Inférence causale sur données observationnelles : développement et application pour les soins critiques

Maxime Leger

▶ **To cite this version:**

Maxime Leger. Inférence causale sur données observationnelles : développement et application pour les soins critiques. Médecine humaine et pathologie. Nantes Université, 2022. Français. NNT : 2022NANU1006 . tel-03917164

HAL Id: tel-03917164
https://theses.hal.science/tel-03917164

Submitted on 1 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DOCTORAT BIOLOGIE BRETAGNE SANTE LOIRE

Nantes Université

THESE DE DOCTORAT DE

NANTES UNIVERSITE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : Biostatistique

Par
Maxime LÉGER

Inférence causale sur données observationnelles : développement et application pour les soins critiques

Thèse présentée et soutenue à Angers, le 13 mai 2022
Unité de recherche : UMR INSERM 1246 – SPHERE « methodS in Patient-centered outcomes & HEalth ResEarch"

Rapporteurs avant soutenance :

David HAJAGE — Maître de conférence des universités – Praticien hospitalier, Sorbonne Université, France
Sophie HAMADA — Professeur des universités – Praticien hospitalier, Université de Paris, France

Composition du Jury :

Président :
Sophie HAMADA — Professeur des universités – Praticien hospitalier, Université de Paris, France

Examinateurs :
Emilie LANOY — Ingénieur de recherche – Chargé de recherche, INSERM, France
Antoine ROQUILLY — Professeur des universités – Praticien hospitalier, Université de Nantes, France

Dir. de thèse :
Yohann FOUCHER — Maître de conférence des universités, Université de Nantes, France

Co-dir. de thèse :
Sigismond LASOCKI — Professeur des universités – Praticien hospitalier, Université d'Angers, France

Remerciements

En premier lieu, je tiens à remercier Sophie Hamada, Emilie Lanoy, David Hajage et Antoine Roquilly, qui me font l'honneur de s'intéresser à mon travail et de constituer ce jury. Vos précieuses remarques me permettent d'améliorer la qualité scientifique de cette thèse.

Je tiens naturellement à poursuivre ces remerciements avec une pensée particulière pour mes directeurs de thèse. Sigismond, vous êtes depuis mon arrivée à Angers un mentor pour ma pratique de l'anesthésie-réanimation. Votre intelligence, votre dynamisme et votre sens du travail sont exemplaires. Merci de me faire confiance, cela compte beaucoup pour moi. Avoir un mentor inspirant est rare, mais en avoir deux est inespéré. Yohann, depuis que tu as accepté de me guider, je n'ai fait que grandir. Ta rigueur scientifique et ton ouverture d'esprit sont des sources d'inspiration. Merci aussi d'avoir joué au lièvre pendant nos sorties le long de la Sèvre Nantaise !

Je souhaiterais ensuite remercier Véronique Sébille et Bruno Giraudeau pour m'avoir accueilli au sein de l'unité SPHERE. De manière plus large, j'aimerais aussi remercier toutes les personnes passionnées de biostatistiques que j'ai pu rencontrer et découvrir dans l'unité, et tout particulièrement Florent et Arthur, vous avez souvent été mes phares dans la brume.

Ma pratique de la recherche est indissociable de ma pratique clinique. En ce sens, je tiens à remercier mes collègues du terrain au bloc opératoire des urgences qui s'adaptent au mieux pour me permettre de faire mon bout de chemin. Merci à vous, Claire, Guillaume, Mathieu ; et sans oublier Florian, Pierre et Vincent.

Merci à tous les membres de l'équipe du DAR Angevin pour cette ambiance de travail agréable au quotidien. C'est aussi grâce à vous qu'il fait bon vivre à Angers.

Je tiens à remercier mes plus proches amis, ceux qui me suivent depuis un bon bout de temps, comme ceux arrivés plus récemment, ou même ceux avec qui je partage le goût de la pédale. Je sais que vous ne comprenez pas toujours ce que je fais, néanmoins votre soutien reste sans faille.

A mes parents, mes grand-parents, ma soeur, Valérie, la famille Verhaeghe, merci d'être toujours là et de continuer à m'encourager comme vous le faites.

A Caroline, ma femme, Faustine et Gustave, mes enfants, votre amour est mon moteur. Je tiens à vous remercier pour votre compréhension dans les moments qui ont été les plus exigeants. Avancer sans vous n'aurait aucun sens.

Valorisations scientifiques

Publications issues de la thèse

1. **Léger M**, Frasca D, Roquilly A, Seguin P, Cinotti R, Dahyot-Fizelier C, Asehnoune K, Le Borgne F, Gaillard T, Foucher Y, Lasocki S, for Atlanréa Group. Early use of barbiturates is associated with increased mortality in traumatic brain injury patients from a propensity score-based analysis of a prospective cohort (2022). *PLOS ONE* 17(5) : e0268013. DOI : 10.1371/journal.pone.0268013
2. **Léger M**, Chatton A, Le Borgne F, Pirracchio R, Sigismond L, Foucher Y. Causal inference in case of near-violation of positivity : comparison of methods (2021). *Biometrical Journal* DOI : 10.1002/bimj.202000323. Epub ahead of print.

