinq chapitres, chacun jouant un rôle complémentaire :

— **Chapitre 1 - Introduction** : situe le contexte de l'inférence causale, ses fondements historiques et conceptuels, et expose les motivations et objectifs de la recherche.
— **Chapitre 2 - État de l'art** : présente les cadres d'inférence (résultats potentiels, SCM, approche informationnelle), les *concepts causaux clés* (ATE, CATE, effets direct/total), les méthodes d'estimation classiques et modernes, ainsi que les *algorithmes de découverte causale*.

1. Dans cet exemple, la température agit comme une variable de confusion qui augmente à la fois la probabilité de consommer de la crème glacée et celle d'aller nager (donc de se noyer). La corrélation observée entre les ventes de glace et les cas de noyade est donc fallacieuse.

— **Chapitre 3 – Fondements variationnels et méthode PEACE** : introduit la variation totale, formalise la méthode PEACE, ses propriétés et variantes (directionnelles et moyennes), et la compare à d'autres métriques causales.
— **Chapitre 4 – Évaluation empirique** : appliquée à une base de données de santé, décrit la base, l'analyse exploratoire, le prétraitement, l'analyse statistique de l'importance des variables (SHAP), les tentatives de découverte causale, l'implémentation de PEACE (avec incertitude par bootstrap), puis l'analyse et l'interprétation des résultats obtenus.
— **Chapitre 5 – Conclusion et perspectives** : synthétise les contributions et propose des pistes de recherche futures.

CHAPITRE 2
ÉTAT DE L'ART

Pour comprendre et estimer des relations causales dans des ensembles de données complexes qu'ils soient réels ou synthétiques, il est essentiel d'examiner comment les cadres d'inférence causale ont évolué historiquement et conceptuellement. Au cours des dernières décennies, le domaine s'est enrichi de théories et modèles influents, passant des approches contrefactuelles classiques à des cadres plus sophistiqués exploitant l'apprentissage automatique et la modélisation probabiliste [2, 14].

Historiquement, trois grands cadres ont posé les fondations de l'inférence causale moderne : (1) le cadre des résultats potentiels (modèle de Rubin-Neyman), (2) le modèle structurel causal (SCM) proposé par Judea Pearl, et (3) l'inférence causale fondée sur la théorie de l'information. Ces trois approches restent dominantes en raison de leur solidité théorique et de leur large adoption. Cependant, elles présentent des limites dans le traitement des données empiriques complexes, en particulier dans les cas impliquant des variables continues, des événements rares ou des structures causales imbriquées à plusieurs niveaux.

Ce chapitre examine de manière critique l'évolution de ces cadres, met en lumière leurs points forts et leurs limites, puis présente les contributions méthodologiques récentes en particulier les approches fondées sur l'apprentissage automatique, telles que le modèle *Probabilistic Easy Variational Causal Effect* (PEACE) qui permettent de surmonter certaines de ces limites et d'établir une base robuste pour l'analyse causale dans des environnements complexes.

2.1 Cadres d'inférence causale

2.1.1 Le cadre des résultats potentiels

Aussi appelé modèle de Rubin-Neyman, ce cadre conceptualise la causalité en comparant deux résultats hypothétiques pour chaque individu : $Y(1)$ si l'individu est traité, et $Y(0)$ s'il ne l'est pas. Comme un seul de ces résultats est observable, l'inférence causale devient un problème de données manquantes.

Une quantité causale centrale dans ce cadre est l'**effet moyen du traitement (ATE)** :

$$ATE = \mathbb{E}[Y(1) - Y(0)], \tag{2.1}$$

Cette mesure évalue la différence moyenne des résultats si toute la population recevait le traitement comparé à si aucun individu ne le recevait.

Ce cadre repose sur deux hypothèses fondamentales :

- **Ignorabilité (ou absence de confusion)** : l'assignation au traitement est indépendante des résultats potentiels conditionnellement aux covariables observées.
- **Cohérence (ou consistance)** : le résultat observé pour un individu correspond exactement à son résultat potentiel sous le traitement effectivement reçu.

Sous ces hypothèses, des effets causaux comme l'ATE peuvent être identifiés et estimés à partir de données observationnelles [1].

2.1.2 Les modèles causaux structurels (SCM)

Proposé par Judea Pearl, le modèle structurel causal (SCM) représente explicitement les relations causales à l'aide d'équations structurelles et de graphes orientés acycliques. Ce formalisme fournit des outils puissants pour raisonner sur les structures causales et les conditions d'identifiabilité des effets [2].

Un élément central du SCM est l'opérateur **do**, noté :

$$P(Y \mid do(X = x)), \tag{2.2}$$

qui représente la distribution de Y lorsque X est fixé artificiellement à une valeur x, rompant ainsi ses dépendances causales naturelles.

Les concepts clés de ce cadre incluent :

- **d-séparation** : critère graphique permettant d'identifier les indépendances conditionnelles à partir du graphe causal.
- **Critère du chemin arrière (Backdoor)** : permet l'identification des effets causaux en ajustant les variables confondantes, c'est-à-dire en bloquant tous les chemins "backdoor" menant de la variable de traitement à la variable d'intérêt.
- **Critère du chemin avant (Frontdoor)** : permet l'identification même en présence de confusion non observable, via l'utilisation de variables intermédiaires.

La formalisation graphique du SCM rend les hypothèses sous-jacentes à l'inférence causale plus explicites et vérifiables.

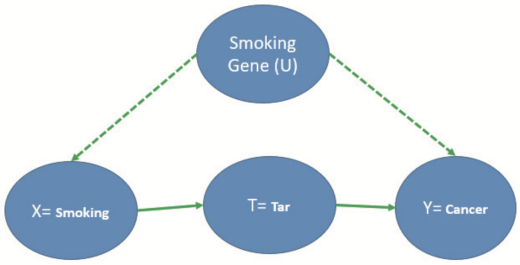


FIGURE 2.1 – Un diagramme causal de l'exemple « Tabac et Cancer » tiré de [15]. Le gène de la dépendance au tabac (U) agit comme un facteur de confusion, influençant à la fois le comportement tabagique et le cancer, tandis que le goudron joue le rôle de médiateur entre le tabagisme et le cancer.

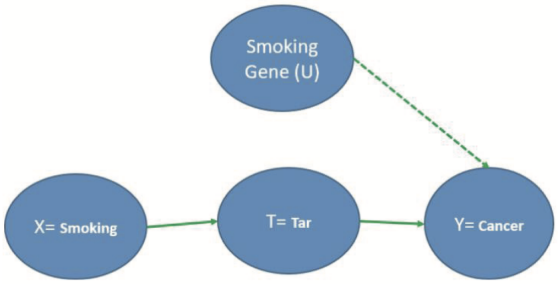


FIGURE 2.2 – Illustration de la distinction entre observation et intervention. Tandis que $P(Y \mid X)$ peut refléter une association statistique, l'effet causal pertinent est représenté par $P(Y \mid do(X))$. L'intervention sur X supprime l'influence naturelle du facteur de confusion U, rompant ainsi le chemin arrière $X \leftarrow U \rightarrow Y$. Cette figure met en évidence la différence fondamentale entre observer une variable et l'intervenir activement. Adapté de [15].

2.1.3 Inférence causale fondée sur la théorie de l'information

Janzing et al. [3] ont introduit une approche de l'inférence causale fondée sur la théorie de l'information, qui quantifie l'influence causale à partir de dépendances statistiques directionnelles, sans recourir à des interventions ou à des hypothèses structurelles. Ce cadre se concentre sur les effets causaux à l'échelle macro, en évaluant comment les variations dans la distribution des variables d'entrée affectent le contenu informationnel des variables de sortie. Bien que moins répandue dans les applications pratiques, cette approche offre des perspectives précieuses sur le comportement causal à l'échelle des systèmes et permet de situer les méthodes récentes comme PEACE, qui cherchent à unifier les raisonnements causaux aux niveaux micro et macro.

2.2 Concepts causaux clés

Pour illustrer clairement ces concepts, nous nous appuyons sur un exemple canonique présenté dans [16], impliquant quatre variables : un gène associé au tabagisme, le comportement tabagique, l'exposition au goudron et l'incidence du cancer. Cet exemple est représenté graphiquement à la Figure 2.1.

2.2.1 Effet Moyen du Traitement (ATE)

L'**effet moyen du traitement (ATE)** quantifie l'impact global d'un traitement en comparant les résultats moyens entre les populations traitées et non traitées. Formellement, il s'écrit :

$$ATE = \mathbb{E}[Y(1) - Y(0)], \tag{2.3}$$

où $Y(1)$ et $Y(0)$ représentent respectivement les résultats potentiels sous traitement et sous contrôle. Cette mesure capture le changement attendu dans le résultat si l'ensemble de la population était soumis au traitement plutôt qu'à la condition de contrôle [1].

Dans le cadre des Modèles Causaux Structurels (SCM), l'ATE s'interprète à l'aide de l'opérateur *do* de Pearl comme suit :

$$ATE = \mathbb{E}[Y \mid do(T = 1)] - \mathbb{E}[Y \mid do(T = 0)], \tag{2.4}$$

ce qui met en évidence la différence entre les résultats attendus dans des scénarios d'intervention hypothétiques [2].

Dans l'exemple du tabagisme (Figure 2.1), l'ATE reflète la différence moyenne d'incidence du cancer si toute la population fumait, comparée à une situation où personne ne fumerait. Comme l'illustre la Figure 2.2, l'opérateur $do(X)$ introduit par Pearl formalise ce type d'intervention, qui permet d'isoler l'effet causal en supprimant l'influence des variables de confusion.

2.2.2 Effet Moyen Conditionnel du Traitement (CATE)

L'**effet moyen conditionnel du traitement (CATE)** mesure l'effet d'un traitement conditionnellement à certaines covariables observées X. Il permet de capturer l'hétérogénéité des effets causaux au sein des différents sous-groupes d'une population. Formellement, dans le cadre des résultats potentiels, il est défini comme suit :

$$CATE(X) = \mathbb{E}[Y(1) - Y(0) \mid X], \tag{2.5}$$

ce qui quantifie la différence attendue des résultats pour les individus ayant les covariables X, selon qu'ils reçoivent le traitement ou non [1].

Dans le cadre des Modèles Causaux Structurels (SCM), le CATE peut être exprimé à l'aide de l'opérateur *do* de Pearl :

$$CATE(X) = \mathbb{E}[Y \mid do(T = 1), X] - \mathbb{E}[Y \mid do(T = 0), X], \tag{2.6}$$

ce qui offre une interprétation graphique des effets causaux conditionnels sous intervention [2].

Par exemple, comme illustré à la Figure 2.1, le CATE permet d'observer comment l'effet causal du tabagisme sur le cancer varie selon la présence ou l'absence du gène associé au tabac.

2.2.3 Effet Direct

L'**effet direct** désigne la composante de l'influence causale d'un traitement sur un résultat qui n'est pas médiée par une variable intermédiaire. Dans un système causal où une variable de traitement T affecte une variable de résultat Y à la fois directement et indirectement via un médiateur M, l'effet direct capture la voie causale non médiée de T vers Y.

Deux définitions couramment utilisées de l'effet direct sont : l'*Effet Direct Contrôlé (CDE)* et l'*Effet Direct Naturel (NDE)*, toutes deux formalisées dans le cadre des résultats potentiels [1] ainsi que dans le cadre des modèles causaux structurels (SCM) à l'aide de l'opérateur *do* [2].

— **Effet Direct Contrôlé (CDE)** : Le CDE mesure l'effet d'un changement de traitement T tout en maintenant le médiateur M fixé à une valeur prédéterminée m. Il répond à la question contrefactuelle suivante : quel serait le changement dans Y si l'on intervenait simultanément sur T et M ?

Notation dans le cadre des résultats potentiels :

$$CDE(m) = \mathbb{E}[Y(1, m)] - \mathbb{E}[Y(0, m)] \tag{2.7}$$

où $Y(t, m)$ représente le résultat potentiel lorsque $T = t$ et $M = m$.

Notation dans le cadre SCM :

$$CDE(m) = \mathbb{E}[Y \mid do(T = 1, M = m)] - \mathbb{E}[Y \mid do(T = 0, M = m)] \tag{2.8}$$

où l'opérateur *do* représente une intervention rompant les liens causaux naturels vers T et M.

— **Effet Direct Naturel (NDE)** : Le NDE isole l'effet direct du traitement en comparant les résultats lorsque T change, tandis que le médiateur M est maintenu à la valeur qu'il aurait naturellement prise en l'absence de traitement. Il mesure ainsi l'effet de T sur Y qui ne passe pas par M.

Notation dans le cadre des résultats potentiels :

$$NDE = \mathbb{E}[Y(1, M(0))] - \mathbb{E}[Y(0, M(0))] \tag{2.9}$$

où $M(0)$ désigne la valeur potentielle de M lorsque $T = 0$.

Notation dans le cadre SCM :

$$NDE = \mathbb{E}[Y \mid do(T = 1, M = M_{T=0})] - \mathbb{E}[Y \mid do(T = 0, M = M_{T=0})] \tag{2.10}$$

où $M_{T=0}$ est la valeur du médiateur sous une intervention fixant $T = 0$.

Ces définitions peuvent être illustrées à l'aide de l'exemple de la Figure 2.1. Le NDE quantifie dans quelle mesure le gène du tabagisme affecte directement l'incidence du cancer, indépendamment de son effet indirect via le comportement tabagique. En revanche, le CDE répond à la question suivante : comment le risque de cancer changerait-il si tout le monde était contraint de fumer ($T = 1$) ou de ne pas fumer ($T = 0$), tout en maintenant constant le niveau de goudron dans les poumons à une valeur fixée m ?

2.2.4 Effet Total

L'**effet total** capture l'impact global d'un traitement sur un résultat, en englobant à la fois les voies causales directes et indirectes. Dans le cadre des résultats potentiels, il est défini comme suit :

$$\text{Effet total} = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)], \tag{2.11}$$

ce qui représente le changement attendu du résultat si toute la population recevait le traitement comparé à une situation où aucun individu ne serait traité [1].

Dans le cadre des Modèles Causaux Structurels (SCM), cette quantité est exprimée à l'aide de l'opérateur *do* de Pearl :

$$\text{Effet total} = \mathbb{E}[Y \mid do(T = 1)] - \mathbb{E}[Y \mid do(T = 0)], \tag{2.12}$$

soulignant ainsi une approche fondée sur l'intervention [2].

Dans l'exemple du tabagisme (Figure 2.1), l'effet total du gène associé au tabac sur le cancer inclut à la fois son influence génétique directe et ses effets indirects via le comportement tabagique et l'accumulation de goudron.

Méthodes d'estimation classiques et approches modernes en apprentissage automatique

L'inférence causale s'est historiquement appuyée sur des méthodes statistiques fondées sur des hypothèses fortes et des modèles dits transparents [1, 14], c'est-à-dire des cadres simples et interprétables où les hypothèses et relations causales sont explicitement formulées. À l'inverse, les approches modernes d'apprentissage automatique offrent une flexibilité accrue dans les contextes à haute dimension, mais sont souvent perçues comme des boîtes noires, leurs mécanismes internes étant difficiles à interpréter [17]. Cette section propose une analyse critique des deux paradigmes.

2.3 Méthodes d'estimation classiques

2.3.1 Variables Instrumentales (IV)

Les méthodes par variables instrumentales (IV) permettent une estimation cohérente des effets causaux en présence de confusion non observée. Un instrument valide influence le traitement mais n'a aucun effet direct sur le résultat, sauf à travers le traitement. Cette méthode est largement utilisée en économétrie et a été formalisée dans le cadre de l'inférence causale par Angrist et Imbens [18].

2.4 Approches modernes en apprentissage automatique

2.4.1 Arbres causaux et forêts causales

Les arbres causaux et les forêts causales étendent les arbres de décision et les forêts aléatoires pour estimer des effets de traitement hétérogènes au sein de sous-populations [19, 20]. Ces méthodes modélisent l'hétérogénéité des effets de traitement de manière flexible, tout en reposant sur l'hypothèse forte d'ignorabilité (2.13) et sur une condition de recouvrement entre les groupes traité et témoin. Les forêts causales, en particulier, utilisent des arbres « honnêtes » où la construction de la structure et l'estimation des effets sont effectuées sur des sous-échantillons distincts, ce qui permet d'obtenir des estimateurs asymptotiquement normaux et des intervalles de confiance valides pour les effets de traitement conditionnels (CATE).

$$\big(Y(0), Y(1)\big) \;\perp\!\!\!\perp\; T \;\big|\; X, \tag{2.13}$$

— $Y(0), Y(1)$: résultats potentiels sans (0) et avec (1) traitement ;
— $T \in \{0, 1\}$: variable d'assignation au traitement ;
— X : vecteur de covariables supposées suffire à lever les biais de confusion ;
— $\perp\!\!\!\perp$: indépendance probabiliste *conditionnelle*.

2.4.2 Réseaux de neurones et apprentissage par représentation

Des modèles d'apprentissage profond comme TARNet et DragonNet utilisent des représentations latentes partagées pour atténuer les biais de confusion dans les données observationnelles. Comme illustré dans les Figures 2.4 et 2.3, TARNet [6] estime les résultats potentiels à l'aide de têtes distinctes conditionnées par le traitement, tandis que DragonNet [7] étend cette architecture en apprenant conjointement un score de propension avec les résultats potentiels. En équilibrant les covariables dans un espace latent appris, ces architectures améliorent l'estimation des effets causaux dans les contextes à haute dimension. Toutefois, leur caractère de « boîte noire » peut rendre les mécanismes causaux difficiles à interpréter [21].

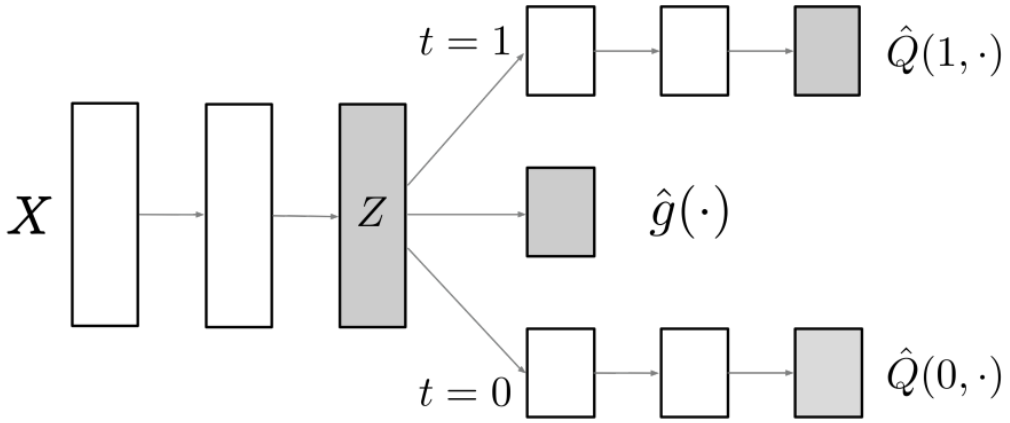


FIGURE 2.3 – Architecture de DragonNet, adaptée de [7]. Le modèle génère une représentation partagée $Z(X)$ à partir des covariables, puis se divise en trois branches : une tête de prédiction du traitement pour les scores de propension $\hat{g}(\cdot)$, et deux têtes de résultats potentiels $\hat{Q}(1, \cdot)$ et $\hat{Q}(0, \cdot)$. La fonction de perte conjointe inclut l'erreur sur les résultats ainsi que l'entropie croisée sur l'assignation au traitement, afin de favoriser des représentations équilibrées.

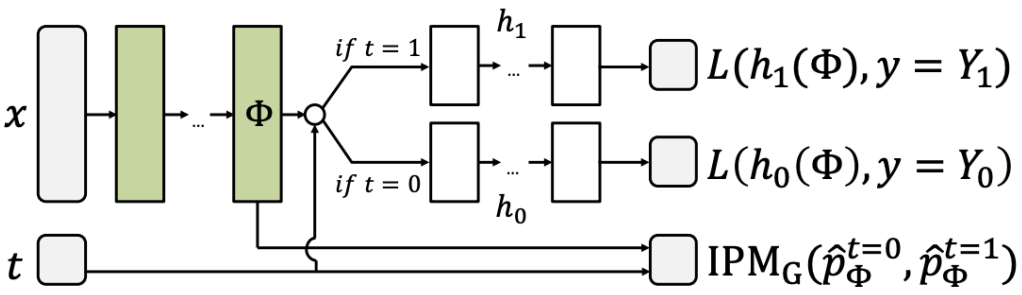


FIGURE 2.4 – Architecture de TARNet, adaptée de [6]. Le modèle projette les covariables x dans une représentation latente partagée $\Phi(x)$, puis dirige cette représentation vers l'une des deux têtes de sortie h_0 pour le résultat sous contrôle, et h_1 pour le résultat sous traitement en fonction du traitement observé t. La fonction de perte n'utilise que la tête factuelle par instance, et intègre une métrique de probabilité intégrale (IPM) pour aligner les distributions des groupes traité et témoin dans l'espace latent.

2.5 Algorithmes de découverte causale

Les algorithmes de découverte causale visent à identifier les relations causales à partir de données observationnelles, généralement représentées sous la forme d'un graphe orienté acyclique (DAG). Chaque nœud correspond à une variable, et chaque arête indique un effet causal direct potentiel. Ces algorithmes se distinguent par leurs hypothèses, leur robustesse et leur capacité à gérer la confusion. Nous présentons ci-dessous cinq approches représentatives.

2.5.1 Algorithme PC

L'algorithme PC (Peter-Clark) est une méthode basée sur des contraintes qui infère le squelette du graphe causal à l'aide d'une série de tests d'indépendance conditionnelle (CI) [11]. Les arêtes sont ensuite orientées à l'aide d'un ensemble de règles déterministes pour construire un graphe acyclique partiellement orienté. La validité de l'algorithme PC repose sur trois hypothèses clés : la condition de Markov causale (chaque variable est indépendante de ses non-effets conditionnellement à ses causes directes), la condition de fidélité (les indépendances statistiques observées reflètent parfaitement le DAG sous-jacent), et la suffisance causale (toutes les causes communes des variables mesurées sont incluses dans l' ensemble de données). Ces hypothèses sont fortes et souvent violées dans les situations réelles, ce qui peut limiter l'applicabilité de l'algorithme.

2.5.2 Fast Causal Inference (FCI)

L'algorithme FCI étend l'algorithme PC pour prendre en compte les variables latentes et les biais de sélection, ce qui le rend adapté à des environnements plus réalistes et partiellement observés [11]. Il produit un graphe ancestral partiel (PAG) qui capture la classe d'équivalence des structures causales compatibles avec les indépendances conditionnelles observées. FCI suppose les conditions de Markov causales et de fidélité, comme PC, mais relâche la suffisance causale en modélisant explicitement les variables non observées. Il suppose également l'acyclicité et soit l'absence de biais de sélection, soit sa modélisation correcte. Ces hypothèses, bien que relâchées, restent strictes et font de FCI un outil puissant pour la découverte causale en contexte d'observation partielle.

2.5.3 NOTEARS

NOTEARS (Non-combinatorial Optimization via Trace Exponential and Augmented lagRangian for Structure learning) formule l'apprentissage de la structure causale comme un problème d'optimisation différentiable en intégrant la contrainte d'acyclicité dans une forme algébrique lisse [22]. Il suppose que les données suivent un modèle d'équations structurelles linéaires (SEM) avec un bruit gaussien et sans confusion latente. Ces hypothèses assurent un paysage d'optimisation bien comporté et garantissent que le DAG appris correspond à la structure causale réelle dans des conditions idéales. Bien qu'efficace et évolutif, sa dépendance à la linéarité et à la suffisance causale limite son application à des systèmes plus complexes, non linéaires ou confondus.

2.5.4 LiNGAM

Le modèle linéaire non-gaussien acyclique (LiNGAM) est une méthode de découverte causale qui identifie la structure causale complète sous l'hypothèse que les données suivent un modèle d'équations structurelles linéaires avec des termes d'erreur non gaussiens [23]. Contrairement à PC ou FCI, LiNGAM peut déterminer la direction de la causalité sans recourir aux tests d'indépendance conditionnelle. Toutefois, sa validité repose sur des hypothèses fortes, incluant la linéarité, la suffisance causale et la non-gaussianité du bruit une combinaison qui limite souvent son applicabilité aux jeux de données réels comportant des relations catégorielles ou non linéaires.

2.5.5 DAG-GNN

DAG-GNN (Directed Acyclic Graph – Graph Neural Network) combine des autoencodeurs variationnels avec des réseaux de neurones graphiques (GNN) pour apprendre conjointement la structure causale et un modèle génératif des données [24]. Un GNN est un type de réseau de neurones conçu pour fonctionner sur des données structurées en graphes, permettant une transmission de messages entre les nœuds pour capturer les dépendances relationnelles. Le modèle DAG-GNN suppose que le processus de génération des données peut être capturé par un modèle latent variationnel avec des priors gaussiens, et que la structure causale est acyclique et entièrement observable. L'efficacité de la méthode dépend de ces hypothèses, en particulier de l'absence de confusion non observée et de la capacité du GNN utilisé à apprendre correctement la structure.

2.6 Discussion

Les méthodes classiques, telles que les variables instrumentales, offrent des bases théoriques solides, mais elles se heurtent souvent à des limites dans les contextes à haute dimension ou purement observationnels [1]. Les approches issues de l'apprentissage automatique relèvent ces défis avec davantage de flexibilité, mais au prix d'une interprétabilité plus faible et de dépendances à des hypothèses fortes comme l'ignorabilité [17].

Des travaux récents soulignent l'intérêt de combiner ces différentes approches, notamment pour estimer non seulement les effets moyens, mais aussi les effets marginaux c'est-à-dire comment de petites variations du traitement influencent localement les résultats [6, 21, 4].

Le cadre PEACE s'inscrit dans cette dynamique en mobilisant des outils probabilistes et variationnels pour capturer à la fois la variabilité causale globale et locale, tout en s'adaptant à la structure des données. Le chapitre suivant formalise cette méthode et en détaille les principaux fondements théoriques.

CHAPITRE 3
FONDEMENTS VARIATIONNELS ET MÉTHODE PEACE

3.1 Variation totale : définitions et interprétations

La **variation totale** (VT) d'une fonction est une mesure de *combien* cette fonction oscille sur son domaine. Elle additionne, en valeur absolue, toutes les augmentations et diminutions locales ; ainsi, plus un signal est « accidenté », plus sa variation totale est grande [25, 26, 27]. Dans le cas discret, plusieurs variantes ont été proposées [28, 29].

3.1.1 Cas univarié – définition classique

Considérons d'abord une fonction réelle $g : [a, b] \to \mathbb{R}$. La VT univariée se définit comme la *borne supérieure* des pas de variation calculés sur *toutes* les partitions finies de l'intervalle :

$$\mathrm{TV}(g) \;=\; \sup_{\mathcal{P}} \sum_{k=1}^{N} \big|g(x_k) - g(x_{k-1})\big|, \quad \mathcal{P} = \{a = x_0 < \cdots < x_N = b\}.$$

— g : fonction réelle définie sur l'intervalle $[a, b]$;
— $a, b \in \mathbb{R}$: bornes de l'intervalle de définition de g ;
— $\mathcal{P}$: partition finie de l'intervalle $[a, b]$, soit l'ensemble de points $\{x_0, x_1, \ldots, x_N\}$ tels que $a = x_0 < x_1 < \cdots < x_N = b$;
— sup : borne supérieure prise sur toutes les partitions possibles $\mathcal{P}$;

Fonction dérivable. Si g est continûment dérivable sur $[a, b]$, i.e., $g \in C^1([a, b])$, alors, lorsque la partition se raffine (le nombre de points N augmente et le pas tend vers zéro), la somme discrète, quand $N \to \infty$, tend vers l'intégrale suivante :

$$\mathrm{TV}(g) = \int_a^b |g'(t)|\, dt,$$

— $g'(t)$: dérivée de g en t, représentant le taux de variation instantané ;
— $t \in [a, b]$: variable d'intégration ;
— $|g'(t)|$: valeur absolue de la pente locale de g ;
— $C^1([a, b])$: fonctions dérivables sur $[a, b]$ dont la dérivée g' est continue.

3.1.2 Cas multivarié – extension variationnelle

Pour $g : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, la généralisation utilise des *champs vectoriels test* et leur divergence.

Principe détaillé. L'idée est d'évaluer la « capacité » de la fonction g à engendrer un flux à travers l'espace Ω. Pour cela, on choisit un *champ test* $\varphi : \Omega \to \mathbb{R}^n$ dont la divergence $\operatorname{div} \varphi$ joue le rôle de *compteur local de flux* :

— Là où $\operatorname{div}\varphi(x) > 0$, le champ φ diverge (flux sortant) et $g(x)$ y contribue positivement.
— Là où $\operatorname{div}\varphi(x) < 0$, le champ converge (flux entrant) et $g(x)$ y contribue négativement.

En pondérant globalement ces contributions par $g(x)$, on mesure l'ampleur totale des variations de g :

$$\mathrm{TV}_n(g) \;=\; \sup_{\substack{\varphi\in C_c^1(\Omega,\mathbb{R}^n)\\ \|\varphi\|_\infty\le 1}} \int_\Omega g(x)\; \underbrace{\operatorname{div}\varphi(x)}_{\substack{\text{mesure locale}\\ \text{de flux}}}\; dx.$$

— $C_c^1(\Omega,\mathbb{R}^n)$: ensemble des *champs test* $\varphi = (\varphi_1,\dots,\varphi_n)$ continûment différentiables (C^1) et à support compact dans Ω (ils s'annulent en bordure et en-dehors).
— $\|\varphi\|_\infty = \sup_{x\in\Omega}|\varphi(x)|$: norme infinie, garantissant que chaque champ test reste borné en amplitude par 1.
— sup : on maximise sur *tous* les champs test admissibles pour capter la variation la plus grande possible.
— $\operatorname{div}\varphi(x) = \sum_{i=1}^n \frac{\partial\varphi_i}{\partial x_i}(x)$: divergence, c'est-à-dire la somme des flux élémentaires sortant le long de chaque direction.
— *Remarque importante :* la fonction g elle-même n'a pas besoin d'être dérivable. Elle doit simplement être de *variation bornée* (BV), c'est-à-dire admettre une valeur finie de $\mathrm{TV}_n(g)$, même en présence de discontinuités.

Si $g \in C^1$:

$$\mathrm{TV}_n(g) = \int_\Omega |\nabla g(x)|\, dx.$$

3.2 Effet causal direct et effet total

Ces définitions, déjà introduites au Chapitre 2, sont rappelées ici pour clarifier le cadre dans lequel s'inscrit la méthode PEACE.

Définition 3.2.1 (Effet causal total) *Variation de Y entre* $\mathrm{do}(X = x_0)$ *et* $\mathrm{do}(X = x_1)$*, toutes voies confondues.*

Définition 3.2.2 (Effet causal direct) *Variation de Y pour le même changement de X, en* bloquant *explicitement tout médiateur.*

3.3 Notions préliminaires

Identifiabilité : L'*identifiabilité* désigne la propriété selon laquelle un effet causal ou, plus généralement, une fonctionnelle causale peut être exprimée de façon unique comme une fonction de la distribution observée des données, compte tenu des hypothèses du modèle causal [2, 30].

Support d'une fonction. Le *support* d'une fonction φ, noté $\mathrm{Supp}(\varphi)$, est défini comme la fermeture topologique de l'ensemble des points $x \in V$ pour lesquels $\varphi(x) \neq 0$ dans V [4, Sec. 2].

3.4 Vue d'ensemble de la méthode PEACE

La méthode *Probabilistic Easy Variational Causal Effect (PEACE)*, proposée par U. Faghihi et A. Saki [4], introduit une approche variationnelle probabiliste pour mesurer l'effet causal direct de la variable de traitement X sur la sortie Y, en tenant compte de la rareté ou de l'abondance des valeurs de traitement. PEACE étend les idées de la variation totale multivariée dans un espace pondéré par les probabilités.

Cette intuition s'appuie sur la représentation visuelle de la divergence et de l'orientation dans la formulation continue de PEACE, illustrée à la Figure 3.1. Celle-ci met en évidence la manière dont la variation totale pondérée peut être comprise à travers une analogie géométrique.

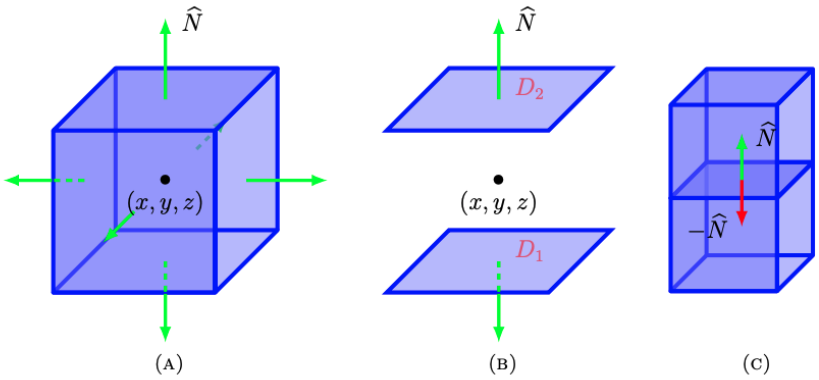


FIGURE 3.1 – Représentation visuelle de la divergence (A) et de l'orientation (B, C) dans la formulation continue de PEACE. Adaptée de la Figure 1 de [4].

L'idée clé est d'utiliser la notion de **variation totale pondérée**, inspirée à la fois des mathématiques (variation totale) et de la physique (flux vectoriel), pour quantifier la sensibilité de Y aux perturbations infinitésimales de X, conditionnellement à Z.

1. **Modèle structurel.** On considère le modèle fonctionnel

$$Y = g(X, Z),$$

où :

— $X \in \mathbb{R}^n$ représente la variable de traitement (qui peut être multidimensionnelle),
— $Z \in \mathbb{R}^m$ est l'ensemble des variables confondantes,
— g est une fonction régulière (C^1) mais non paramétrée, décrivant le lien structurel entre les variables.

2. **Identifiabilité.** Dans le cadre causal de la méthode PEACE, la composante interventionnelle est identifiable comme :

$$g_{\text{in}}(x, z) = \mathbb{E}[Y \mid \text{do}(X) = x, Z = z],$$

C'est-à-dire l'espérance de Y lorsqu'on intervient pour fixer la variable $X = x$, tout en conditionnant sur $Z = z$. Cette quantité exprime l'effet causal de X sur Y dans le contexte $Z = z$.

— $\mathrm{do}(X) = x$: notation d'intervention [2], désignant une manipulation exogène de X plutôt qu'une simple observation.

3. **Critère d'identifiabilité sans graphe causal.** En l'absence de DAG, PEACE et ses dérivés (comme la PEACE Moyenne ou les PEACE Positif et Négatif) doivent satisfaire un critère d'identifiabilité tel que l'ignorabilité ou l'ignorabilité conditionnelle. Cette hypothèse, formulée dans le cadre des résultats potentiels [31, 1], postule que les résultats potentiels $Y(x)$, c'est-à-dire la valeur que prendrait la variable de sortie Y si le traitement X était fixé à x, sont indépendants du traitement effectivement reçu conditionnellement aux covariables Z :

$$Y(x) \perp\!\!\!\perp X \mid Z.$$

Sous cette condition, la composante interventionnelle $g_{\text{in}}(x, z)$, représentant l'espérance de Y sous intervention, devient identifiable comme une espérance conditionnelle observable :

$$g_{\text{in}}(x, z) = \mathbb{E}[Y \mid X = x, Z = z],$$

ce qui permet l'estimation des scores PEACE à partir des données observationnelles, même sans modèle graphique explicite [4, Sec. 5].

4. **Variation interventionnelle.** L'idée centrale est de mesurer la variation totale de la fonction $x \mapsto g_{\text{in}}(x, z)$, en pondérant les contributions locales par $f_{X|Z}(x \mid z)^{2d}$, où $f_{X|Z}$ est la densité conditionnelle de X sachant Z et $d \geq 0$ est un paramètre qui ajuste le « zoom » sur les zones denses ou rares. La quantité définie est :

$$\mathrm{PIEV}_d^z(X \to Y) = \sup_{\substack{\varphi \in C_c^1(U, \mathbb{R}^n) \\ |\varphi(x)| \leq f_{X|Z}(x|z)^{2d}}} \int_U g_{\text{in}}(x, z)\ \mathrm{div}\, \varphi(x)\, dx,$$

où $U \subseteq \mathbb{R}^n$ désigne le support de X (voir Définition 3.3).