Autres publications en lien avec la thèse

1. Le Borgne F, Chatton A, **Léger M**, Lenain R et Foucher Y. G-computation and machine learning for estimating the causal effects of binary exposure statuses on binary outcomes (2021). *Scientific Reports* 11, 1435. DOI : 10.1038/s41598-021-81110-0
2. Chatton A, Le Borgne F, Leyrat C, Gillaizeau F, Rousseau C, Barbin L, Laplaud D-A, **Léger M**, Giraudeau B et Foucher Y. G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets : a comparative simulation study (2020). *Scientific Reports* 10, 9219. DOI : 10.1038/s41598-020-65917-x
3. Danelian G, Foucher Y, **Léger M**, Le Borgne F et Chatton A. Identifying positivity violations using decision trees : introducing the P-CART algorithm. [En préparation]

Table des matières

Table des matières vii

Abréviations ix

1 État de l'art en inférence causale 3
- 1.1 Causalité . . . 4
- 1.2 Intérêt des essais randomisés . . . 4
- 1.3 Facteurs de confusion et biais d'indication . . . 6
- 1.4 Évènements contrefactuels . . . 7
- 1.5 Mesures de l'effet causal . . . 7
- 1.6 Population d'intérêt . . . 8
- 1.7 Trois hypothèses sous-tendant l'identifiabilité . . . 8
- 1.8 Particularités de l'infraction de l'hypothèse de positivité . . . 9
  - 1.8.1 Infraction théorique et infraction aléatoire . . . 9
  - 1.8.2 Problématique de l'extrapolation . . . 9
- 1.9 Modélisations statistiques et inférence causale . . . 10
  - 1.9.1 Scores de propension . . . 10
  - 1.9.2 G-computation . . . 14
  - 1.9.3 Estimateurs doublement robustes . . . 15
  - 1.9.4 Méthodes avec troncature . . . 16
  - 1.9.5 Évaluation du déséquilibre entre les groupes . . . 16

2 Application aux soins critiques 19
- 2.1 Inférence causale dans les soins critiques . . . 20
  - 2.1.1 Les limites des études contrôlées randomisées . . . 20
  - 2.1.2 Données observationnelles et soins critiques . . . 20
  - 2.1.3 Exemple du projet MIMIC . . . 21
  - 2.1.4 Autres bases de données en soins critiques . . . 22
- 2.2 Présentation de la cohorte AtlanRéa . . . 23
  - 2.2.1 Présentation de la cohorte . . . 23
  - 2.2.2 Données collectées . . . 24
  - 2.2.3 Description de la cohorte (extraction de janvier 2021) . . . 26

3 Evaluation de l'impact de l'utilisation précoce des barbituriques chez les patients traumatisés graves **29**

3.1 Manuscrit 29

3.2 Éléments supplémentaires 44

4 Comparaison de méthodes d'estimation causale dans des situations d'infraction aléatoire de la positivité **59**

4.1 Manuscrit 59

4.2 Éléments supplémentaires 75

5 Discussion et perspectives **107**

5.1 Résumé des travaux réalisés 108

5.1.1 Analyse de l'impact des barbituriques 108

5.1.2 Problématique de l'hypothèse de positivité 108

5.1.3 Intérêt de la G-Computation 109

5.1.4 Intérêt des méthodes d'apprentissage automatisé 109

5.2 Perspectives pour les bases de données en soins critiques 110

5.2.1 Limites de la constitution de cohortes 110

5.2.2 Challenge informatique 111

5.2.3 Développement d'aides décisionnelles au lit du patient 112

6 Conclusion **113**

Bibliographie **115**

A Mesures de performance pour les études de simulation **I**

B Présentation du cahier de recueil formalisé pour la cohorte AtlanRéa. **III**

C Étude de Danelian *et al.* Identification automatisée de potentielles violations de l'hypothèse de positivité **XXXV**

D Étude de simulation de Chatton *et al.* Comparaison des principales méthodes d'estimation causale selon les covariables incluses **LIII**

E Étude de simulation de Le Borgne *et al.* Apport de l'apprentissage automatique en g-computation **LXVII**

Abréviations

AIPW : *Augmented Inverse Probability Weighting*
ATE : *Average Treatment effect in the Entire population*
ATT : *Average Treatment effect on the Treated*
ATU : *Average Treatment effect on the Untreated*
CRF : Cahier de Recueil Formalisé
DAG : *Directed Acyclic Graph*
EHDEN : *European Health Data Evidence Network*
EDR : Estimateur doublement robuste
GC : G-Computation
IPTW : *Inverse Probability Treatment Weighting*
MIMIC : *Medical Information Mart for Intensive Care*
ML : *Machine Learning*
OMOP : *Observational Medical Outcomes Partnership*
SL : *Super Learner*
SP : Score de Propension
TMLE : *Targeted Maximum Likelihood Estimation*

Préambule

Il est reconnu actuellement que la pratique de la médecine doit se fonder sur des faits. On fait souvent référence à cette pratique de la médecine sous le terme d'*Evidence Based Medicine* (médecine basée sur les preuves). Ainsi, pour toute question de causalité, l'habitude est de mettre en place des essais cliniques contrôlés randomisés. Cependant, les études interventionnelles ne peuvent pas répondre à toutes les questions causales, s'intéressant à toutes les formes de stratégies thérapeutiques, pour tous les types de critères de jugement, et pour tous les groupes de population. Dans de nombreuses situations, ce type d'essai n'est pas réalisable. Dans ce cadre, l'utilisation de données observationnelles est une ressource possible.

Ces difficultés sont notamment fréquentes pour les patients de soins critiques pour lesquels l'inclusion dans les études est rendue difficile par l'urgence des prises en charge, mais aussi dans la difficulté à consistuer des populations homogènes. Cependant, il s'agit de populations pour lesquelles le recueil de données est prolifique, avec l'enregistrement itératif de paramètres biologiques, physiologiques issus du monitorage, ou même cliniques.

Dans mon exercice clinique d'anesthésiste-réanimateur, je suis confronté à la prise en charge de patients de soins critiques, et je participe à l'animation du réseau AtlanRéa (réseau des réanimations de l'ouest de la France), me permettant d'avoir accès au registre du réseau.

L'objectif de mon travail de thèse s'est donc naturellement orienté vers la question de l'inférence causale et l'utilisation des données observationnelles de patients de soins critiques.

La première problématique visait à analyser l'impact de l'utilisation précoce des barbituriques dans une population de patients traumatisés craniens. Pour se faire, nous avons donc utilisé la cohorte AtlanRéa, regroupant des patients de soins critiques, inclus dans différents centres hospitaliers de l'inter-région Ouest de la France. Dans ce but, nous avons utilisé une méthode statistique basée sur la pondération inverse des scores de propension.