Enfin, l'effet causal est obtenu en moyennant sur la distribution de Z :

$$\mathrm{PEACE}_d(X \to Y) = \mathbb{E}_Z \left[\mathrm{PIEV}_d^Z(X \to Y) \right].$$

Exemple illustratif (intervention). Pour illustrer concrètement la distinction entre observation et intervention, considérons un exemple simple (Figure 3.2). Soit $T \in \{0, 1, \ldots, 10\}$ le *nombre de séances de sport par semaine*, C la *pression au travail*, et Y le *niveau de stress* sur une échelle de 0 à 20. Il s'agit ici d'un exemple illustratif, mais la littérature scientifique montre bien que l'activité physique régulière contribue à réduire le stress psychologique, même en contexte de forte pression professionnelle [32].

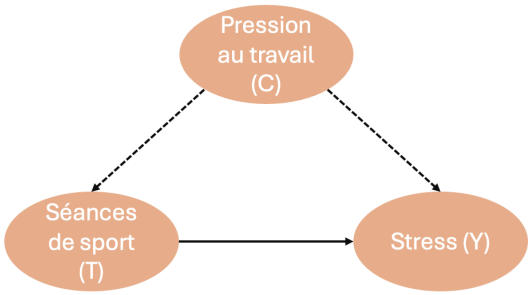


FIGURE 3.2 – Le graphe causal illustre l'effet direct des séances de sport (T) sur le stress (Y), avec la pression au travail (C) comme facteur de confusion.

Sans intervention, le modèle observationnel s'écrit :

$$Y = g(T, C) = a_T + C,$$

où aT désigne le niveau de stress de base associé à T. La différence entre deux niveaux t_0 et t_1 est alors :

$$\big|\mathbb{E}[g(t_1, C)] - \mathbb{E}[g(t_0, C)]\big| = \big|(at_1 + \mathbb{E}[C]) - (at_0 + \mathbb{E}[C])\big| = |a(t_1 - t_0)|.$$

En revanche, dans le cas d'une intervention fixant exogènement le nombre de séances de sport à $T = t$, le modèle devient :

$$Y = g_{\text{in}}(T, C) = a_T + C, \quad T = t,$$

où T et C sont indépendants. La différence causale s'écrit :

$$\big|\mathbb{E}[g_{\text{in}}(t_1, C)] - \mathbb{E}[g_{\text{in}}(t_0, C)]\big| = \big|(a_{t_1} + \mathbb{E}(C)) - (a_{t_0} + \mathbb{E}(C))\big| = |a_{t_1} - a_{t_0}|.$$

Cet exemple met en évidence le fait qu'avant intervention, la variation de Y dépend à la fois du sport et du contexte, tandis qu'après intervention, l'effet propre du sport est isolé.

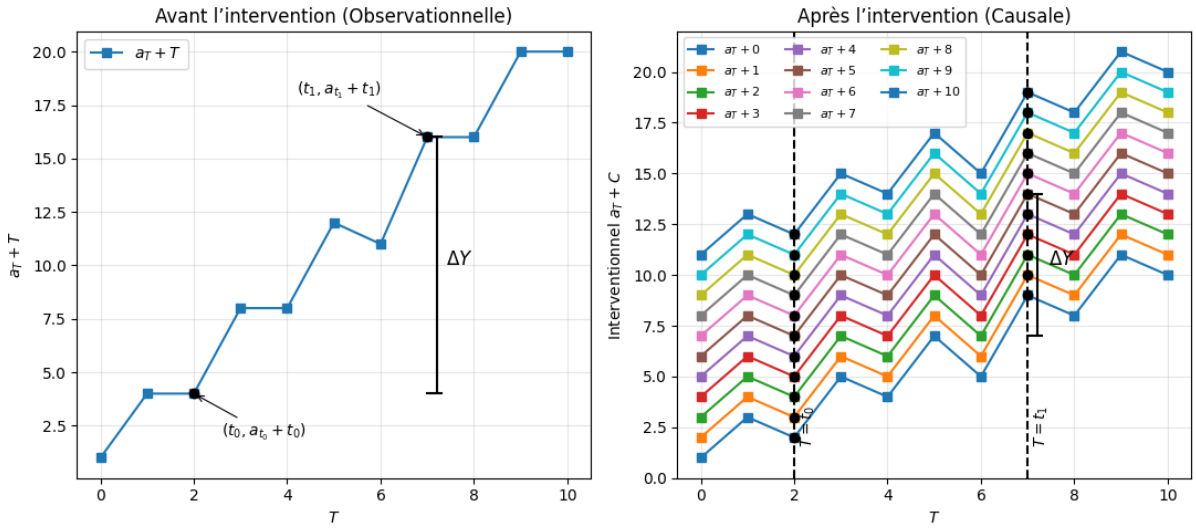


FIGURE 3.3 – Le graphique de gauche représente le niveau de stress Y avant toute intervention, où T correspond au nombre de séances de sport par semaine, C à la pression au travail, et Y au niveau de stress. Dans ce cas, ΔY est défini comme la différence absolue $|\mathbb{E}[g(t_1, C)] - \mathbb{E}[g(t_0, C)]|$. Le graphique de droite illustre Y après l'intervention. Étant donné une intervention fixant le nombre de séances de sport à $T = t$, les valeurs attendues aux points d'intersection avec la ligne verticale en $T = t$ sont représentées par $\mathbb{E}(g_{\text{in}}(t, C))$. Ici, ΔY est défini comme $|\mathbb{E}[g_{\text{in}}(t_1, C)] - \mathbb{E}[g_{\text{in}}(t_0, C)]|$, en supposant que $\mathbb{E}(g_{\text{in}}(t_0, C))$ et $\mathbb{E}(g_{\text{in}}(t_1, C))$ correspondent respectivement aux grands points orange et vert. Inspiré de [33].

3.4.1 Définition intuitive

Cas discret. Lorsque X est discret et unidimensionnel, avec support $\{x_0, \ldots, x_\ell\}$, PEACE se formule :

$$\text{PIEV}_d^z = \sum_{i=1}^{\ell} |g_{\text{in}}(x_i, z) - g_{\text{in}}(x_{i-1}, z)| \, P(x_i \mid z)^d \, P(x_{i-1} \mid z)^d,$$

où chaque transition $x_{i-1} \to x_i$ est pondérée par la probabilité conditionnelle $P(x_i \mid z)$.

Cas continu. Le cas général :

$$Y \;=\; g(X, Z), \qquad (X, Z) \in \Omega \subseteq \mathbb{R}^2, \qquad \Omega_z \;=\; \big\{x \in \mathbb{R} \mid (x, z) \in \Omega\big\}.$$

Pour un niveau fixé de la variable de contexte $Z = z$, la *Probabilistic Intervention Effect Variation* s'écrit

$$\text{PIEV}_d^z(X \to Y) \;=\; \int_{\Omega_z} \Big|\partial_x g_{\text{in}}(x, z)\Big| \, f_{X|Z}(x \mid z)^{2d} \, dx,$$

où $f_{X|Z}$ est la densité conditionnelle de X sachant Z et $d \geq 0$ module l'importance accordée aux régions denses ou rares.

L'effet causal se définit alors par moyennage :

$$\text{PEACE}_d(X \to Y) \;=\; \mathbb{E}_Z\big[\text{PIEV}_d^Z(X \to Y)\big]$$

— $\partial_x g_{\text{in}}(x, z)$ mesure la sensibilité locale de Y aux variations de X pour un contexte $Z = z$;

— $f_{X|Z}(x \mid z)$ est la densité conditionnelle de X ;
— Le paramètre $d \geq 0$ module l'importance accordée aux transitions : pour $d \approx 0$, toutes les transitions sont pondérées de manière équitable, tandis qu'une augmentation de d privilégie les plus probables (macro-effets) au détriment des événements rares.
— Ω_z est la tranche du domaine Ω associée à la valeur z.

Rôle du paramètre d. Le paramètre d agit comme un *zoom* permettant d'ajuster la focalisation de la mesure. Autrement dit le *degré* d agit en élevant la densité conditionnelle $f_{X|Z}(x)$ à la puissance $2d$:

— $d \approx 0$: traite toutes les transitions de façon équitable, y compris les événements rares (micro-effets),
— Augmentation de la valeur d : privilégie les transitions les plus probables, mettant l'accent sur les effets dominants (macro-effets).

Pourquoi parle-t-on d'« effet causal » ? L'intégrale qui définit PEACE peut être justifiée à partir du *do-calculus* ou du cadre des *résultats potentiels* : de manière intuitive, elle agrège les dérivées directionnelles qui apparaîtraient si l'on perturbait légèrement X tout en maintenant les covariables Z constantes.

3.4.2 PEACE directionnel et variantes moyennes

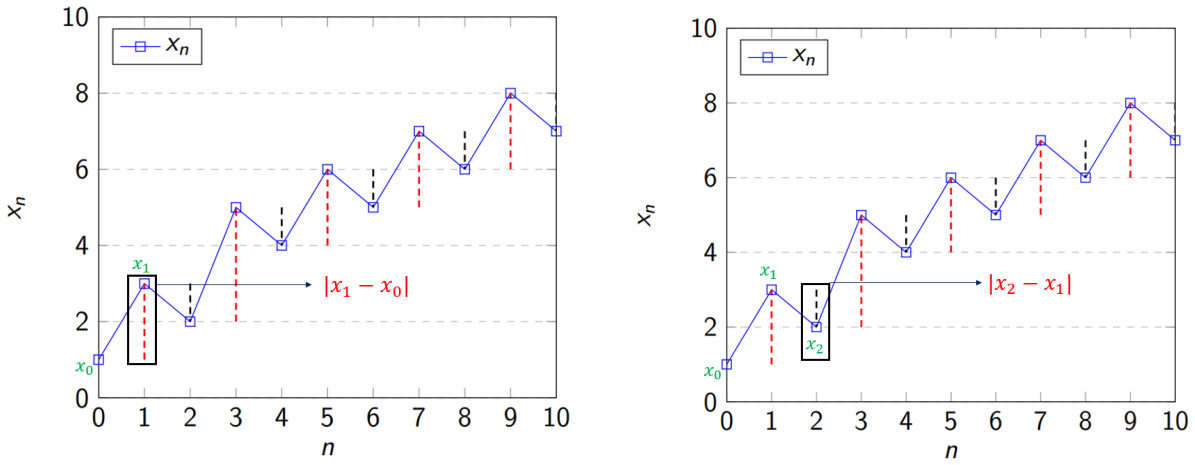


FIGURE 3.4 – Illustrations intuitives des contributions directionnelles. (A) montre une contribution positive où le traitement augmente le résultat. (B) montre une contribution négative où le traitement diminue le résultat. Ces visualisations reflètent comment PEACE décompose les variations causales. Adapté du carnet explicatif PEACE [33].

Bien que $\text{PEACE}_d(X \to Y)$ quantifie l'intensité globale, elle ne renseigne pas sur la direction (augmentation ou diminution de Y). Pour cela, dans le cas où X et Y sont unidimensionnels, la méthode PEACE introduit des décompositions directionnelles, illustrées à la Figure 3.4.

Effets positifs et négatifs (amplitude totale) Pour un $Z = z$ fixé :

$$\mathcal{PIEV}_d^{z,+}(X \to Y) = \int \left(\frac{\partial g_{\text{in}}}{\partial x}(x, z) \right)^+ f_{X|Z}^{2d}(x \mid z)\, dx, \tag{3.1}$$

$$\mathcal{PIEV}_d^{z,-}(X \to Y) = \int \left(\frac{\partial g_{\text{in}}}{\partial x}(x, z) \right)^- f_{X|Z}^{2d}(x \mid z)\, dx, \tag{3.2}$$

où :

— $(\cdot)^+ = \max(\cdot, 0)$,

— $(\cdot)^- = \max(-\cdot, 0)$,

— $f_{X|Z}(x \mid z)$ est la densité conditionnelle de X sachant $Z = z$.

Agrégées sur Z, les composantes directionnelles deviennent :

$$\text{PEACE}_d^+(X \to Y) = \mathbb{E}_Z \left[\mathcal{PIEV}_d^{z,+}(X \to Y) \right], \tag{3.3}$$

$$\text{PEACE}_d^-(X \to Y) = \mathbb{E}_Z \left[\mathcal{PIEV}_d^{z,-}(X \to Y) \right]. \tag{3.4}$$

La variation causale totale se décompose ainsi :

$$\text{PEACE}_d(X \to Y) = \text{PEACE}_d^+(X \to Y) + \text{PEACE}_d^-(X \to Y). \tag{3.5}$$

PEACE moyen (effet directionnel net). Contrairement à la variation totale, le *PEACE moyen* conserve le signe de la dérivée et le normalise par la masse totale des poids $f_{X|Z}^{2d}$. Le score permet ainsi d'indiquer si, en moyenne, les interventions sur X tendent à faire croître $(+)$ ou décroître $(-)$ la variable Y.

Dans le cas **continu univarié**, pour un contexte fixé $Z = z$, on a :

$$\text{PEACE}_d^{\text{mean}}(X \to Y) = \mathbb{E}_Z \left[\frac{\int \partial_x g_{\text{in}}(x, z)\, f_{X|Z}(x \mid z)^{2d}\, dx}{\int f_{X|Z}(x \mid z)^{2d}\, dx} \right] \tag{3.6}$$

Il se décompose en parties positives et négatives :

$$\text{PEACE}_d^{\text{mean}}(X \to Y) = \text{PEACE}_d^{\text{mean},+}(X \to Y) + \text{PEACE}_d^{\text{mean},-}(X \to Y) \tag{3.7}$$

$$\text{PEACE}_d^{\text{mean},+}(X \to Y) = \mathbb{E}_Z \left[\frac{\int \left(\partial_x g_{\text{in}}(x, z) \right)^+ f_{X|Z}(x \mid z)^{2d}\, dx}{\int f_{X|Z}(x \mid z)^{2d}\, dx} \right] \tag{3.8}$$

$$\text{PEACE}_d^{\text{mean},-}(X \to Y) = \mathbb{E}_Z \left[\frac{\int \left(\partial_x g_{\text{in}}(x, z) \right)^- f_{X|Z}(x \mid z)^{2d}\, dx}{\int f_{X|Z}(x \mid z)^{2d}\, dx} \right] \tag{3.9}$$

où $(\cdot)^+$ et $(\cdot)^-$ désignent respectivement les parties positive et négative de la dérivée. Un score positif indique que les interventions sur X augmentent Y en moyenne, et inversement.

Résumé Cette décomposition permet d'isoler :

— Les transitions causales croissantes et décroissantes (+, −),
— L'effet net signé (PEACE moyen),
— La capacité totale d'impact causal (PEACE absolu).

Un récapitulatif de ces différentes variantes est présenté dans le Tableau 3.1, qui synthétise leurs caractéristiques principales en termes d'amplitude et de sensibilité au signe.

Métrique	**Amplitude**	**Sensibilité au signe**
PEACE_d	Totale (absolue)	Non
PEACE_d^+, PEACE_d^-	Totale par direction	Non
$\text{PEACE}_d^{\text{mean}}$	Moyenne signée (uniquement transitions locales permises)	Oui
$\text{PEACE}_d^{\text{mean},+}$, $\text{PEACE}_d^{\text{mean},-}$	Moyenne par direction	Oui

Table 3.1 – Résumé des variantes directionnelles de PEACE

3.4.3 Propriétés théoriques principales

— **Fondements mesure-théoriques** : PEACE s'étend à une mesure régulière de Borel, assurant l'additivité et la monotonie (Théorème 4.4 de [4]).
— **Formule explicite via les gradients** : lorsque $g \in C^1$, PEACE admet la forme intégrale ci-dessus (Théorème 4.6 de [4]).
— **Caractérisation de l'effet nul** : PEACE est nul si et seulement si g est (localement) constant sur le support de X (Corollaire 4.8 [4]).
— **Invariance par transformation** : PEACE est invariant par reparamétrisation isométrique de X (Proposition 4.9 [4]).
— **Identifiabilité** : sous conditions de séparabilité et d'exogénéité, PEACE s'exprime par la dérivée de l'espérance conditionnelle :

$$\frac{\partial \mathbb{E}[Y \mid X, Z]}{\partial X},$$

ce qui le rend identifiable à partir des données observationnelles (Théorème 5.1 [4]).

— Y : variable de sortie (ou réponse) ;
— X : variable de traitement ;
— Z : ensemble des variables de confusion (ou contexte observé) ;
— $\mathbb{E}[Y \mid X, Z]$: espérance conditionnelle de Y sachant X et Z ;
— $\frac{\partial}{\partial X}$: dérivée partielle par rapport à la variable de traitement X, indiquant la sensibilité locale de Y à une variation infinitésimale de X.

3.4.4 Extensions et applications

— **PEACE positif / négatif** : permet de décomposer l'effet causal en contributions *bénéfiques* (PEACE^+) et *néfastes* (PEACE^-) [4, Sec. 7].

— **Variables instrumentales** : s'adapte à des scénarios avec confondeurs non mesurés via les conditions d'identifiabilité par IV [4, Sec. 5].
— **Domaines d'application** :
  — **Apprentissage par renforcement** : régularisation des réseaux de neurones profonds (DQN) pour respecter les structures causales [4, Sec. 13].
  — **Diagnostic médical** : différenciation des sous-types de tumeurs en mesurant l'influence causale des caractéristiques [4, Sec. 13].
  — **Économie comportementale** : évaluation des stratégies d'incitation à la réabonnement en présence de confusions complexes [4, Sec. 13].

Conclusion. PEACE constitue une approche flexible et mathématiquement rigoureuse pour quantifier les effets causaux, offrant un compromis entre robustesse aux événements rares et focalisation sur les phénomènes dominants. Son fondement variationnel assure une interprétabilité forte et une grande adaptabilité aux données.

3.4.5 Comparaison avec d'autres métriques causales

Le tableau 3.2 résume les principales différences entre PEACE et d'autres approches classiques :

Méthode	Principe	Hypothèses	Limites
ATE	Effet moyen du traitement : $\mathbb{E}[Y(1) - Y(0)]$	Ignorabilité, absence de confondeurs non observés	Sensible aux variables latentes
CATE	Effet moyen *conditionnel* : $\tau(z) = \mathbb{E}[Y(1) - Y(0) \mid Z = z]$	Idem ATE + stratification sur Z	Dépend de la qualité du modèle de stratification
SHAP	Décomposition de l'effet des caractéristiques dans un modèle prédictif	mesure d'importance corrélationnelle	Non causal, sensible à la colinéarité entre variables
PEACE	L'effet causal local pondérée par la densité conditionnelle $f_{X\|Z}$	Prise en compte des cas rares et fréquents, critère d'identifiabilité (voir l'élément **2.**)	Paramétrage du degré d

Table 3.2 – Comparaison de PEACE avec ATE, CATE et SHAP

Ce que PEACE apporte par rapport aux cadres classiques

Les cadres classiques de l'inférence causale exposées au Chapitre 2, tels que celui de Rubin (résultats potentiels) et celui de Pearl (SCM et do-calculus), ont permis d'importants progrès mais présentent certaines limitations :

— **Rubin (résultats potentiels + ignorabilité)**

    - *Covariables manquantes* : si un confondeur n'est pas mesuré, l'ignorabilité est violée.
    - *Effets sur sous-populations rares* : des strates peu fournies peuvent biaiser fortement l'ATE.
- **Pearl (SCM + backdoor/frontdoor)**
    - *Graphe exact requis* : toute erreur de structure (confounders latents...) compromet l'identification.
    - *Complexité du do-calculus* : pour des graphes denses, trouver les bons ensembles devient ardu.

La méthode **PEACE** ne rejette pas ces cadres, mais les étend en intégrant des éléments issus des deux approches et en introduisant des innovations majeures.

- **Poids probabiliste adaptatif :** PEACE introduit une pondération nouvelle via le terme d.
    - Ce poids reflète la fréquence des transitions de traitement dans la population observée.
    - Le paramètre $d \geq 0$ permet de moduler l'analyse :
        - $d \approx 0$: toutes les transitions ont le même poids, ce qui le rend sensible aux *micro-effets*, même rares.
        - Augmentation de la valeur d : on concentre l'analyse sur les transitions les plus probables , ce qui met en avant les *macro-effets* dominants.
- **Effets directionnels :** Contrairement aux approches classiques qui donnent un effet net unique, PEACE propose :
    - *Positive PEACE* : met en évidence les effets bénéfiques.
    - *Negative PEACE* : isole les effets délétères.
- **Une métrique générique :** Fondée sur la variation totale, la métrique *PEACE* s'applique à toute nature de traitement binaire, catégoriel, discret ou continu. Cette granularité dépasse les métriques binaires comme l'*Average Treatment Effect* (ATE) [14], ainsi que les métriques dérivées telles que l'*Average Derivative Effect* (ADE), qui se limite à la pente signée moyenne d'un traitement continu [1]
- **Ordre des valeurs du traitement :** Le calcul des effets variationnels dans PEACE repose sur l'existence d'un ordre défini entre les modalités du traitement.

3.5 Intégration avec les modèles prédictifs

Dans le cadre de PEACE, la fonction $g(X, Z)$, qui décrit comment les variables de traitement et de confusion déterminent conjointement le résultat, est supposée différentiable. Dans le développement théorique de la méthode, cette fonction est considérée comme connue [4].

Dans ce travail, nous adoptons la même structure formelle et considérons le processus d'estimation causale comme reposant sur l'accès à une telle fonction. Pour chaque variable de traitement X_j, l'effet causal est calculé en observant comment la sortie de $g(X, Z)$ varie lorsque X_j est perturbée, les autres variables étant maintenues constantes.

1. Les ADE et leurs versions pondérées WADE estiment $\mathbb{E}[w(A, Z)\, \partial_A \mu(A, Z)]$, c'est-à-dire une moyenne pondérée de la dérivée locale de la fonction de régression. Comme le rappellent Hines *et al.* : « *Weighted average derivative effects (WADEs) are non-parametric estimands with uses in economics and causal inference.* » [34].

Ces variations dans les résultats sont ensuite agrégées sur le domaine d'entrée selon la formulation de PEACE, avec une pondération déterminée par la densité de données estimée. Cela permet de calculer des scores causaux qui reflètent à la fois la sensibilité à l'intervention et la structure de la distribution observée.

3.6 Évaluation des estimations causales

Comme présenté dans la formulation originale de Faghihi et Saki [4], l'évaluation des estimations causales dans le cadre de PEACE repose sur une analyse de sensibilité et une interprétation des dynamiques des scores pour différentes variables de traitement. Étant donné que la méthode est variationnelle et indépendante du modèle, elle ne dépend pas de valeurs causales « vérité terrain » pour la validation. L'objectif est plutôt d'examiner le comportement interne des scores PEACE et d'évaluer leur robustesse sous différentes hypothèses de modélisation et valeurs de paramètres.

— **Intensité relative des effets :** Les scores PEACE sont calculés pour plusieurs variables de traitement par rapport à une même variable de sortie. Comparer leurs amplitudes permet d'identifier les variables ayant un effet direct plus fort, en tenant compte de la distribution des données.

— **Sensibilité au paramètre de degré (d) :** Afin d'étudier comment la force causale évolue avec la densité des données, le paramètre de degré d est systématiquement varié. Comme l'illustre la Figure 3.5, l'évolution des scores **PEACE** en fonction de d permet d'identifier si les effets sont concentrés dans les régions à forte densité ou répartis plus uniformément.

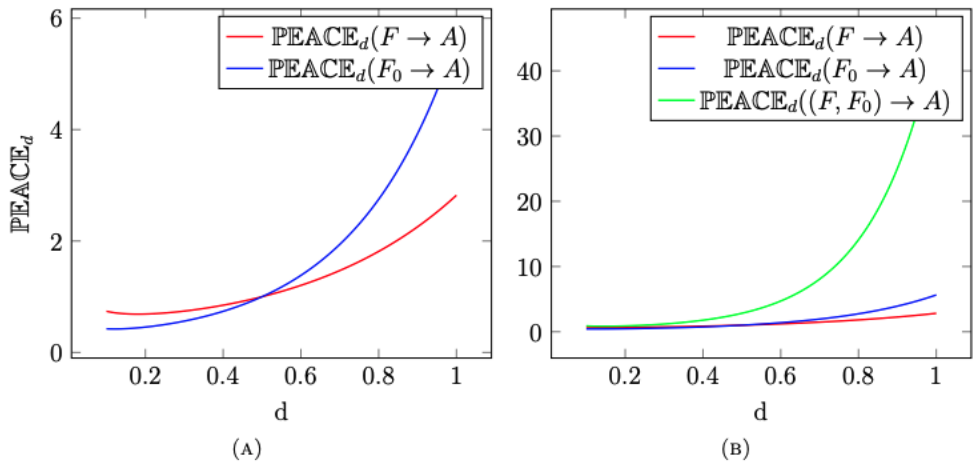


FIGURE 3.5 – Évolution des scores PEACE en fonction de l'augmentation du paramètre de degré d, pour différentes variables de traitement. Le panneau (A) compare l'influence causale des variables F et F_0 sur la variable de sortie A, tandis que le panneau (B) considère leur contribution conjointe (F, F_0). Lorsque d augmente, les scores PEACE croissent plus rapidement, indiquant que les zones les plus denses de la distribution des données dominent l'estimation. Cette évolution illustre comment PEACE s'adapte à la structure des données grâce à la sensibilité à la densité. Adapté de la Figure 7 de [4].

— **Composition et interaction des effets :**

PEACE permet de comparer les effets causaux marginaux et conjoints. Toute divergence entre le score conjoint et la somme des scores marginaux signale qu'il existe une dépendance ou interaction entre les variables de traitement, dont la nature doit ensuite être interprétée à la lumière de la densité des données et du contexte.

— **Interprétation des scores selon le contexte :**

Au lieu d'une validation par référence externe, PEACE met l'accent sur une interprétation contextuelle. La direction, l'intensité et la variabilité des scores sont analysées à la lumière des connaissances du domaine et des mécanismes causaux attendus.

Ce chapitre a survolé les fondements théoriques et méthodologiques du cadre PEACE. Nous avons exposé sa formulation mathématique, précisé les hypothèses sous-jacentes, et décrit comment il s'intègre aux modèles d'apprentissage supervisé. Le chapitre suivant détaillera l'application empirique de PEACE à un ensemble de données synthétique, y compris la préparation des données, l'entraînement des modèles et la mise en œuvre de la procédure d'estimation des effets causaux.

CHAPITRE 4
ÉVALUATION EMPIRIQUE

Ce chapitre présente l'évaluation empirique de la méthode PEACE appliquée à des données de santé synthétiques, en mettant en évidence sa capacité à détecter des effets causaux directs, non linéaires et spécifiques à certaines sous-populations, là où les approches classiques comme l'ATE/CATE ou les modèles neuronaux comme TARNet [6] et DragonNet [7] atteignent leurs limites.

4.1 Description de la base de données

La base de données utilisée dans cette étude est un jeu de données synthétique, librement accessible, publié sur Kaggle par Anthony Therrien [5]. Elle a été conçue pour simuler des tendances réelles observées dans les domaines de la santé, du mode de vie et des études socio-économiques, en mettant l'accent sur des variables potentiellement liées aux résultats en matière de santé mentale et physique.

Ce jeu de données contient **413 768 enregistrements individuels** et **16 variables**, ce qui le rend particulièrement adapté à des analyses statistiques et causales robustes. Chaque observation correspond à un individu unique (une seule ligne par personne, incluant des caractéristiques démographiques comme l'âge) et inclut une combinaison d'informations comportementales et cliniques. Fait important, la base ne présente **aucune valeur manquante**, ce qui élimine le besoin d'imputation et permet de se concentrer directement sur la structure causale.

La variable cible étudiée est la **présence de maladies chroniques**, un indicateur binaire signalant si un individu présente une ou plusieurs pathologies médicales. Cette variable est traitée comme la variable de sortie Y, tandis que les autres attributs sont considérés comme des traitements candidats X_j ou des variables de conditionnement Z, selon le contexte du modèle.

Ce jeu de données comporte **16 variables**, couvrant des informations démographiques, comportementales et cliniques :

- **Nom** : identifiant textuel (non utilisé dans l'analyse)
- **Âge** : variable quantitative continue
- **Revenu annuel (USD)** : variable quantitative continue
- **Nombre d'enfants** : variable discrète ordinale (valeurs de 0 à 4)
- **État civil** : Célibataire, Marié(e), Divorcé(e), Veuf(ve)
- **Niveau d'éducation** : École secondaire, Diplôme d'associé, Licence, Maîtrise, Doctorat
- **Tabagisme** : Fumeur, Ancien fumeur, Non-fumeur
- **Activité physique** : Sédentaire, Modérée, Active
- **Statut professionnel** : En emploi, Sans emploi
- **Consommation d'alcool** : Faible, Modérée, Élevée
- **Habitudes alimentaires** : Saines, Modérées, Mauvaises
- **Qualité du sommeil** : Bonne, Moyenne, Mauvaise
- **Antécédents de troubles mentaux** : Oui, Non

— **Antécédents de toxicomanie** : Oui, Non
— **Antécédents familiaux de dépression** : Oui, Non
— **Maladies chroniques (cible)** : Oui, Non

Ce jeu synthétique ne se limite pas à une simple construction artificielle : il intègre des variables dont l'influence causale sur les maladies chroniques est bien documentée dans la littérature, notamment le statut tabagique [35], le niveau d'activité physique [36], les habitudes alimentaires [37], la qualité du sommeil [38], ou encore l'historique de santé mentale [39]. Comme c'est souvent le cas dans les bases observationnelles en santé [30], ce jeu ne fournit ni générateur explicite ni graphe causal sous-jacent. Cette absence ne constitue pas une limitation, mais reflète les conditions réelles dans lesquelles les méthodes causales sont appliquées. Il constitue ainsi un banc d'essai réaliste pour évaluer la méthode PEACE, dans un environnement contrôlé sans bruit de mesure ni valeurs manquantes tout en reflétant la complexité multidimensionnelle typique des données en santé publique.

4.2 Analyse exploratoire des données

Une analyse préliminaire a été réalisée afin de comprendre la distribution des variables et leur relation avec la variable cible : **Maladies chroniques**.

Répartition des classes

Environ **33 %** des individus de l'échantillon déclarent souffrir d'une maladie chronique, comme illustré dans la Figure 4.1, ce qui traduit un déséquilibre modéré des classes. Cette asymétrie a motivé l'application ultérieure de techniques d'équilibrage telles que SMOTETomek [40] durant la phase de prétraitement.

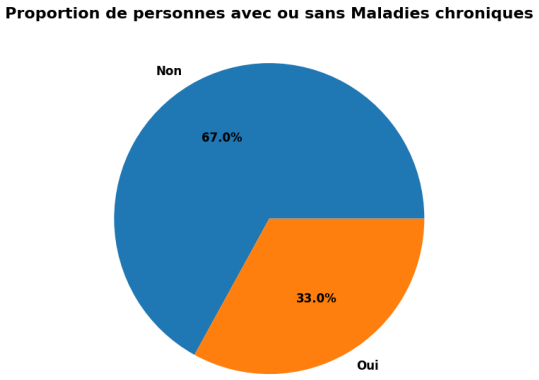
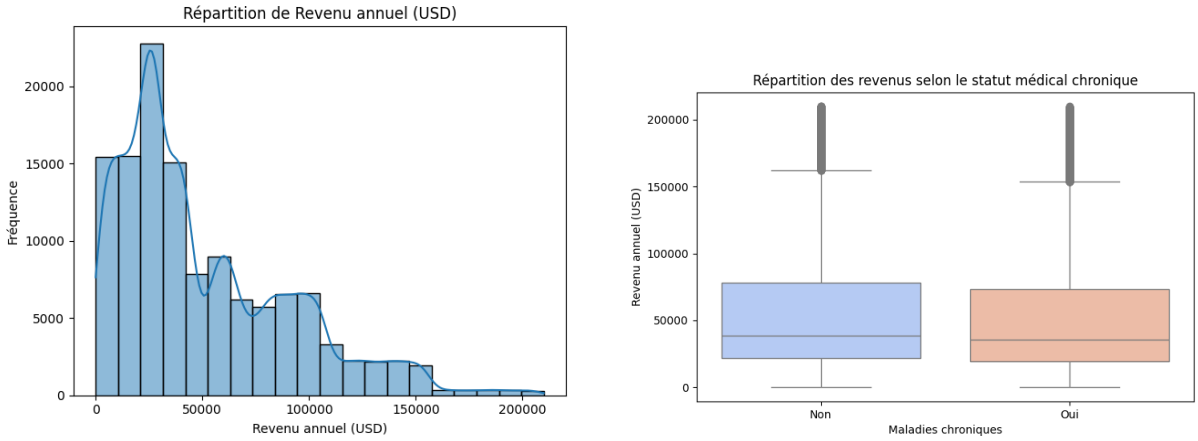


FIGURE 4.1 – Proportion de personnes avec ou sans *Maladies chroniques*. Environ 33 % des individus de l'échantillon présentent au moins une maladie chronique.

Distributions des variables et tendances observées

La majorité des répondants sont *mariés*, *non-fumeurs* et *actifs professionnellement*. Les habitudes rapportées incluent une alimentation *peu saine ou modérément saine*, une *consommation d'alcool modérée* et une *qualité de sommeil moyenne*. Le **revenu annuel** est *fortement asymétrique à droite* : quelques revenus exceptionnellement élevés allongent la distribution vers les grandes valeurs et tirent la moyenne au-dessus de la

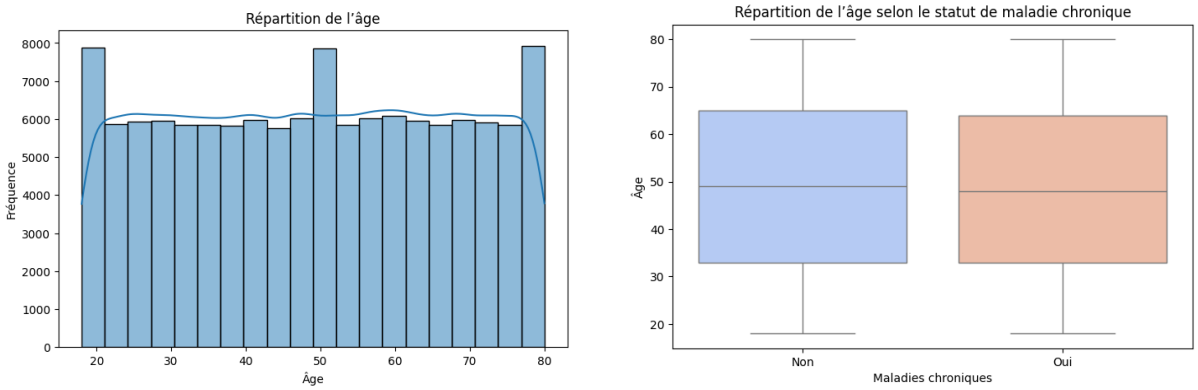
médiane (Figure 4.2 A). Les individus sans maladies chroniques présentent un revenu médian légèrement plus élevé (Figure 4.2 B). Les graphiques et le code d'analyse sont disponibles dans le notebook complémentaire [41].



(A) **(B)**

FIGURE 4.2 – Caractéristiques des revenus dans la base de données. (A) montre la distribution globale, asymétrique vers la droite avec une majorité des individus gagnant moins de 50 000$. (B) compare les revenus selon la présence de maladies chroniques, montrant un revenu médian légèrement plus élevé chez les individus en bonne santé.

La distribution de l'**âge** s'étend sur l'ensemble de l'âge adulte (Figure 4.3 A). La comparaison selon le statut de maladie chronique (Figure 4.3 B) montre des médianes très proches et des boîtes largement chevauchantes ; la différence est faible.



(A) **(B)**

FIGURE 4.3 – Caractéristiques de l'âge. (A) Distribution globale des âges dans l'échantillon, étalée sur l'ensemble de l'âge adulte. (B) Répartition des âges selon la présence de maladies chroniques (*Oui/Non*) : médianes proches et boîtes largement chevauchantes, indiquant une différence faible.

Nous avons ensuite exploré des relations bivariées entre certaines caractéristiques et la prévalence des maladies chroniques. Les tendances visualisées aux Figures 4.4 et 4.5 incluent à la fois des résultats attendus, conformes à la littérature, et des anomalies qui motivent une analyse causale plus approfondie.

— **Tendances attendues :** Être *employé* (rapport Non/Oui [1]= 2,174) et bénéficier d'un *bon sommeil* (rapport Non/Oui = 2,147) sont associés à de meilleurs résultats de santé, ce qui est cohérent avec la littérature [42, 38].

— **Résultats inattendus :** De manière contre-intuitive, les *individus sédentaires* (rapport Non/Oui = 2,219) ou ayant une *alimentation malsaine* (rapport Non/Oui = 2,061) présentent une prévalence comparable ou inférieure de maladies chroniques, en contradiction apparente avec les recommandations sanitaires classiques [36, 37].