Au cours de cette analyse d'inférence causale, nous avons rencontré différentes problématiques dans la vérification des hypothèses permettant d'approcher un lien de causalité. Nous avons notamment approché la problématique de l'infraction de l'hypothèse de positivité. Ce contexte nous a mené au deuxième objectif de cette thèse : effectuer des comparaisons de différentes méthodes statistiques d'inférence causale, dans différentes situations d'infraction de l'hypothèse de positivité, à partir de données simulées et par une application sur des données réelles.

Dans ce manuscrit de travail de thèse, nous ferons en introduction un état de l'art de l'inférence causale et qu'elles peuvent être ses intérêts dans le domaine des soins critiques. Nous poursuivrons par une présentation des résultats des travaux de cette thèse

(celui évaluant l'impact des barbituriques, puis celui comparant différentes méthodes statistiques dans des contextes d'infraction de l'hypothèse de positivité). Nous conclurons par une discussion et les perspectives envisageables dans un futur proche pour l'inférence causale dans le domaine des soins critiques.

Chapitre 1

État de l'art en inférence causale

Sommaire

1.1 Causalité 4
1.2 Intérêt des essais randomisés 4
1.3 Facteurs de confusion et biais d'indication 6
1.4 Évènements contrefactuels 7
1.5 Mesures de l'effet causal 7
1.6 Population d'intérêt 8
1.7 Trois hypothèses sous-tendant l'identifiabilité 8
1.8 Particularités de l'infraction de l'hypothèse de positivité 9
1.8.1 Infraction théorique et infraction aléatoire 9
1.8.2 Problématique de l'extrapolation 9
1.9 Modélisations statistiques et inférence causale 10
1.9.1 Scores de propension 10
1.9.2 G-computation 14
1.9.3 Estimateurs doublement robustes 15
1.9.4 Méthodes avec troncature 16
1.9.5 Évaluation du déséquilibre entre les groupes 16

1.1 Causalité

Afin de pouvoir étudier l'effet causal d'un traitement ou d'une exposition, les chercheurs ont recours à l'expérimentation standardisée d'un essai contrôlé randomisé permettant l'obtention de groupes comparables. Ces groupes ainsi créés ne se différencieront que par l'allocation à une des modalités du traitement expérimental. [1] Cependant, ce type d'essai est associé à certaines limites : la difficulté d'inclusion pour obtenir la population adéquate, le risque de déviations de protocole entre les groupes, la problématique de la généralisation des résultats, ou même les enjeux financiers et éthiques rencontrés. [2] L'utilisation de données observationnelles pouvant être issues de cohorte est une alternative intéressante. Dans ce cas, devant l'absence de comparabilité entre les groupes d'étude, l'inférence causale vise à mimer un essai contrôlé randomisé à partir de données observationnelles. [3]

Afin d'étudier une association causale d'une exposition et un évènement, des critères empiriques ont été listés par Austin Bradford Hill en 1965. [4] Aucun de ces critères n'est nécessaire, ni suffisant pour conclure à un lien causal :

— La force : la taille de l'association (valeur d'une différence ou rapport de deux probabilités) est importante.

— La reproductibilité : l'association a été reproduite indépendamment dans différentes études, dans différents lieux et à différents moments.

— La spécificité : l'association n'est dûe qu'à une seule et même exposition, en l'absence d'autres explications.

— La temporalité : l'exposition précède son effet.

— la relation dose-effet : il existe une relation entre la dose de l'exposition et le risque de l'évènement.

— La plausibilité : il existe un mécanisme expliquant l'association observée.

— La cohérence : l'association est en accord avec les connaissances déjà obtenues, et ne doit pas les contredire.

— L'expérimentation : l'association a été observée au moyen d'un plan expérimental.

— L'analogie : l'association ressemble à d'autres relations causales déjà décrites.

Á l'occasion du cinquantième anniversaire de ces critères, John P. A. Ioannidis soulignent que les critères les plus importants dans les relations de causalité dans le domaine de la santé sont la reproductibilité et l'expérimentation. [5] Cependant, la temporalité reste un critère primordial : la cause précède l'effet. [6] Par ailleurs, plusieurs auteurs proposent de se baser sur la plausibilité, et de recourir à des graphiques acycliques orientés (Directed Acycliq Graphs, DAG) pour guider l'interprétation de l'association. [7] En effet, les DAG sont fréquemment utilisés dans le domaine de l'épidémiologie pour représenter un système de relations.

1.2 Intérêt des essais randomisés

Considérons une intervention thérapeutique à deux modalités A (1 : sujet traité, 0 : sujet non traité), et un critère de jugement binaire Y (1 : survenue de l'évènement, 0 :

absence de survenue de l'évènement). Définissons $Y^{A=1}$ comme le devenir qui serait observé sous la modalité de traitement A = 1, et inversement $Y^{A=0}$ comme le devenir qui serait observé sous la modalité d'absence de traitement (A = 0).

L'existence d'un effet causal pour le traitement A suppose que le devenir individuel Y est différent entre le régime $a = 1$ que $a = 0$ (nous notons le statut de A par a lorsque l'on s'intéresse au devenir individuel d'un sujet i pour lequel son devenir serait Y_i^a). Ainsi, l'effet causal pour un individu i, suppose que $Y_i^{a=1} \neq Y_i^{a=0}$.

L'essai contrôlé randomisé est un type d'étude permettant de comparer les différentes modalités thérapeutiques de A, sur le critère de jugement Y. Dans la situation la plus répandue d'essai randomisé avec deux bras de traitement, et avec une randomisation d'un ratio 1 : 1, chaque individu reçoit une des deux modalités thérapeutiques avec une probabilité égale à 0,5. Cette probabilité est la même pour tous les individus. Ainsi, au moment de la randomisation, les deux groupes constitués apparaissent comparables sur leurs caractéristiques, et ne diffèrent que par leur statut de traitement expérimental. Cette notion fait référence à l'*échangeabilité* entre les groupes : les individus (qu'ils soient traités ou non) ont le même risque moyen de faire l'évènement avant de recevoir une modalité de traitement. Ces groupes sont échangeables, et l'effet causal serait aussi mis en évidence si nous intervertissions les modalités de traitement entre les groupes. La notion d'échangeabilité permet de conclure qu'une différence observée sur le critère de jugement (Y) peut être considérée comme uniquement liée à la modalité de traitement attribuée. La mesure de la différence au niveau du critère de jugement permet l'estimation de l'effet causal du traitement. Cette mesure est effectuée au niveau de la population, ce qu'on définit usuellement comme une mesure marginale.