— **Effets subtils :** Des corrélations plus faibles que prévu ont été observées pour des variables telles que les *antécédents de santé mentale* (rapport Non/Oui = 2,044) et le *statut tabagique*, suggérant la présence de confusions latentes ou d'interactions non linéaires, malgré des preuves établies de leur lien avec les maladies chroniques [39, 43].

Ces résultats soulignent la nécessité d'une analyse causale rigoureuse. Les simples corrélations peuvent induire en erreur, ce qui justifie l'adoption du cadre PEACE, orienté vers la sensibilité causale au-delà des relations statistiques superficielles.

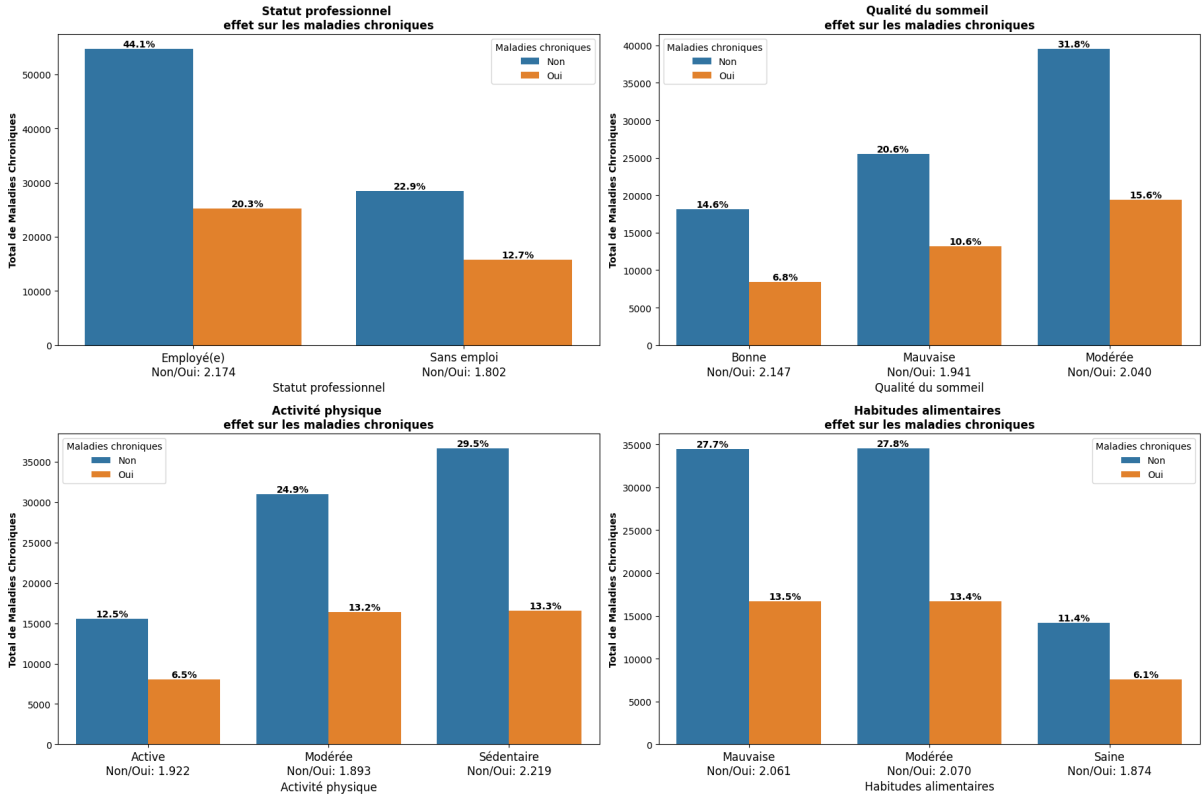


FIGURE 4.4 – Relations bivariées sélectionnées entre les variables individuelles et la prévalence des maladies chroniques. **Haut :** Tendances attendues, avec emploi et bon sommeil associés à une incidence plus faible. **Bas :** Tendances inattendues, modes de vie sédentaires et alimentation peu saine non systématiquement liés à une détérioration de la santé.

1. Le rapport Non/Oui indique le rapport entre la proportion d'individus atteints de maladies chroniques parmi ceux ayant répondu "Non" versus ceux ayant répondu "Oui" à la variable correspondante.

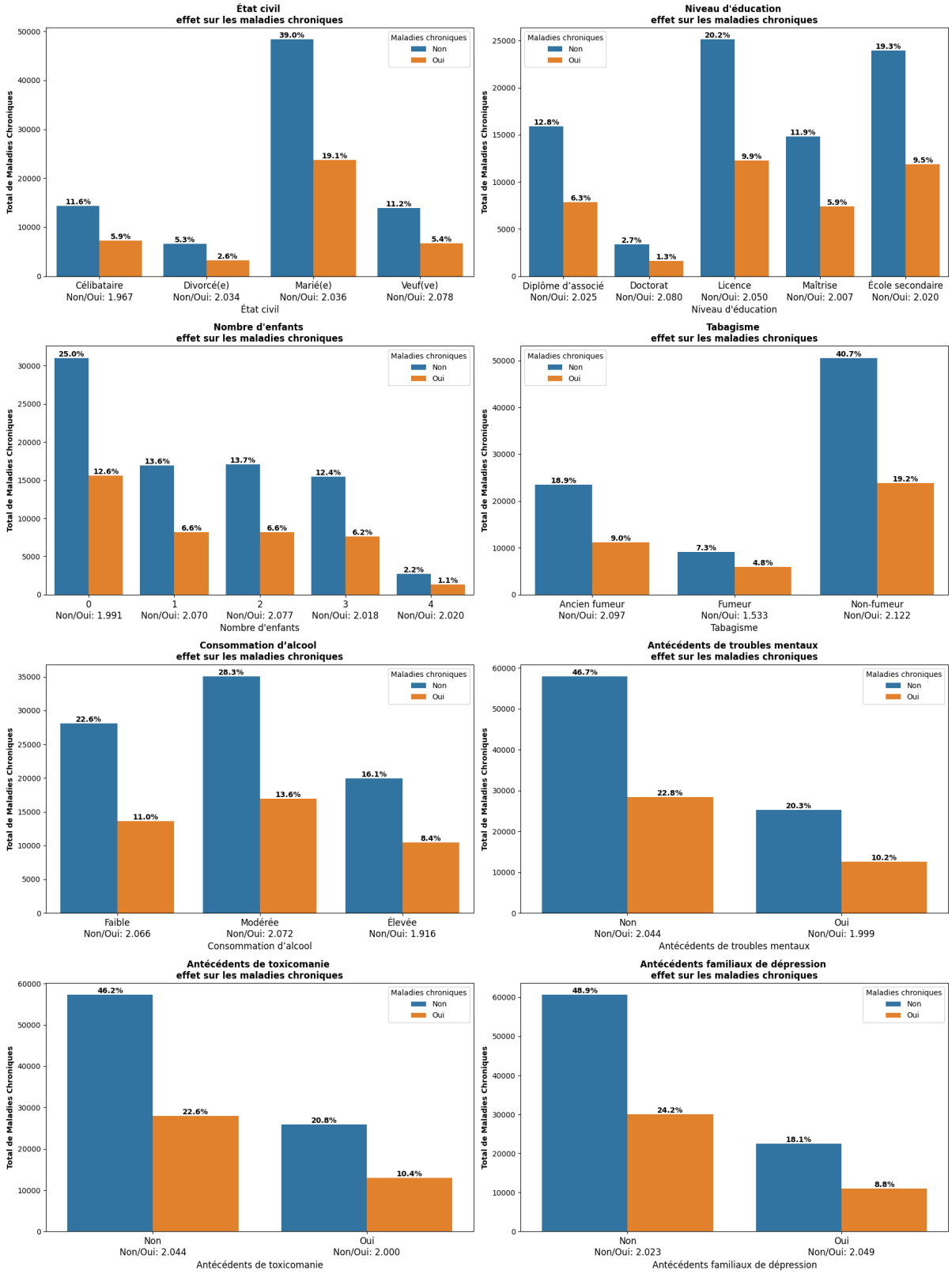


FIGURE 4.5 – Comparaisons bivariées entre la présence de maladies chroniques (*Oui*/*Non*) et plusieurs variables sociodémographiques et cliniques.

4.3 Prétraitement des données

Plusieurs variables catégorielles possèdent un ordre naturel pertinent dans un contexte de santé. Ces variables ont été encodées de manière ordinale afin de préserver leur structure hiérarchique, tout en évitant l'explosion dimensionnelle induite par l'encodage one-hot. Par exemple, une variable ordinale comme le *niveau d'éducation* a été encodée de *Secondaire (1)* à *Doctorat (5)* pour refléter une progression croissante du niveau de formation.

Les types de données ont été optimisés pour des raisons d'efficacité mémoire à l'aide de conversions numériques et de typages catégoriels. La variable cible binaire a été extraite et transformée en vecteur entier.

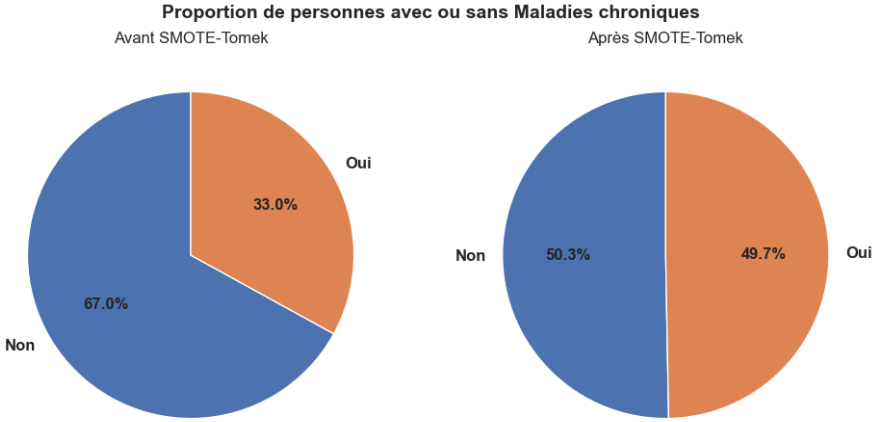


FIGURE 4.6 – Proportion de personnes atteintes ou non de maladies chroniques, avant et après application de SMOTE-Tomek.

Pour corriger le déséquilibre initial de classes (environ 33 % de cas positifs), nous avons appliqué la méthode SMOTE-Tomek [40] uniquement sur les données d'entraînement [2], avec un ratio cible de 0,99 (minorité ≈ majorité). Cette procédure a permis d'équilibrer les classes de manière quasi symétrique, portant la proportion de cas positifs à environ 49,7 % après suppression des doublons frontaliers par les liens de SMOTE-Tomek. Afin de garantir que cette augmentation artificielle n'introduise pas de duplications ou artefacts, nous avons vérifié manuellement un échantillon des instances générées. Les exemples synthétiques présentaient des interpolations crédibles entre observations voisines, sans duplication excessive ni concentration anormale.

2. La méthode SMOTE-Tomek n'a été utilisée que pour l'entraînement des modèles prédictifs. Elle n'intervient à aucun moment dans l'évaluation causale via PEACE, laquelle repose sur la distribution empirique initiale des données, sans sur-échantillonnage.

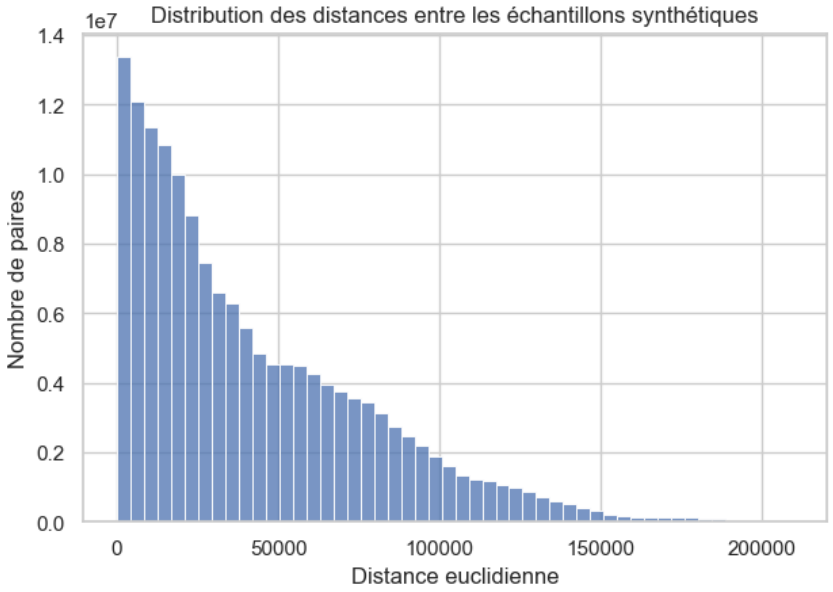


FIGURE 4.7 – Histogramme des distances euclidiennes entre paires de points synthétiques (axe y en $\times 10^7$ paires). Absence de masse en 0 $\Rightarrow$ pas de duplicats exacts ; distances majoritairement faibles à modérées (interpolation locale), avec une queue droite ténue.

Une vérification empirique complémentaire a confirmé l'absence de duplication exacte ou de quasi-duplicats (distance $< 10^{-5}$) parmi les échantillons synthétiques. Nous avons utilisé la distance euclidienne car SMOTE-Tomek génère des valeurs intermédiaires, plaçant les données dans un espace continu où cette mesure de proximité est naturelle. Comme le montre la distribution des distances entre les points générés (voir Figure 4.7), l'algorithme a produit une interpolation fluide, sans concentrations artificielles de points trop proches. De telles concentrations auraient créé des pics de densité non réalistes et biaisé l'estimation par noyau utilisée dans la méthode PEACE.

Enfin, l'ensemble de données a été divisé en deux sous-ensembles d'apprentissage et de test selon une répartition 80/20, en conservant la distribution des classes via un échantillonnage stratifié.

4.4 Analyse statistique de l'importance des variables (SHAP)

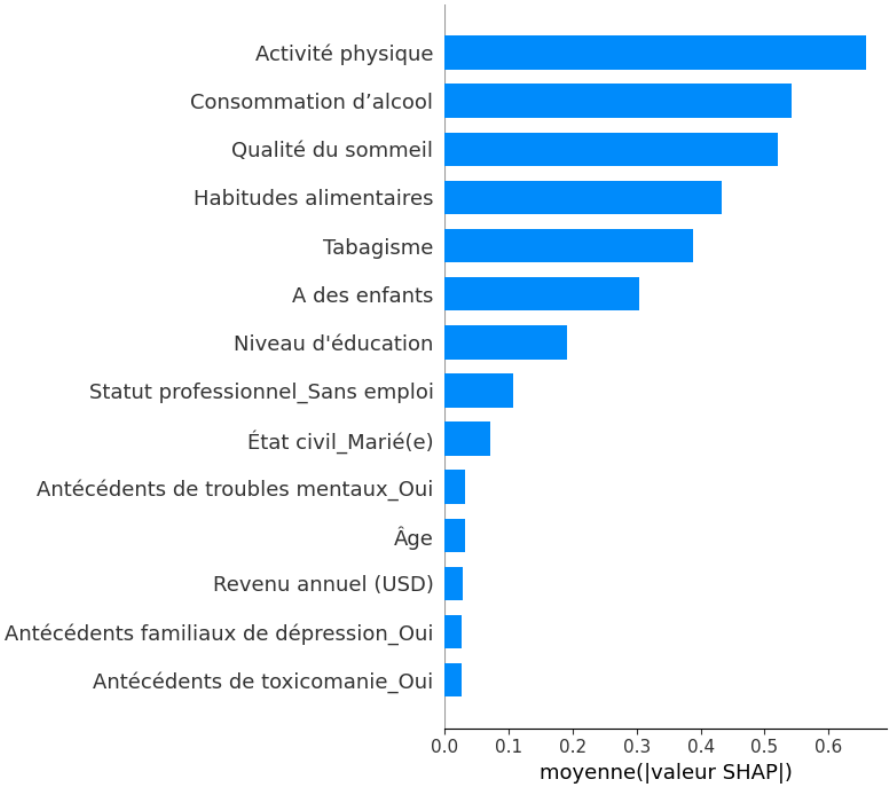


FIGURE 4.8 – Importance globale des variables obtenue avec SHAP sur LGBMClassifier [44].

La Figure 4.8 présente l'importance prédictive des variables obtenue à partir du LGBMClassifier [44] en utilisant les valeurs SHAP (Shapley Additive Explanations) [45]. Cette mesure reflète l'importance corrélationnelle des variables, c'est-à-dire leur contribution moyenne aux prédictions du modèle. Les résultats montrent que *l'activité physique*, *la consommation d'alcool* et *la qualité du sommeil* sont les trois prédicteurs les plus influents du modèle. Ces variables liées au mode de vie dominent le paysage prédictif, suggérant que les comportements individuels constituent les principaux moteurs des prédictions. *Les habitudes alimentaires* et *le tabagisme* apparaissent également avec une importance relativement élevée, confirmant l'influence des comportements quotidiens.

Au-delà des comportements, certaines variables sociodémographiques comme *le fait d'avoir des enfants* et *le niveau d'éducation* présentent une influence modérée. Leur contribution indique que la structure familiale et l'éducation, bien que moins déterminantes que les variables comportementales, participent encore à la formation des résultats prédits.

Dans l'ensemble, l'analyse SHAP met en évidence une hiérarchie claire : les facteurs comportementaux modifiables (activité physique, consommation d'alcool, sommeil, alimentation, tabac) expliquent la plus grande part de la capacité prédictive corrélationnelle du modèle, tandis que les caractéristiques de contexte comme *le statut professionnel*, *l'état civil*, *l'âge*, *le revenu annuel*, ainsi que les *antécédents personnels ou familiaux de troubles mentaux, de dépression ou de toxicomanie* ne jouent qu'un rôle mineur.

4.5 Tentatives d'algorithmes de découverte causale

Avant d'adopter le cadre PEACE, nous avons exploré plusieurs algorithmes de découverte causale afin d'inférer directement des relations structurelles à partir des données. Les méthodes testées incluent l'algorithme PC (Peter-Clark), LiNGAM, et DAG-GNN. Cependant, aucune de ces approches ne s'est révélée adaptée à la nature de notre base de données.

L'**algorithme PC**, qui repose sur des tests d'indépendance conditionnelle sous hypothèse de normalité gaussienne, s'est avéré inadapté. Les tests de normalité ont confirmé que nos variables numériques ne suivent pas une loi normale : le test de Shapiro–Wilk ($W = 0.7097$, $p < 0.001$) et le test de Kolmogorov–Smirnov ($D = 0.3312$, $p < 0.001$) rejettent tous deux l'hypothèse nulle de normalité. De plus, la majorité des variables sont catégorielles ou binaires, pour lesquelles les tests d'indépendance de type Pearson ne sont pas valides. Étant donné que PC repose sur des hypothèses fortes, notamment la fidélité, la suffisance causale et la gaussienneté, il a produit des graphes causaux instables et peu fiables.

L'**algorithme LiNGAM** a également été envisagé. Bien que séduisant sur le plan théorique en raison de son identifiabilité sous des modèles linéaires non-gaussiens, il suppose (i) des relations causales strictement linéaires, (ii) l'absence de confusion latente, et (iii) des termes d'erreur indépendants et non-gaussiens. Or, notre jeu de données présente des relations non linéaires et contient des variables catégorielles encodées via des vecteurs one-hot. Ces hypothèses étant violées, LiNGAM n'a pas permis de générer une structure causale interprétable ou cohérente.

Enfin, nous avons expérimenté **DAG-GNN** (voir figure 4.9), un modèle génératif profond basé sur des réseaux de neurones graphiques au sein d'une architecture de type autoencodeur variationnel. Bien que ce modèle ait produit des DAG complets, les graphes obtenus contenaient des relations causales peu plausibles entre des variables sémantiquement indépendantes (ex. : lien entre *tabagisme* et *âge* sans justification théorique). Cela indique qu'en l'absence de biais inductifs solides ou de contraintes spécifiques au domaine, DAG-GNN peut surajuster ou mal interpréter les relations.

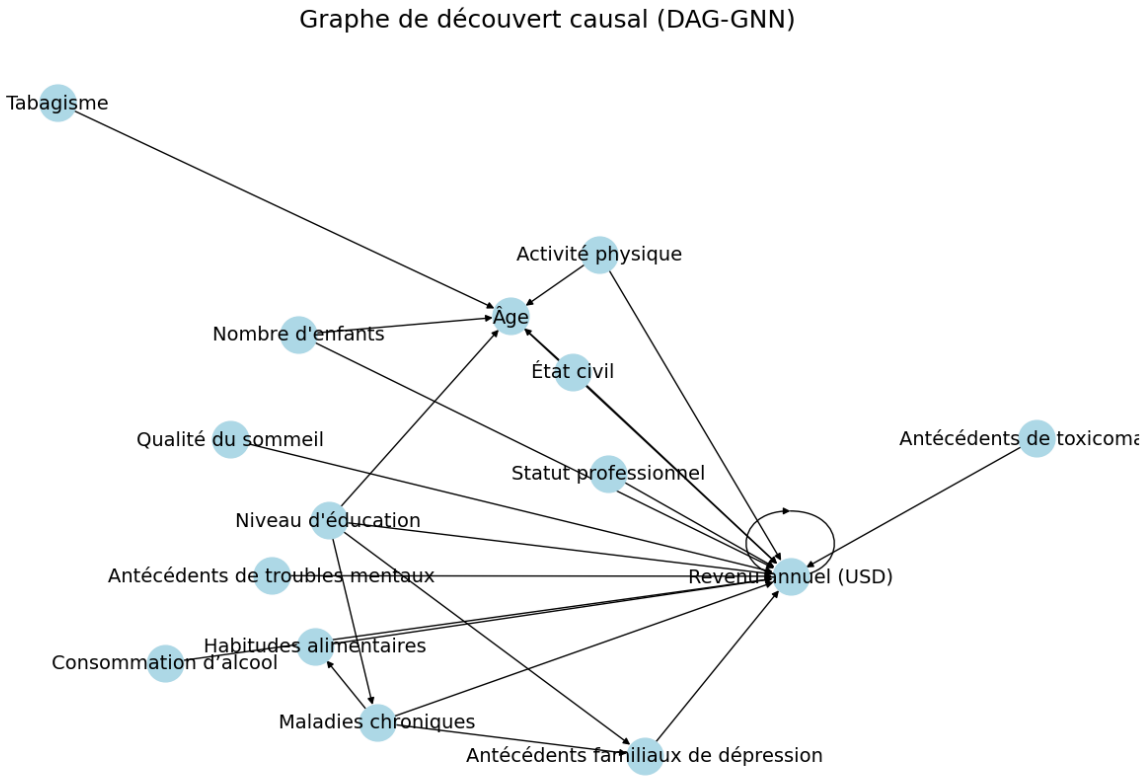


FIGURE 4.9 – Graphe de dépendances causales estimé à l'aide de DAG-GNN. Certaines relations détectées peuvent manquer de plausibilité théorique en raison de l'absence de contraintes structurelles ou de biais inductifs.

Ces explorations ont confirmé la difficulté d'appliquer des méthodes de découverte structurelle sur des données mixtes de haute dimension sans hypothèses fortes. La méthode PEACE propose une alternative robuste en introduisant un paramètre appelé *degré d (*$d \geq 0$*)*, qui permet de contrôler l'importance accordée aux transitions fréquentes ou rares. Lorsque $d \approx 0$, toutes les transitions sont pondérées de façon équitable, y compris les événements rares (micro-effets) ; à mesure que la valeur de d augmente, l'analyse met davantage l'accent sur les transitions les plus fréquentes, ce qui fait ressortir les *macro-effets* dominants. PEACE inclut également des variantes directionnelles Positive PEACE et Negative PEACE qui permettent de distinguer les effets bénéfiques des effets délétères, ce qui facilite l'interprétation. Enfin, *la forme des courbes* ***Mean PEACE*** *fournit des indications précieuses* : si la courbe est **croissante**, cela signifie que les effets causaux plus élevés sont plus probables ; si elle est **décroissante**, ce sont les effets causaux plus faibles qui sont les plus probables. En ce sens, la méthode PEACE se révèle plus interprétable que d'autres méthodes causales [4]. Elle peut ainsi être utilisée même en l'absence de graphe causal, sous l'hypothèse d'ignorabilité (voir l'élément **3.**)

4.6 Métriques d'évaluation

Cette section présente les métriques utilisées pour évaluer à la fois la performance prédictive des modèles et la qualité de l'inférence causale.

Résultats de comparaison des modèles

Tous les modèles ont été évalués à l'aide d'une validation croisée stratifiée à 5 plis, en reportant la moyenne et l'écart-type des métriques standards : exactitude, précision et AUC-ROC[3]. Comme indiqué dans le tableau 4.1, les modèles de type gradient boosting (XGBoost [47], LGBMClassifier [44, 48]) affichent les meilleures performances globales, avec **LGBMClassifier** atteignant l'exactitude la plus élevée (76,34 ± 0,12 %) et le meilleur score AUC-ROC (0,762 ± 0,001). Contrairement aux résultats précédents, TabNet présente une performance compétitive, notamment en termes de précision (91,50 ± 4,12 %).

TABLE 4.1 – Comparaison des performances des modèles (moyenne ± écart-type sur validation croisée 5-plis)

Modèle	**Exactitude (%)**	**Précision (%)**	**AUC-ROC**
Naive Bayes	55,66 ± 0,27	54,00 ± 0,21	0,558 ± 0,003
Régression logistique	59,46 ± 0,12	59,49 ± 0,04	0,594 ± 0,001
K plus proches voisins	64,87 ± 0,32	62,75 ± 0,27	0,649 ± 0,003
Arbre de décision	67,71 ± 0,11	67,09 ± 0,16	0,677 ± 0,001
TabNet	68,51 ± 0,97	91,50 ± 4,12	0,715 ± 0,006
Forêt aléatoire	72,98 ± 0,11	80,57 ± 0,32	0,729 ± 0,001
XGBClassifier	75,99 ± 0,10	96,21 ± 0,34	0,759 ± 0,001
LGBMClassifier	**76,34 ± 0,12**	**99,69 ± 0,15**	**0,762 ± 0,001**

Parmi l'ensemble des modèles évalués, LGBMClassifier a obtenu les meilleures performances globales, avec une exactitude moyenne de 76,34 ± 0,12 %, une précision de 99,69 ± 0,15 % et un score AUC-ROC de 0,762 ± 0,001. Ces résultats en font un candidat naturel pour modéliser la fonction $g(X, Z)$ dans le cadre de la méthode PEACE. Outre ses performances empiriques élevées, LGBMClassifier est également reconnu pour sa rapidité d'entraînement et d'inférence, ce qui est particulièrement avantageux dans des procédures intensives comme le bootstrap ou l'estimation répétée de scores causaux. Sa stabilité, sa scalabilité et sa capacité à gérer des données hétérogènes renforcent encore sa pertinence comme choix de classifieur principal dans ce contexte [44, 48].

4.7 Implémentation de la méthode PEACE

Pour estimer les effets causaux, nous avons mis en œuvre la méthode PEACE en suivant le guide pratique fourni dans le carnet PEACE de Faghihi et Saki [33]. Cette implémentation comprend les étapes suivantes :

— L'entraînement du LGBMClassifier servant de fonction de sortie $g(X, Z)$,
— La génération de courbes d'effets marginaux en perturbant chaque variable de traitement selon des quantiles empiriques, tout en maintenant les autres variables constantes,

3. L'AUC-ROC (Area Under the Receiver Operating Characteristic curve) mesure la capacité d'un classificateur à distinguer entre les classes positives et négatives. Une valeur proche de 1 indique une excellente séparation [46].

— L'estimation des densités des variables de traitement à l'aide de l'estimation par noyau (KDE, *Kernel Density Estimation*) pour les variables continues [49], ou de fonctions de masse de probabilité empiriques (PMF, *Probability Mass Function*) pour les variables catégorielles [50],
— Le calcul des scores PEACE à l'aide d'une approximation discrète de l'intégrale, pondérée par la densité empirique,
— Le calcul des variantes totale, positive et négative du score PEACE, ainsi que de leurs versions normalisées par le nombre de transitions.

L'implémentation a été réalisée en Python, à l'aide de fonctions personnalisées développées pour l'estimation de densité, le calcul des effets marginaux et l'agrégation des scores. L'ensemble du code utilisé dans cette analyse est disponible publiquement dans le carnet Kaggle associé [41].

4.7.1 Incertitude statistique des scores PEACE par bootstrap

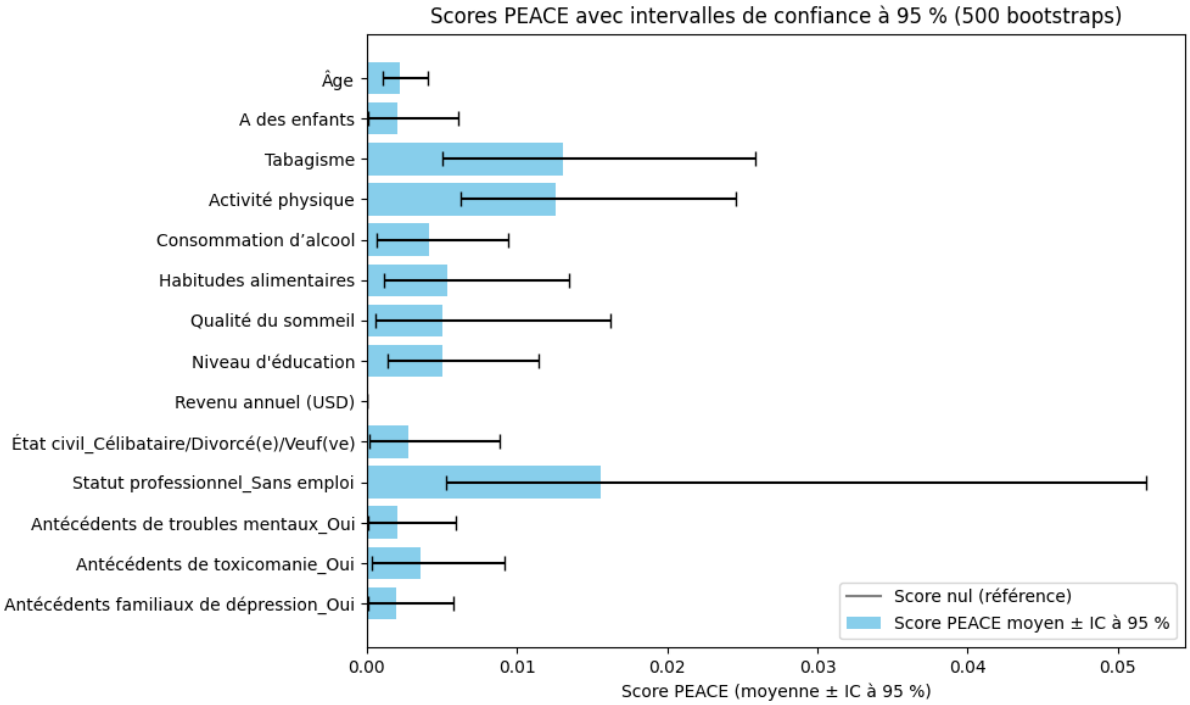


FIGURE 4.10 – Scores PEACE moyens avec intervalles de confiance à 95 % calculés par bootstrap (500 réplications). La ligne grise représente un score nul (référence).

Nous avons appliqué une procédure de bootstrap (500 réplications) afin d'estimer les intervalles de confiance à 95 % pour chaque score PEACE, et ainsi quantifier l'incertitude statistique. Comme illustré dans la Figure 4.10, tous les effets mesurés sont statistiquement significatifs, à l'exception de la variable *Revenu annuel (USD)*, dont le score et l'intervalle sont nuls. Cela pourrait s'expliquer par des interactions complexes entre le revenu, l'éducation et l'emploi, qui influencent peut-être conjointement les résultats de santé [51].

4.8 Analyse des scores PEACE

Une fois le pipeline d'implémentation établi, nous procédons à l'analyse des scores PEACE (Probabilistic Easy Variational Causal Effect) pour chaque variable de traitement. Pour chaque caractéristique X_j, nous calculons les scores PEACE, $\text{PEACE}_{\text{positif}}$, $\text{PEACE}_{\text{négatif}}$, ainsi que leurs versions normalisées par la moyenne, sur un intervalle de paramètres de sensibilité à la densité $d \in [0, 1]$. Ce paramètre d agit comme une lentille sur la distribution des données : des valeurs proches de zéro pondèrent toutes les transitions de manière équitable, y compris les événements rares (*micro-effets*), tandis que l'augmentation de d privilégie les transitions fréquentes ou typiques (*macro-effets*).

Lecture des courbes rôle du paramètre d : Le paramètre d module la pondération des sous-populations selon leur densité. Des valeurs faibles de d favorisent une lecture *micro*, centrée sur les effets dans des groupes rares de la distribution. Inversement, des valeurs élevées conduisent à une lecture *macro*, mettant en lumière les tendances globales dans les zones denses. Cette interprétation s'applique à l'ensemble des figures de cette section.

Les courbes obtenues offrent un aperçu de la direction, de l'intensité et de la stabilité des effets de chaque variable sur le résultat. Elles ne reflètent pas seulement l'effet causal moyen, mais révèlent aussi la structure probabiliste de l'apparition de ces effets. Les sous-sections suivantes interprètent les visualisations produites à l'aide de la méthode PEACE selon cette logique.

4.8.1 Scores PEACE totaux

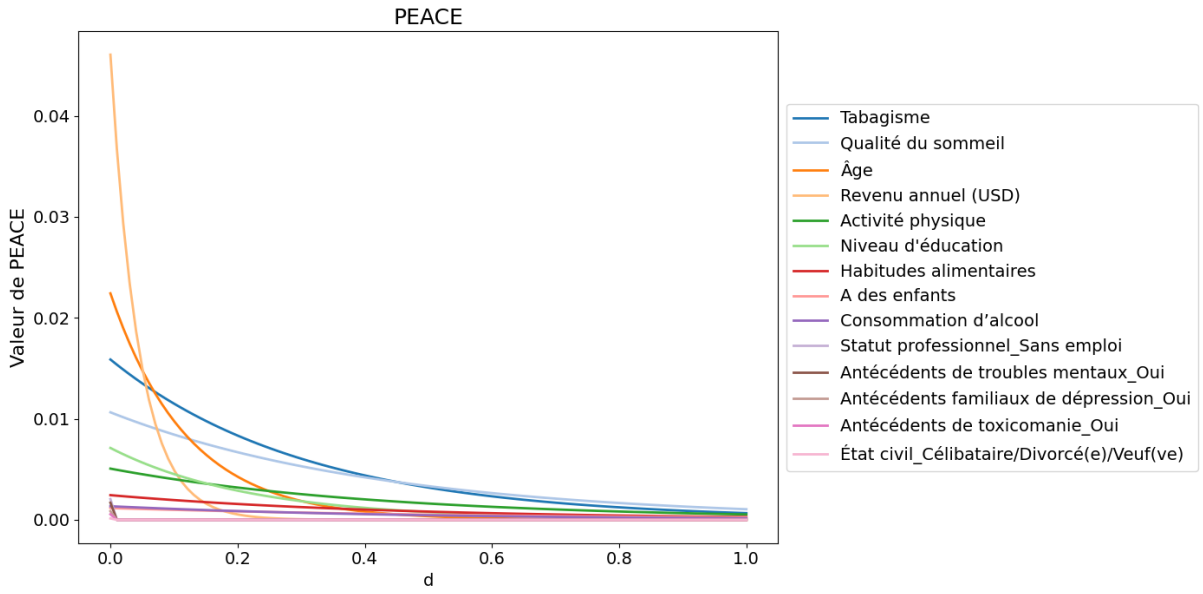


FIGURE 4.11 – Scores PEACE totaux selon le paramètre d, mesurant l'intensité cumulative des effets causaux sur les maladies chroniques (voir l'encart sur le paramètre d en début de section).

Les scores *PEACE totaux* (Figure 4.11) mettent en évidence le **revenu annuel**, l'**âge**, le **statut tabagique** et la **qualité du sommeil** comme variables à forte influence causale globale, particulièrement dans les sous-populations à faible densité.