La Figure 1 représente la différence entre ce qui s'apparente à l'*association* et ce qui représente la *causalité*. La définition de *causalité* implique une comparaison entre tous les individus qui seraient traités par rapport à tous les individus que ne seraient pas traités. De manière différente, l'association se focalise sur la comparaison entre les sujets traités et ceux qui ne le sont pas. Comme nous venons de le définir, lorsque l'on s'intéresse à la population entière, nous parlons d'une probablité marginale, qui est inconditionnelle : $P(Y^{A=a} = 1)$. A l'inverse, lorsque l'*association* est étudiée, nous nous intéressons à deux sous-ensembles de la population, déterminés par les modalités observées individuelles de traitement, faisant référence à une probabilité conditionnelle : $P(Y = 1|A = a)$.

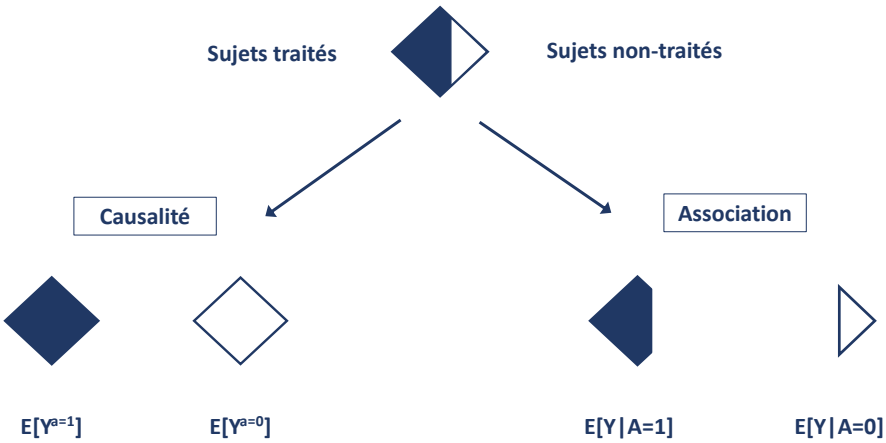


FIGURE 1 – Différence entre les principes de causalité et d'association, issu et traduit de Hernan, 2020. [3]

1.3 Facteurs de confusion et biais d'indication

Dans le cadre des données observationnelles, la modalité thérapeutique ou l'exposition étudiée n'est pas randomisée, et donc la comparaison des groupes de traitement est impactée par le biais d'indication.

En effet, les analyses à partir de données observationnelles peuvent faire intervenir des facteurs de confusion (ensemble de variable noté L), définis comme des facteurs causant à la fois à l'exposition d'intérêt (le statut de traitement A) et au critère de jugement Y. L'estimation de l'effet causal de A pourrait être biaisée, si son statut de traitement dépend des facteurs L. Dans le cadre des données observationnelles, il est d'ailleurs fréquent que les sujets ayant un risque plus élevé de présenter l'évènement d'intérêt Y (selon par exemple leur âge) aient une probabilité plus importante d'être traités. Cette situation fait référence au biais d'indication.

En conceptualisant ces relations par un DAG, via des nœuds représentant les variables et des flèches représentant les effets causaux, nous pouvons obtenir l'illustration représentée dans la Figure 2.

Nous pouvons ainsi définir des *variables instrumentales* qui sont des causes du traitement A mais sans relation sur l'évènement Y, et inversement les *facteurs de risque* qui sont des causes de Y sans impacter A. La *variable médiatrice* est une variable présente dans la relation causale de deux variables, alors que la *variable collider* est une conséquence commune à deux autres variables. Nous définissons la *variable proxy* comme une variable mesurée pouvant estimer une variable non-mesurée.

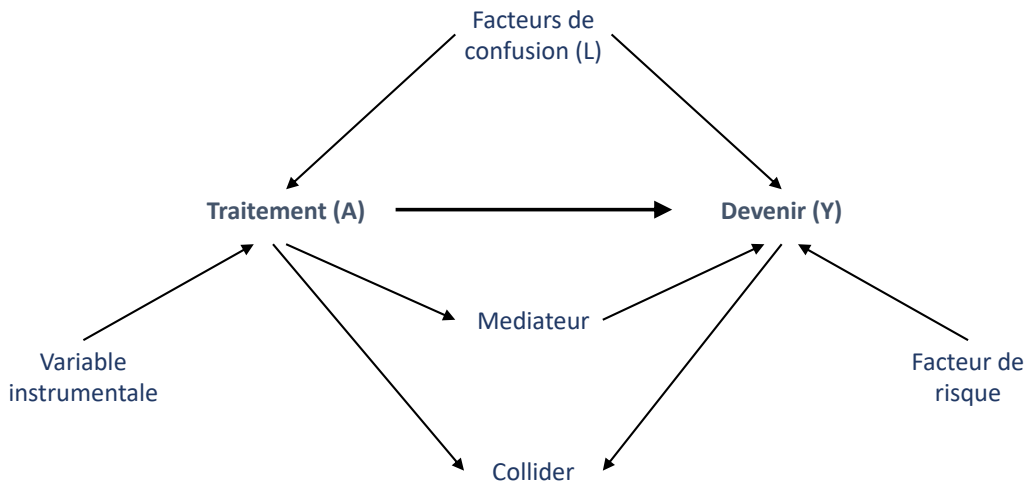


FIGURE 2 – Représentation des rôles des variables au sein d'un diagramme causal

1.4 Évènements contrefactuels

Afin d'estimer la causalité, Rubin propose en 1974 le concept des mondes contrefactuels. [8] Ce concept se base sur l'estimation de l'effet si le patient reçoit une autre modalité du traitement (par exemple, dans le cadre du traitement A à deux modalités : A = 1 en place de A = 0). Un seul évènement peut réellement être observé, les autres seront alors définis comme étant contrefactuels. Ainsi, pour chaque individu, il existe un évènement potentiel selon la modalité d'exposition étudiée.