4.8.2 Scores PEACE moyens

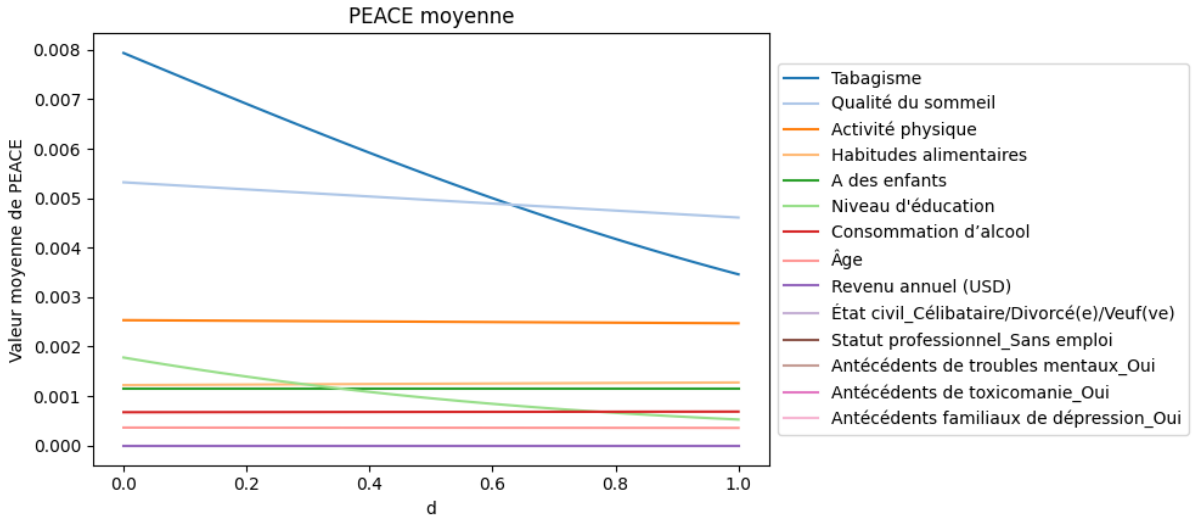


FIGURE 4.12 – Scores PEACE moyens selon le paramètre d, estimant l'effet causal moyen par opportunité d'intervention (voir l'encart sur le paramètre d en début de section).

Les scores *PEACE moyens* (Figure 4.12) représentent l'intensité causale moyenne, obtenue en normalisant les effets totaux par la fréquence des interventions possibles. Ils mettent en évidence le **statut tabagique**, la **qualité du sommeil** et **l'activité physique** comme facteurs d'influence moyenne stable sur les maladies chroniques. On observe que les courbes du **statut tabagique**, de la **qualité du sommeil** et du **niveau d'éducation** sont *décroissantes*, ce qui indique que des effets causaux faibles sont plus probables. Les autres variables présentent des courbes *horizontales*, ce qui suggère une distribution uniforme des effets causaux.

— **Le statut tabagique** présente le score *PEACE moyen* le plus élevé, en particulier dans les régions de faible densité (correspondant à un degré d proche de zéro), ce qui indique une forte influence causale moyenne sur les maladies chroniques au sein des sous-populations rares.
— **Les habitudes de sommeil** se placent en deuxième position sur l'ensemble des valeurs de d, traduisant une influence moyenne soutenue et cohérente sur la distribution.
— **Le niveau d'activité physique** montre des scores *PEACE moyens* stables sur tout l'intervalle, suggérant une influence causale bien répartie dans la population.
— **Les habitudes alimentaires** présentent des valeurs modérées, avec un impact légèrement plus élevé pour les faibles d, ce qui suggère une pertinence dans les transitions peu fréquentes.
— **Le fait d'avoir des enfants**, **le niveau d'éducation**, **la consommation d'alcool** et **l'âge** montrent des effets moyens plus faibles mais non négligeables.
— **Le revenu annuel**, **le statut marital**, **les antécédents de troubles mentaux**, **de toxicomanie** et **familiaux de dépression** affichent des scores faibles sur l'ensemble des d, ce qui indique une influence causale moyenne par transition limitée.

4.8.3 Scores PEACE positifs

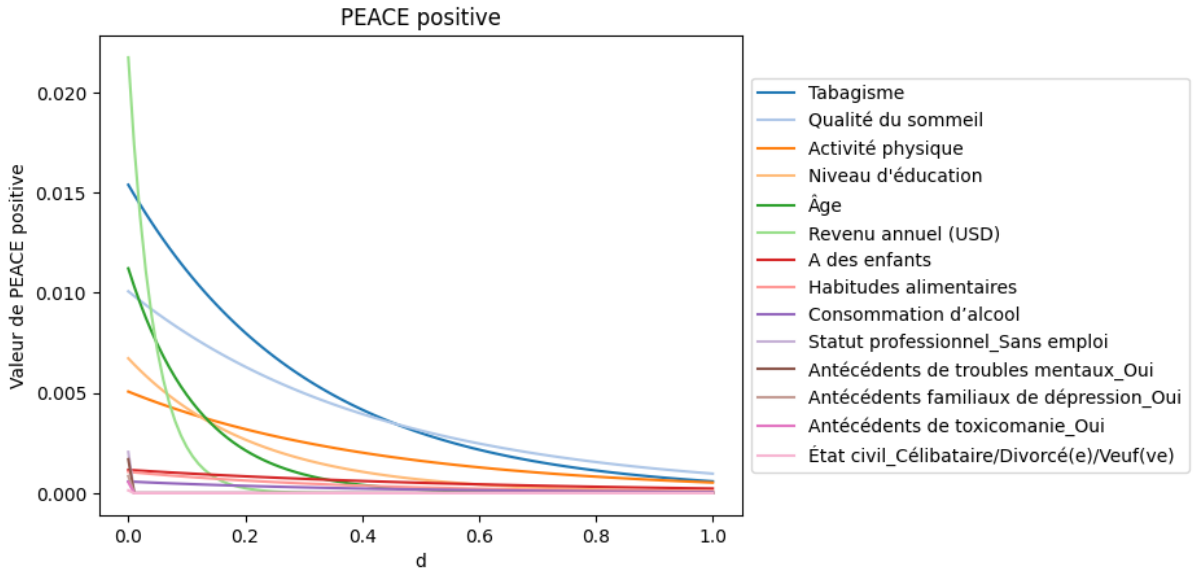


FIGURE 4.13 – Scores PEACE positifs selon le paramètre d, représentant l'intensité des effets causaux délétères sur les maladies chroniques (voir l'encart sur le paramètre d en début de section).

Les scores *PEACE positifs* (Figure 4.13) représentent l'intensité cumulative des effets causaux qui *augmentent* la probabilité de souffrir de maladies chroniques. À faible valeur de d, les **revenus annuels** présentent un score *PEACE positif* particulièrement élevé, indiquant que, dans les sous-populations rares, par exemple des revenus très bas ou, possiblement, des revenus extrêmes, comme le suggère la distribution observée 4.2, sont fortement associés à une augmentation du risque de maladies chroniques. Ce facteur est suivi par **l'âge** et **le statut tabagique**, qui contribuent également à une élévation du risque. Toutefois, le **statut tabagique** et la **qualité du sommeil** maintiennent des effets aggravants élevés sur l'ensemble des valeurs de d, traduisant une influence constante de ces comportements sur le développement des maladies chroniques, même dans les sous-populations majoritaires.

4.8.4 Scores PEACE positifs moyens

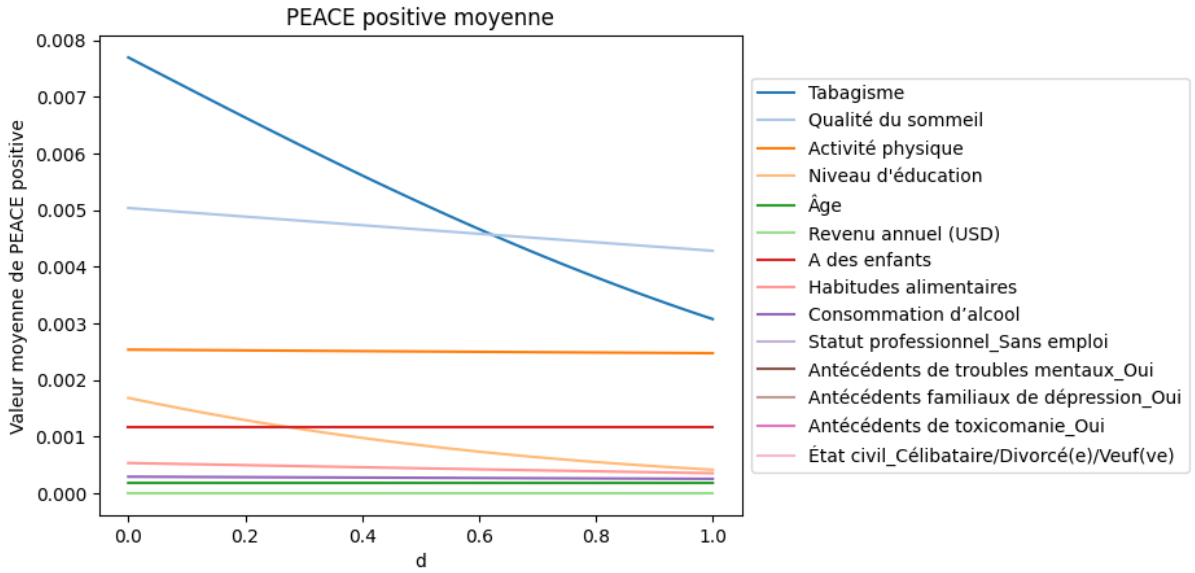


FIGURE 4.14 – Scores PEACE positifs moyens selon le paramètre d, représentant l'effet causal moyen par transition aggravante sur les maladies chroniques (voir l'encart sur le paramètre d en début de section).

Les scores *PEACE positifs moyens* (Figure 4.14) mesurent la régularité avec laquelle chaque variable contribue à l'augmentation du risque de souffrir de maladies chroniques, en normalisant les effets causaux aggravants par la fréquence des interventions. Le **statut tabagique** et la **qualité du sommeil** émergent comme les principaux facteurs aggravants, avec un **pic de l'effet de l'âge** dans les transitions rares. On observe que les courbes du **statut tabagique**, de la **qualité du sommeil** et du **niveau d'éducation** sont *décroissantes*, indiquant que leurs effets causaux aggravants les plus faibles sont les plus probables. **L'activité physique** présente un effet moyen constant et modéré sur l'ensemble des valeurs de d, tandis que les autres variables montrent des courbes *horizontales*, suggérant que leurs effets aggravants, lorsqu'ils existent, sont répartis de manière plus uniforme dans la population.

— **Le statut tabagique** affiche le score *PEACE positif moyen* le plus élevé pour les faibles d, révélant un effet aggravant fort par transition, concentré dans les sous-populations rares.
— **La qualité du sommeil** arrive en deuxième position, avec une courbe stable mais décroissante, témoignant d'un effet aggravant moyen significatif.
— **L'activité physique** et **les habitudes alimentaires** montrent des scores modérés, bien répartis le long de l'intervalle d.
— **Le niveau d'éducation**, **le revenu annuel**, et **le fait d'avoir des enfants** affichent des effets faibles mais non négligeables, plus visibles dans les zones de faible densité.
— **L'âge**, **la consommation d'alcool**, **le statut professionnel** et **le statut marital** ont des scores proches de zéro, indiquant un effet aggravant moyen marginal.
— Enfin, **les antécédents de troubles mentaux**, **de toxicomanie** et **familiaux de dépression** demeurent plats et négligeables, suggérant une absence d'effet causal moyen sur l'augmentation du risque.

4.8.5 Scores PEACE négatifs

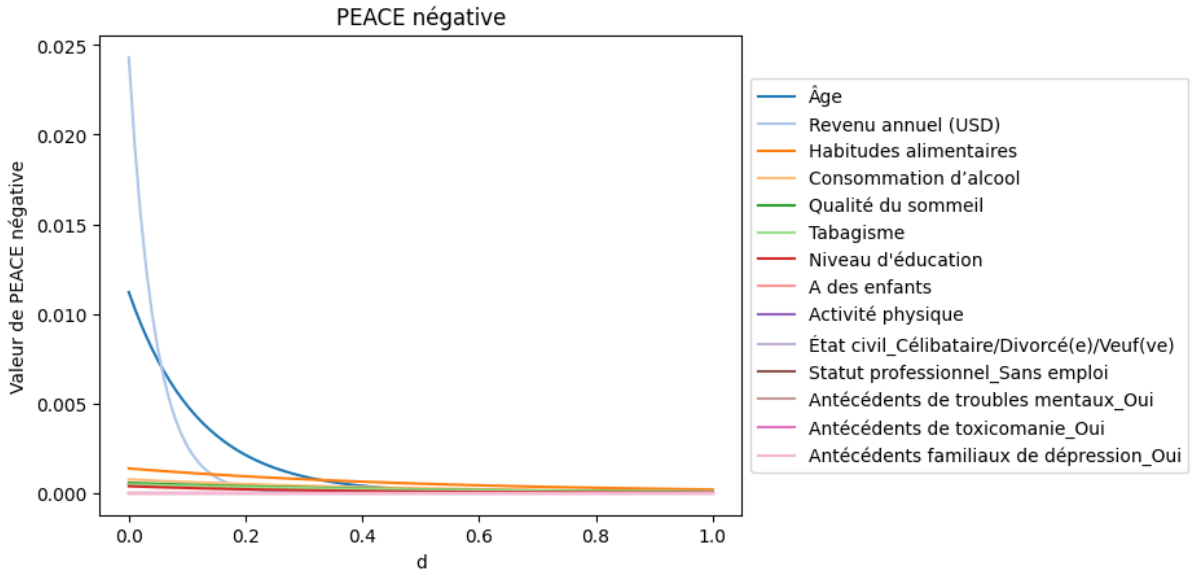


FIGURE 4.15 – Scores PEACE négatifs selon le paramètre d, représentant l'intensité cumulative des effets causaux protecteurs contre les maladies chroniques (voir l'encart sur le paramètre d en début de section).

Les scores *PEACE négatifs* (Figure 4.15) représentent l'intensité cumulative des effets causaux qui réduisent la probabilité de souffrir de maladies chroniques. Le **revenu annuel**, l'**âge** et les **habitudes alimentaires** se démarquent comme les principaux facteurs protecteurs dans les sous-populations rares, avec des effets qui s'estompent à mesure que la densité augmente. La **consommation d'alcool** et, dans une moindre mesure, le **statut tabagique**, présentent des effets protecteurs plus faibles mais persistants. Les autres variables n'exercent qu'une influence négligeable sur la diminution du risque.

4.8.6 Scores PEACE négatifs moyens

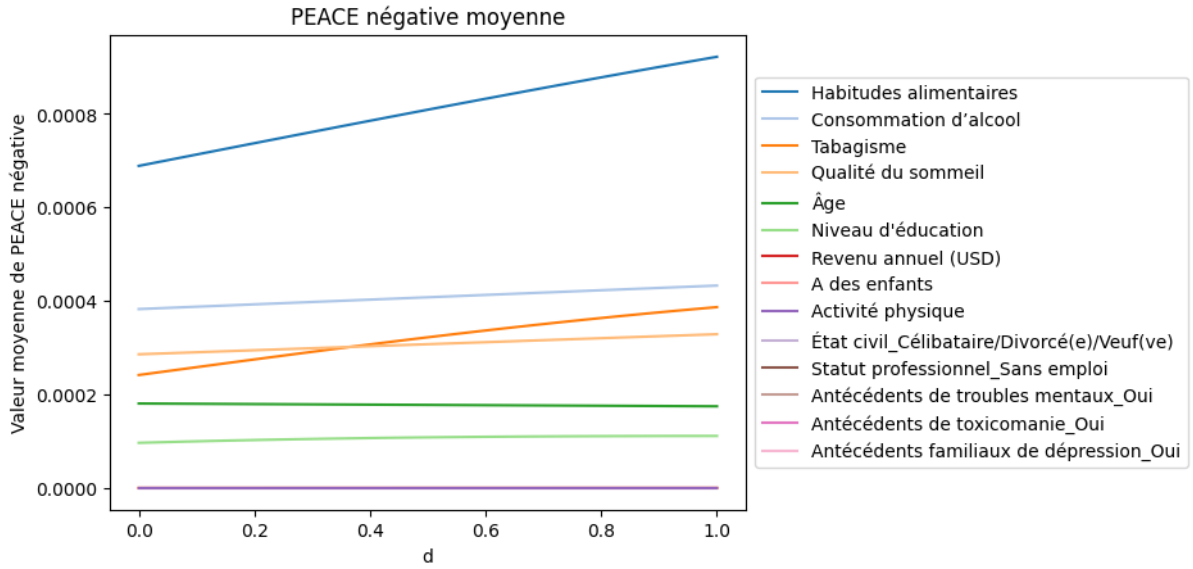


FIGURE 4.16 – Scores PEACE négatifs moyens selon le paramètre d, représentant l'effet causal moyen par transition réduisant le risque de maladies chroniques (voir l'encart sur le paramètre d en début de section).

Les scores *PEACE négatifs moyens* (Figure 4.16) révèlent la force avec laquelle certaines variables contribuent, en moyenne, à la réduction du risque de souffrir de maladies chroniques. Les **habitudes alimentaires**, la **consommation d'alcool**, le **statut tabagique** et **la qualité du sommeil** montrent des effets protecteurs moyens modestes mais cohérents, particulièrement visibles dans les zones de forte densité. On observe que leurs courbes sont *croissantes*, indiquant que les effets causaux protecteurs les plus élevés sont plus probables. Les autres variables présentent des courbes *horizontales*, ce qui suggère que leurs effets protecteurs sont répartis de manière uniforme.

— **Les habitudes alimentaires** présentent les scores les plus élevés sur l'ensemble de d, avec une tendance légèrement croissante, traduisant un effet protecteur moyen fort et étendu.
— **La consommation d'alcool** et **le statut tabagique** suivent avec des scores modérés croissants, suggérant un effet protecteur qui devient plus présent dans les régions de forte densité.
— **La qualité du sommeil** montre une courbe relativement plate, mais se distingue par un effet protecteur moyen cohérent sur tout l'intervalle.
— **L'âge**, **le revenu annuel** et **le niveau d'éducation** présentent des scores faibles et stables, indiquant une influence protectrice moyenne marginale.
— Toutes les autres variables, notamment **le fait d'avoir des enfants**, **le statut marital**, **le statut professionnel**, **les antécédents de troubles mentaux**, **de toxicomanie** et **familiaux de dépression**, restent proches de zéro pour tous les d, ce qui indique une contribution protectrice moyenne négligeable.

4.8.7 Synthèse des analyses PEACE

Les analyses PEACE révèlent comment les variables de traitement influencent causalement les maladies chroniques, tant par des effets aggravants que protecteurs. Ces observations sont résumées dans le Tableau 4.2, qui présente exclusivement les variantes de Mean PEACE en distinguant les contributions positives et négatives.

Les scores PEACE totaux mettent en évidence le **revenu annuel**, l'**âge**, le **statut tabagique** et la **qualité du sommeil** comme variables à forte influence causale globale, particulièrement dans les sous-populations à faible densité. En revanche, **les scores PEACE moyens** soulignent le **statut tabagique**, la **qualité du sommeil** et **l'activité physique** comme facteurs d'influence moyenne stable, reflétant des effets causaux plus uniformément répartis.

Les composantes positives des scores PEACE, qu'elles soient totales ou moyennes, isolent les effets aggravants sur le risque. Le **statut tabagique** et la **qualité du sommeil** apparaissent systématiquement comme les contributeurs dominants, avec un pic de l'effet de l'**âge** dans les transitions rares, tandis que **l'activité physique** montre un effet moyen large et constant. Alors que le score PEACE positif reflète l'ampleur totale des transitions nuisibles, le score moyen met en lumière la régularité avec laquelle une variable augmente le risque à chaque intervention.

Les composantes négatives des scores PEACE capturent les effets causaux diminuant légèrement le risque. Toutes les variables présentent des valeurs faibles, proches de zéro, ce qui indique que les effets protecteurs sont globalement modestes. Parmi elles, les **habitudes alimentaires** se distinguent par une valeur un peu plus élevée, suivies par la **consommation d'alcool**, le **statut tabagique**, la **qualité du sommeil** et l'**âge**, mais leurs contributions restent limitées. La majorité des autres variables sont quasi nulles, confirmant l'absence d'effet protecteur marqué.

Notons que certaines variables, comme l'**âge**, présentent à la fois des effets aggravants (positifs) dans certaines transitions rares et des effets protecteurs (négatifs) dans d'autres, révélant un comportement causal ambivalent selon les sous-populations considérées.

Une asymétrie claire émerge : les effets causaux augmentant le risque sont plus forts et plus largement répartis que les effets protecteurs. Ce déséquilibre semble en partie lié à une distribution des données biaisée. Par exemple, les profils protecteurs tels que les *habitudes alimentaires saines* (17,5 %), l'*activité physique active* (19 %) ou le *bon sommeil* (21 %) sont sous-représentés, tandis que les catégories à risque comme le *mauvais sommeil* (31 %) et les *habitudes alimentaires malsaines* (41 %) prédominent. En conséquence, PEACE est plus sensible aux transitions nocives car elles sont plus fréquentes ou concernent des régions de données plus denses. Ce biais structurel peut masquer le potentiel causal réel de facteurs bénéfiques plus rares.

Métrique	Variables dominantes	Interprétation
$\text{PEACE}_d^{\text{mean}}$ (Totale)	Statut tabagique, Qualité du sommeil, Activité physique	Effets causaux moyens stables, uniformément répartis dans les transitions.
$\text{PEACE}_d^{\text{mean},+}$ (Positive)	Statut tabagique, Qualité du sommeil, Activité physique (effet constant)	Facteurs aggravants du risque ; mettent en évidence la régularité avec laquelle une variable augmente le risque à chaque intervention.
$\text{PEACE}_d^{\text{mean},-}$ (Négative)	Habitudes alimentaires (plus élevé), puis Consommation d'alcool, Tabagisme, Qualité du sommeil et Âge	Tous les effets protecteurs restent faibles, proches de zéro. Toutefois, les habitudes alimentaires se distinguent légèrement au-dessus des autres variables, tandis que les contributions de l'alcool, du tabagisme, du sommeil et de l'âge demeurent très modestes.

TABLE 4.2 – Synthèse des variantes de Mean PEACE : variables dominantes et interprétation

4.8.8 Importance corrélationnelle vs importance causale

Alors que SHAP 4.4 mesure une *importance corrélationnelle* des variables dans le modèle LGBMClassifier, la méthode **PEACE** révèle leur *importance causale*, en distinguant les effets positifs et négatifs et en tenant compte de la distribution des traitements. Ainsi, SHAP reflète la contribution prédictive moyenne dans un cadre corrélationnel, tandis que PEACE vise à capturer les véritables relations causales sous-jacentes.

Concrètement, SHAP met en avant l'*activité physique*, la *consommation d'alcool* et la *qualité du sommeil* comme principaux prédicteurs. En revanche, les **scores PEACE moyens** identifient surtout le **statut tabagique**, la **qualité du sommeil** et l'**activité physique** comme facteurs causaux stables et réguliers par transition. Des variables comme le **revenu annuel** ou l'**âge**, quasiment négligeables dans SHAP, présentent une influence marquée dans les **scores PEACE totaux**, mais concentrée dans des sous-populations rares.

4.8.9 Lien avec les politiques de santé publique.

Les résultats obtenus via la méthode PEACE mettent en évidence le tabagisme, les troubles du sommeil et le manque d'activité physique comme facteurs de risque comportementaux majeurs, avec des scores PEACE positifs élevés. Ces conclusions renforcent la pertinence d'interventions ciblées, déjà identifiées comme prioritaires dans les politiques actuelles de santé publique. La *Stratégie canadienne sur le tabac* vise à réduire la prévalence du tabagisme à moins de 5 % d'ici 2035, en soutenant l'abandon du tabac dans les populations vulnérables [52]. En parallèle, le gouvernement fédéral soutient activement la santé du sommeil à travers un programme de recherche pancanadien lancé en 2022 par les Instituts de recherche en santé du Canada (IRSC), qui finance le développement d'interventions novatrices pour prévenir, détecter et gérer l'insomnie [53]. Au Québec, le *Programme national de santé publique 2023–2027* fait explicitement référence à la réduction du tabagisme, à la promotion du sommeil et à l'augmentation de l'activité physique comme priorités d'action [54]. Ainsi, l'utilisation des scores PEACE permet d'objectiver l'allocation des ressources vers ces déterminants modifiables, tout en s'alignant sur les

priorités stratégiques actuelles des systèmes de santé. Les scores PEACE offrent un appui empirique aux choix d'intervention en santé publique, en priorisant les facteurs de risque les plus influents pour les maladies chroniques.

4.9 Avantages de la méthode PEACE comparée aux approches classiques et modernes

Comme discuté dans le Chapitre 2, les méthodes classiques d'inférence causale, telles que le cadre des résultats potentiels [14, 1], les modèles causaux structurels (SCM) [2] ou les variables instrumentales [55, 18], offrent chacune des outils précieux pour identifier des relations causales. Cependant, elles peinent à s'adapter à des données complexes et de haute dimension, et restent restreintes dans leur capacité à modéliser l'hétérogénéité des effets du traitement. Par ailleurs, elles se focalisent généralement sur des effets moyens ou des sous-groupes particuliers, comme l'effet causal local moyen (LATE) dans le cas des variables instrumentales [18], ce qui compromet leur généralisabilité à l'ensemble de la population et réduit leur pertinence dans les contextes où l'on vise des estimations causales individualisées à partir de bases de données observationnelles riches et multidimensionnelles.

Au-delà des approches classiques, plusieurs méthodes modernes basées sur l'apprentissage automatique ont été proposées pour surmonter ces limites. Mais, elles posent de nouveaux défis, notamment en matière d'interprétabilité. L'une des principales limites des méthodes telles que DragonNet [7], TARNet [6] et les forêts causales [20] réside dans l'interprétabilité de leurs sorties. DragonNet et TARNet produisent des prédictions contrefactuelles $\hat{Y}(0)$ et $\hat{Y}(1)$, mais fournissent peu d'éléments sur les variables explicatives qui influencent ces prédictions, ni sur la manière dont les variations du traitement modifient localement le résultat. Leur architecture neuronale profonde agit comme une boîte noire, rendant difficile toute interprétation causale fine. Les forêts causales, quant à elles, estiment des effets de traitement conditionnels en des points spécifiques de l'espace des covariables, mais ne permettent pas de distinguer si l'effet provient d'une augmentation ou d'une diminution du traitement. De plus, l'agrégation des décisions au sein de multiples arbres rend opaques les chemins causaux individuels.

Face à ces difficultés d'interprétation et de généralisation, la méthode **PEACE** fournit des sorties interprétables et constitue pour nous une alternative robuste, notamment grâce à l'introduction d'un paramètre appelé *degré d*, qui permet de moduler l'importance accordée aux transitions fréquentes ou rares dans l'espace des traitements. La méthode offre également des variantes directionnelles, *Positive PEACE* et *Negative PEACE*, qui nous permettent de distinguer clairement les effets bénéfiques des effets délétères, améliorant ainsi la lisibilité des résultats. Enfin, la *forme de la courbe* ***Mean PEACE*** nous fournit des indications précieuses sur la distribution des effets causaux : une courbe croissante suggère que les effets élevés sont plus probables, tandis qu'une courbe décroissante indique une prédominance des effets faibles. Ces propriétés renforcent, selon nous, le caractère interprétable et exploratoire de la méthode PEACE par rapport à d'autres approches causales.

Empiriquement, PEACE capture à la fois les structures causales globales et locales, tout en mettant en évidence des asymétries d'influence. Par exemple, des variables comme le

statut tabagique et la **qualité du sommeil** émergent comme des facteurs dominants d'augmentation du risque, tandis que les effets protecteurs, bien que présents, restent plus localisés et confinés à des sous-populations peu denses. Ces subtilités seraient probablement ignorées par les estimateurs classiques de type ATE ou par les mesures d'importance des variables issues de modèles purement prédictifs, qui ne tiennent pas compte de la direction ni de la forme locale des effets causaux.

CHAPITRE 5
CONCLUSION ET PERSPECTIVES

Ce mémoire s'est structuré en plusieurs étapes complémentaires. Nous avons d'abord mené une analyse statistique exploratoire, incluant l'évaluation de l'importance des variables à l'aide de SHAP, puis expérimenté certaines méthodes de découverte causale afin d'estimer des structures potentielles entre les variables. Nous avons ensuite présenté les principaux modèles causaux existants, avant d'introduire et de formaliser la méthode **PEACE** (*Probabilistic Easy Variational Causal Effect* [4]), que nous avons finalement mise en œuvre dans une étude empirique.

Cette étude a appliqué le cadre PEACE à une base de données de santé synthétique complexe afin d'étudier comment les variables de traitement influencent causalement la présence de maladies chroniques. Les scores PEACE ont permis d'identifier non seulement les variables exerçant des effets causaux, mais aussi la manière dont ces effets varient en magnitude et en direction selon les sous-populations.

L'analyse a montré que les effets causaux augmentant le risque sont substantiellement plus forts et plus étendus que ceux réduisant le risque. Des variables comme le **statut tabagique**, la **qualité du sommeil** et l'**âge** se révèlent être des moteurs dominants du risque accru, tandis que les influences protectrices notamment liées aux **habitudes alimentaires** ou à la **consommation d'alcool** sont plus modestes et localisées. Cette asymétrie semble en partie induite par la sous-représentation des profils « sains » dans la distribution des données.

Comparée à des approches classiques et modernes d'inférence causale, PEACE offre un cadre flexible, interprétable et particulièrement adapté à l'exploration de l'hétérogénéité causale dans des contextes à forte complexité structurelle ou à données biaisées. Sa capacité à décomposer les effets en composantes positives et négatives, et à les évaluer selon la densité des données, permet une lecture plus riche et nuancée de l'hétérogénéité causale.

Parmi les perspectives de recherche, on peut citer l'application du cadre PEACE à des bases réelles de santé clinique ou socio-économique, l'extension de la méthode à des données temporelles ou longitudinales, ou encore son intégration à des architectures d'apprentissage profond telles que TARNet ou DragonNet pour améliorer ses performances dans des contextes à haute dimensionnalité et à confondement partiel.

Bibliographie

[1] G. W. Imbens et D. B. Rubin, *Causal Inference for Statistics, Social, and Biomedical Sciences*. 2015.

[2] J. Pearl, *Causality : Models, Reasoning and Inference*. Cambridge University Press, 2009.

[3] D. Janzing, L. Minorics et B. Schölkopf, « Quantifying causal influences », 2013.

[4] U. Faghihi et A. Saki, « Probabilistic easy variational causal effect (peace) », 2024.

[5] A. Therrien, « Depression dataset », 2024.

[6] U. Shalit, F. D. Johansson et D. Sontag, « Estimating individual treatment effect : generalization bounds and algorithms », 2017.

[7] C. Shi, D. M. Blei et V. Veitch, « Adapting neural networks for the estimation of treatment effects », 2019.

[8] Aristotle, *Physics*. 350 BC. Translated and edited by Robin Waterfield, Oxford University Press, 2021.

[9] A. Falcon, « Aristotle on causality », *The Stanford Encyclopedia of Philosophy*, 2019.

[10] D. Hume, *An Enquiry Concerning Human Understanding*. 1748. Available at `https://www.gutenberg.org/ebooks/9662`, Project Gutenberg.

[11] P. Spirtes, C. Glymour et R. Scheines, *Causation, Prediction, and Search*. MIT Press, 2nd éd., 2000.

[12] M. A. Hernán et J. M. Robins, *Causal Inference*. 2010. Unpublished book available at `https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/`.

[13] B. Neal, « Introduction to causal inference ». `https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf`, 2020. Accessed April 2025.

[14] D. B. Rubin, « Estimating causal effects of treatments in randomized and nonrandomized studies », 1974.

[15] U. Faghihi, S. Robert, P. Poirier et Y. Barkaoui, « From association to reasoning, an alternative to pearl's causal reasoning », 2020.

[16] J. Pearl et D. Mackenzie, *The Book of Why : The New Science of Cause and Effect*. 2018.

[17] S. Athey, « The impact of machine learning on economics », 2019.

[18] J. D. Angrist et G. W. Imbens, « Two-stage least squares estimation of average causal effects in models with variable treatment intensity », 1995.

[19] S. Athey et G. Imbens, « Recursive partitioning for heterogeneous causal effects », 2016.

[20] S. Wager et S. Athey, « Estimation and inference of heterogeneous treatment effects using random forests », 2018.

[21] A. M. Alaa et M. van der Schaar, « Causal inference benchmarking via algorithmic independence of cause and mechanism », 2022.

[22] X. Zheng, B. Aragam, P. Ravikumar et E. P. Xing, « Dags with no tears : Continuous optimization for structure learning », 2018.

[23] S. Shimizu, P. O. Hoyer, A. Hyvärinen et A. Kerminen, « A linear non-gaussian acyclic model for causal discovery », 2006.

[24] Y. Yu, Z. Zhang, Y. Rong, T. Xu et J. Huang, « Dag-gnn : Dag structure learning with graph neural networks », 2019.

[25] L. I. Rudin, S. Osher et E. Fatemi, « Nonlinear total variation based noise removal algorithms », 1992.

[26] L. C. Evans et R. F. Gariepy, *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics, CRC Press, 1992.

[27] G. B. Folland, *Real Analysis : Modern Techniques and Their Applications*. 1999.

[28] R. Abergel et L. Moisan, « The shannon total variation », 2017.

[29] A. Chambolle, « Total variation minimization and a class of binary mrf models », 2005.

[30] M. A. Hernán et J. M. Robins, *Causal Inference : What If*. 2020.

[31] P. R. Rosenbaum et D. B. Rubin, « The central role of the propensity score in observational studies for causal effects », 1983.

[32] T. Abe, Y. Shibata, T. Ito, K. Suzuki, A. Kudo, H. Hara, M. Nakao, G. Sugihara et H. Takahashi, « Association between exercise-based physical activity and stress responses according to work-related moderate-to-vigorous physical activity : A cross-sectional study », 2024.

[33] U. Faghihi, « Probabilistic easy variational causal effect – explanation notebook ». `https://github.com/joseffaghihi/Probabilistic-Easy-Variational-Causal-Effect-As-a-New-Causal-Inference`, 2024.

[34] O. Hines, K. Diaz-Ordaz et S. Vansteelandt, « Optimally weighted average derivative effects », 2024.

[35] Centers for Disease Control and Prevention, « Health effects of cigarette smoking ». `https://www.cdc.gov/tobacco/about/`, 2020. Accessed : 2025-05-20.

[36] D. E. Warburton, C. W. Nicol et S. S. Bredin, « Health benefits of physical activity : the evidence », *Canadian Medical Association Journal*, 2006.

[37] World Health Organization, « Healthy diet ». https://www.who.int/news-room/fact-sheets/detail/healthy-diet, 2020.

[38] J. M. Krueger *et al.*, « Sleep quality and health : A review of the relationship between sleep and health outcomes », *Journal of Clinical Sleep Medicine*, vol. 5, no. 2, p. 123–129, 2009.

[39] M. Kivimäki, G. D. Batty, M. Hamer, J. E. Ferrie, J. Vahtera et M. Virtanen, « Association between common mental disorders and subsequent chronic physical conditions : a meta-analysis », 2017.

[40] G. E. Batista, R. C. Prati et M. C. Monard, « Tomek links and smote for data cleaning and class imbalance », 2004.

[41] K. Faustin, « Depression causal inference with peace ». https://www.kaggle.com/code/kagabofaustin/depression-causal-inference-with-peace, 2025. Kaggle notebook accompanying the master's thesis.

[42] M. Marmot, « Employment conditions and health inequalities », *The Lancet*, vol. 372, no. 9650, p. 1153–1163, 2008.

[43] U. S. General, « Smoking cessation : A report of the surgeon general ». U.S. Department of Health and Human Services, 2020.

[44] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye et T.-Y. Liu, « Lightgbm : A highly efficient gradient boosting decision tree », 2017.

[45] S. M. Lundberg et S.-I. Lee, « A unified approach to interpreting model predictions », 2017.

[46] T. Fawcett, « An introduction to roc analysis », 2006.

[47] T. Chen et C. Guestrin, « Xgboost : A scalable tree boosting system », 2016.

[48] Microsoft, « Lightgbm documentation ». Disponible à https://lightgbm.readthedocs.io/en/stable/.

[49] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. 1986.

[50] L. Wasserman, *All of Statistics : A Concise Course in Statistical Inference*. 2004.

[51] P. Braveman, « Health disparities and health equity : the issue is justice », 2005.

[52] S. Canada, « Stratégie canadienne sur le tabac », 2025. Disponible à https://www.canada.ca/fr/sante-canada/services/publications/vie-saine/strategie-tabac-canada.html.

[53] C. I. of Health Research, « Government of canada invests in research to improve sleep for canadians », 2022. Disponible à https://www.canada.ca/en/institutes-health-research/news/2022/06/government-of-canada-invests-in-research-to-improve-sleep-for-canadians.html.

[54] M. de la Santé et des Services sociaux du Québec, « Programme national de santé publique 2023–2027 », 2023. Disponible à https://publications.msss.gouv.qc.ca/msss/fichiers/2022/22-297-05W.pdf.

[55] J. D. Angrist et A. B. Krueger, « Identification of causal effects using instrumental variables », 1996.