Ces évènements contrefactuels ne sont pas mesurables, mais permettent d'approcher un effet causal ciblé, appelé *estimand théorique*; qui est à différencier de l'effet mesuré à partir des données observées, appelé *estimand empirique*. [9, 10]

1.5 Mesures de l'effet causal

Nous nous concentrerons sur les expositions de type binaire qui représentent le type de variable le plus fréquemment utilisé dans la littérature médicale. Il existe différentes manières de matérialiser un effet causal binaire. Nous pouvons présenter :

— une différence de risque : $P(Y^{A=1} = 1) - P(Y^{A=0} = 1)$;

— un risque relatif : $\frac{P(Y^{A=1}=1)}{P(Y^{A=0}=1)}$; ou

— un odds-ratio : $\frac{P(Y^{A=1}=1)/(1-P(Y^{A=0}=1))}{P(Y^{A=0}=1)/(1-P(Y^{A=1}=1)}$.

Chacune de ces mesures de l'effet pourra être utilisée en fonction de l'objectif de l'inférence. Le risque relatif, comme l'odds-ratio représentent une échelle multiplicative alors que la différence de risque est basée sur une échelle additive.

1.6 Population d'intérêt

Si l'on considère l'ensemble de la population d'intérêt, nous pouvons mesurer l'effet du passage de l'ensemble de la population du statut de "non-traité" à "traité". Cet effet est défini dans la littérature comme l'effet dans la population entière (*average treatment effect on the entire population* - ATE). Il représente la différence moyenne des évènements dans les deux mondes contrefactuels : $E(Y^1 - Y^0)$.

Si l'on s'intéresse à la mesure de l'effet du passage de l'ensemble de la population effectivement traitée (A = 1) du statut de "non-traité" à "traité", nous estimerons l'effet dans la population des individus traités (*average treatment effect on the treated* - ATT). Dans ce cas, l'intérêt serait de mettre en évidence que le traitement a un effet chez les individus qui le reçoivent. De la même manière, nous pouvons obtenir l'effet causal moyen chez les individus non-traités (*average treatment effect on the untreated* – ATU), se déterminant dans la sous-population des individus non-traités (A = 0). [11]

Afin que l'estimand théorique corresponde bien à l'estimand empirique, et qu'il soit considéré identifiable, plusieurs hypothèses successives nécessitent d'être vérifiées. Dans ce but, Petersen *et al.* propose un processus à plusieurs étapes. [12] La première étape n'implique pas les données, et nécessite de recourir aux connaissances pour estimer le modèle causal à partir des hypothèses d'association (via l'utilisation d'une équation structurale non-paramétrique, [13] ou via l'aide d'un DAG). Les étapes suivantes demanderont de spécifier les caractéristiques de recueil des données, spécifier la grandeur causale cible, pour ensuite vérifier l'identifiabilité (*i.e.* l'effet causal cible peut-être représenté par la distribution des données observées). Il s'en suit des étapes de vérification d'estimation statistique en choisissant un estimand théorique qui permet d'approximer l'effet causal d'intérêt, tout en respectant les limites de la modélisation. L'interprétation finale du résultat dépendra des niveaux précédents et pourra aller de l'interprétation statistique isolée issue des données observées, au résultat qui serait observé dans un essai randomisé hypothétique.

Au cours de ce processus, nous pouvons schématiquement rencontrer deux grands types de biais : le *biais d'identification*[1] ou le *biais d'estimation*. [14] Le biais d'identification correspond au non-respect d'au moins une des conditions d'identifiabilité, alors que le biais d'estimation est en lien avec un problème de modélisation (*i.e.* une différence entre l'estimation obtenue à partir des données et l'estimand empirique). Ainsi, le biais d'identification peut être atténué en optimisant les modèles théoriques de causalité et demande la connaissance des mécanismes physiopathologiques sous-tendant la question analysée. [15] Une analyse d'inférence causale ne pourra pas se passer de connaissances préalables sur les associations théoriques présupposées de la relation étudiée. Concernant le biais d'estimation, il pourra être optimisé en améliorant les méthodes statistiques d'estimation.

1.7 Trois hypothèses sous-tendant l'identifiabilité

Afin de vérifier les conditions d'identifiabilité permettant de conceptualiser une étude observationnelle comme une expérimentation conditionelle randomisée, trois hypothèses devront être vérifiées : la *consistance*, l'*échangeabilité* et la *positivité*.

1. le *biais d'identification* est aussi appelé *biais causal*

La *consistance* suppose que les évènements observés concordent avec ceux qui sont potentiels (contrefactuels). Pour un individu réellement traité ($A_i = 1$), son devenir contrefactuel sous le statut traité ($Y_i^{a=1}$) est équivalent à son devenir observé Y_i.

Comme décrit précédemment, l'*échangeabilité* suppose que les individus traités et non-traités ont le même risque moyen de faire l'évènement avant d'être exposé à une des modalités du traitement A : $Y^a \perp\!\!\!\perp A$, où $\perp\!\!\!\perp$ désigne l'indépendance [16]. Cela sous-entend que la probabilité conditionnelle de recevoir chaque modalité du traitement ne dépend que des covariables mesurées.

La *positivité* suppose que n'importe quel individu de la population a une probabilité non-nulle de recevoir une des deux modalités du traitement A. Ainsi, chaque individu doit théoriquement pouvoir recevoir une des deux modalités du traitement étudié. Nous pouvons dire que l'hypothèse de positivité est vérifiée si $0 < P(A = a|L = l) < 1$, où la positivité est nécessaire que pour les variables L impliquées dans l'hypothèse d'échangeabilité. Imaginons une analyse de causalité évaluant un traitement sur la survenue de décès hospitalier, il ne sera pas nécessaire de s'assurer que la probabilité de recevoir le traitement est positive pour les sujets avec des cheveux roux, si le fait d'avoir des cheveux roux n'est pas nécessaire pour assurer l'échangeabilité. Le fait d'avoir des cheveux roux ne devrait pas être considéré comme un statut de prédicteur indépendant du devenir conditionnel de mortalité hospitalière.

1.8 Particularités de l'infraction de l'hypothèse de positivité

1.8.1 Infraction théorique et infraction aléatoire

L'infraction ou la violation de l'hypothèse de positivité peut survenir dans deux cadres différents.

D'une part, nous identifions la *violation structurelle* (aussi appelée violation théorique) de la positivité impliquant une strate d'individus dans la population qui ne peut pas théoriquement recevoir le traitement (si par exemple on étudie l'impact de la césarienne dans une population intégrant des hommes).

D'autre part, nous pouvons mettre en évidence des situations de *violations aléatoires* (appelées en anglais *near* ou *practical violation*) pouvant survenir par échantillonnage. Ces situations se rencontrent lorsque la taille de la population est faible ou que la prévalence de traitement est rare. [17]

La *violation structurelle* de la positivité est une conséquence d'un problème de conception dans la modélisation de la question causale, et doit faire discuter la possibilité de restreindre la population d'étude. [17, 18] Par contre, dans les cas d'*infraction aléatoire* à l'hypothèse de positivité, la population cible de l'étude est bien définie, et la solution peut résider dans l'optimisation de l'impact de l'échantillonnage sur l'estimateur statistique.

1.8.2 Problématique de l'extrapolation

Lorsque sont utilisées des méthodes statistiques dont la construction cherche à modéliser le critère de jugement, une extrapolation va être nécessaire dans la strate d'individus affectés par la violation de la positivité.

Imaginons, en reprenant le traitement binaire A, le critère de jugement binaire Y et par exemple un facteur de confusion quantitatif L comme l'âge. Nous pouvons définir f(L|A) comme la fonction de densité de la variable quantitative L conditionnellement à A. Nous supposons que L est un vrai facteur de confusion, étant une cause de l'allocation du traitement de A (un sujet agé aura moins tendance à être traité), mais aussi à la survenue de l'évènement (un sujet âgé aura une plus grande probabilité de présenter l'évènement, comme par exemple un décès hospitalier). Considérons une violation aléatoire de la positivité lorsque L dépasse un certain seuil α avec une augmentation théorique de la probabilité de survenue du décès hospitalier chez les sujets traités (A = 1) pour des valeurs supérieures au seuil α. A cause du manque d'information due à l'infraction de l'hypothèse de positivité avec peu d'individus dans l'échantillon ayant un âge supérieur au seuil α, l'estimation de l'effet thérapeutique est extrapolé à partir des données observées (les individus avec $L_{âge}$ inférieur au seuil α). Et même si la modélisation statistique est parfaitement spécifiée pour les données observées, elle pourrait être inadaptée pour cette région souffrant d'une infraction de la positivité. [18]

La difficulté de cette situation est qu'aucun moyen statistique ne permet d'identifier une problématique d'extrapolation. Seule l'infraction de l'hypothèse de positivité sera identifiable.

Le choix de la méthode statistique et des covariables introduites dans la modélisation sont des paramètres primordiaux dans ces situations problématiques de violation aléatoire de l'hypothèse de positivité, pouvant aboutir à une problématique d'extrapolation.

1.9 Modélisations statistiques et inférence causale

Actuellement, il existe différentes alternatives statistiques pour modéliser une question causale. Tout d'abord, les méthodes basées sur les scores de propension (SP) sont celles qui sont actuellement les plus utilisées et rapportées dans la littérature médicale. Le recours aux scores de propension continue d'augmenter au fil des années récentes (Figure 3).

Les méthodes basées sur les SP ont pour objectif de modéliser l'allocation au traitement. [19] De manière alternative, la *g-computation* ou *g-standardisation* est une méthode basée sur le principe des mondes contrefactuels et vise à prédire l'évènement contrefactuel. [20–22] Enfin des méthodes plus récentes se basent sur une combinaison des deux méthodes précédentes pour aboutir à un estimateur considéré comme doublement robuste. [23]

1.9.1 Scores de propension

Le principe des scores de propension

Comme décrit précédemment, le SP vise à représenter la probabilité d'allocation au traitement en fonction des variables d'ajustement L afin d'obtenir deux groupes équilibrés. Nous définissons donc le SP comme la probabilité conditionnelle aux variables d'ajustement : P(A = 1|L).

Les analyses basées sur le SP permettront d'avoir une similarité sur les variables d'ajustement L entre le groupe d'individus traités et celui des individus non-traités dans l'ob-

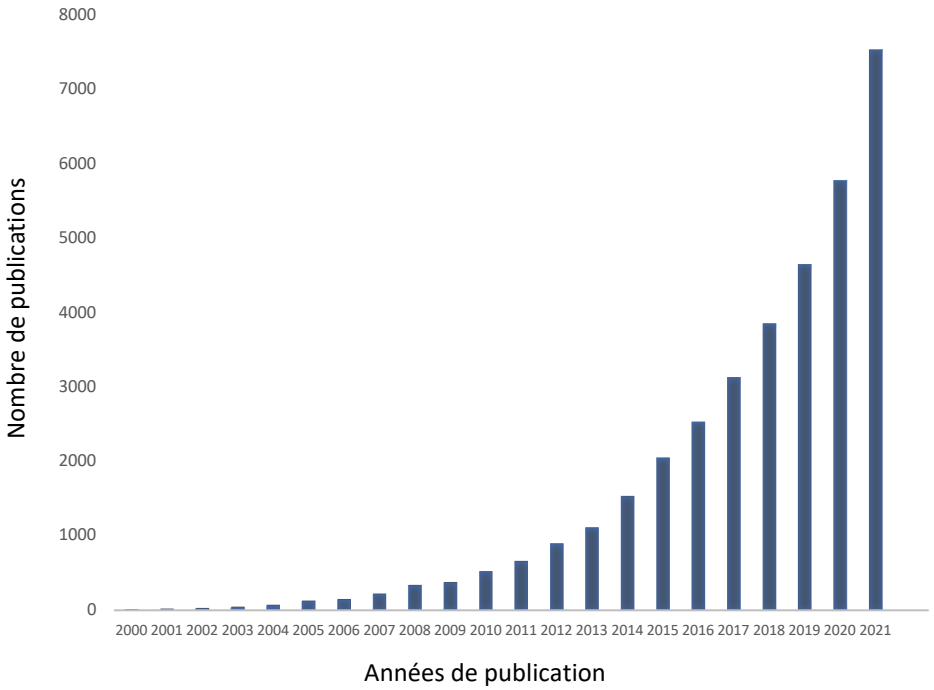


FIGURE 3 – Evolution temporelle des publications dans PubMed® faisant référence aux scores de propension soit dans leur titre, soit dans leur abstract

jectif de diminuer le biais de confusion. Ainsi, nous obtiendrons une seule valeur représentant la probabilité d'être traité et résumant la distribution de l'ensemble des facteurs de confusion. Le recours au SP permettra de se positionner dans le cadre d'un hypothétique essai randomisé (*i.e.* vérifiant l'hypothèse d'échangeabilité), et facilitant l'inférence causale. Dans ce cadre, il est important de souligner que l'objectif du recours au SP doit être d'équilibrer les covariables dans la population contrefactuelle, et non de prédire parfaitement l'allocation au traitement. [24]

Différentes méthodes statistiques peuvent être utilisées pour estimer les SP. La régression logistique est la technique la plus souvent décrite. Des méthodes plus récentes et basées sur des algorithmes d'intelligence artificielle (apprentissage automatisé, *machine learning* - ML) peuvent aussi être utilisées. [17, 24, 25]

Une fois les SP estimés, quatre approches différentes peuvent être considérées : l'*ajustement conditionnel*, la *stratification*, l'*appariement* et la *pondération*. [19, 26]

Aussi séduisante qu'est la méthodologie des SP, il persiste des limites à son utilisation. Le SP ne pourra pas corriger les facteurs de confusion non mesurés et non intégrés dans leur modélisation. [27]

Mais cela ne doit pas effacer les intérêts potentiels du recours au SP. L'utilisation d'une méthode basée sur les SP est notamment préférable quand on peut facilement identifier le modèle d'allocation au traitement, par rapport à celui de survenue de l'évènement. Par ailleurs, l'intérêt des SP est aussi de vérifier la performance d'une méthode à équilibrer les covariables entre les groupes. Il est assez aisé de tester l'hypothèse de positivité via l'utilisation des SP et de redéfinir la population en excluant ceux qui sont assurés de ne recevoir qu'une des deux modalités de l'intervention thérapeutique. En effet, en interprétant la distribution des SP, chez les individus traités et ceux non traités, l'investigateur peut investiguer le chevauchement (appelé *overlap*), qui pourrait apparaître insuffisant

entre ces groupes.

Analyse du chevauchement des scores de propension

La représentation graphique des SP dans chaque groupe est une technique simple et rapide pour mettre en évidence un chevauchement entre les probabilités d'être traité entre chacun des groupes. [18, 28]

Pour aboutir à une représentation d'une infraction de l'hypothèse de positivité, Westreich et Cole proposaient d'utiliser des tableaux de contingence pour l'ensemble des strates possibles des variables d'ajustement. [17] Cette approche très fastidieuse ne peut être envisagée que dans le cadre d'un très faible nombre de facteurs de confusion, et notamment pour des variables catégorielles. L'autre alternative existante pour quantifier l'absence de chevauchement est proposée par Petersen, puis a été modifiée par Bahamyirou. [18, 29] Cette approche se base sur l'utilisation du bootstrap paramétrique comme un outil diagnostique d'infraction de l'hypothèse de positivité, et donc d'un manque de chevauchement. La Figure 4 représente cette absence de chevauchement dans une représentation de distribution des SP dans deux groupes de sujets traités et non-traités.

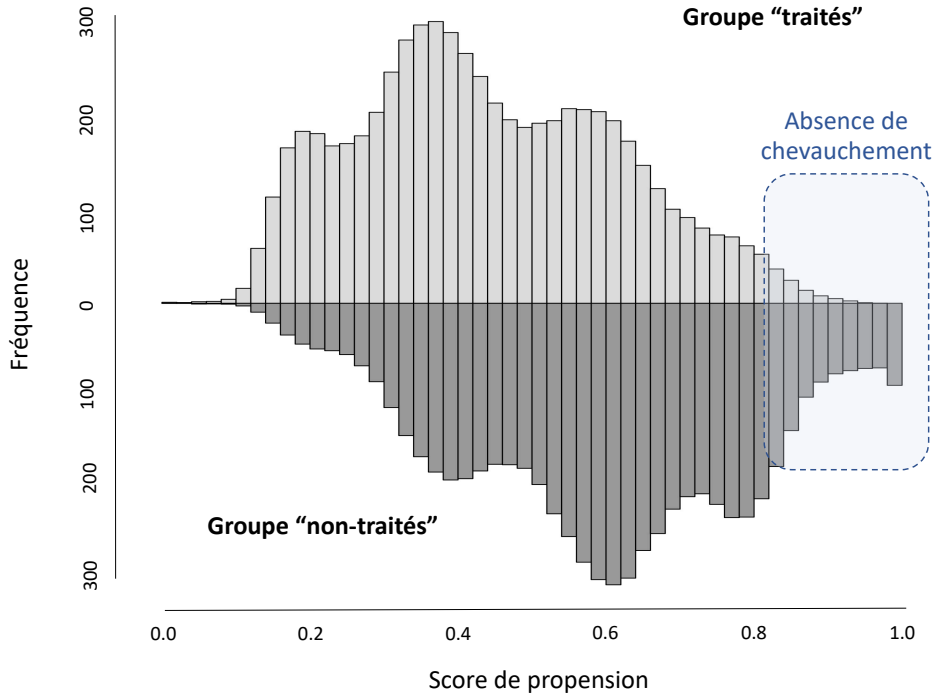


FIGURE 4 – Représentation de l'absence de chevauchement pour des distributions de scores de propension

Ajustement sur les scores de propension

L'ajustement vise à utiliser le SP comme variable d'ajustement dans un modèle statistique représentant la relation entre le traitement et l'évènement : $E(Y|A,L) = \beta + \alpha_1 A + \alpha_2 SP$. Les limites de l'ajustement se posent dans le cadre du risque de mauvaise spécification du modèle, dans l'estimation d'un effet qui est conditionnel (*sujet-spécific*), et

surtout dans la difficulté à obtenir une comparabilité entre les groupes. [30] De ce fait, l'ajustement sur le SP est actuellement non-recommandé. [31]

Stratification sur les scores de propension

La stratification sur le SP est une méthode alternative. Le rationnel réside dans l'organisation de la population en strates d'individus ayant des valeurs proches de SP, permettant d'estimer au sein de chaque strate un effet causal du traitement et de moyenner un effet global. Du fait de la méthode, la stratification sur le SP estimera un effet conditionnel. Plus le découpage de la population par strate sera fin, plus faible sera la confusion résiduelle au sein de chaque strate. Pour une variable d'exposition binaire, Rosenbaum et Rubin proposaient d'utiliser au moins cinq strates basées sur les quintiles du SP pour éliminer le biais. [19] La difficulté va se poser lorsque la prévalence au traitement est rare, le SP moyen sera faible et l'information se concentrera dans les strates avec les SP les plus importants (celles où il existera moins d'individus). Cela peut mener à une confusion résiduelle restant importante au sein des strates des SP les plus faibles. Pour pallier ce problème, il faudra s'assurer d'avoir exclu les sujets pouvant être à risque de violation de la positivité, augmenter le nombre de strates et vérifier que l'équilibre est atteint au sein de chaque strate.

Appariement sur le score de propension

L'appariement sur le SP fait aussi parti des premières méthodes décrites. Après avoir assigné à chaque individu une valeur de SP, un groupe est désigné comme le groupe ciblé (généralement le groupe avec le plus faible effectif, qui est souvent le groupe des individus traités). Pour chacun des sujets du groupe cible, un individu similaire du groupe comparateur sera identifié. La difficulté réside dans la définition de la similarité entre les groupes (basée sur la valeur du SP). L'analyste pourra choisir d'attribuer aléatoirement à un individu cible un individu similaire du groupe comparateur jusqu'à avoir attribué un sujet semblable à tous les individus du groupe cible. Cette méthode est appelée l'*appariement selon le plus proche voisin*. Sinon, l'analyste pourra préférer minimiser, via un algorithme, la distance entre les valeurs de SP des individus du groupe cible (e.g. les individus traités) et ceux du groupe comparateur, et cette stratégie correspond à l'*appariement optimal*.

Des limites sont souvent utilisées pour éviter les différences trop importantes de valeur de SP entre deux individus similaires : *les calipers*. Ces seuils représentent le degré de tolérance de la similarité entre deux patients. [32, 33]

Par ailleurs, l'appariement pourra être d'un pour un individu, ou d'un individu pour plusieurs individus. Et l'appariement pourra être avec remise (*i.e.* un individu du groupe comparateur pourra être apparié à plusieurs individus du groupe cible), ou sans remise. [34]

La méthode d'appariement selon le plus proche voisin, avec un caliper fixé à 0.2 écart-type du logit du SP, et sans remise de l'individu, est une stratégie performante d'appariement. [32] Une fois l'appariement réalisé, le devenir dans chacun des deux groupes créés pourra être comparé sachant que l'hypothèse d'exchangeabilité est supposée et que la confusion est contrôlée. [35]

La limite de l'appariement se pose dans les situations de risque d'infraction de l'hypothèse de positivité, où la proportion de la population cible qui pourra être appariée sera

faible, et dont l'estimation de l'effet traitement aboutira à une mauvaise approximation. Être moins strict sur les limites des *calipers* peut être une solution pour conserver une taille suffisante de la population cible mais au risque de perdre le contrôle de la confusion. Le recours aux *calipers* peut être considérée comme une optimisation automatique de la population d'étude mais au risque de ne pas réussir à définir la population analysée *in fine*. L'issue résidera souvent dans la redéfinition par le clinicien de la question causale et de la population analysée. [36, 37]

Il est important de noter que la méthode utilisant l'appariement telle que décrite tend à approcher une estimation de l'ATT ou de l'ATU, et que pour estimer un ATE, il faudrait utiliser une approche de type *optimal full matching* (combinant l'appariement de strates de la population dont la somme absolue des distances dans chaque strate est optimisée, et dont des poids pour chaque individu sont estimés en fonction de l'appartenance à la strate). [38]

Pondération sur le score de propension

La pondération par le SP peut être utilisée pour créer des pseudo-populations de sujets traités et non-traités dont la balance des covariables mesurées serait assurée. L'analyste peut modifier le poids de chaque sujet à la hausse ou à la baisse, avec le poids comme fonction du SP. [26, 39]

Différentes populations d'intérêt peuvent être utilisées : la population totale (pondération inverse sur la probabilité de traitement – inverse probability of treatment weighting, IPTW), la population des individus traités (ATT), ou par exemple la population avec un intérêt pour la région de chevauchement (*overlap weighting*). [40]

Les poids peuvent être stabilisés pour améliorer la précision des estimateurs issus de la pondération, et pour obtenir un pseudo-échantillon pondéré d'une taille similaire à l'échantillon initial. [26, 41] Lors de l'estimation de la variance, il sera nécessaire de considérer la nature pondérée de l'échantillon, et de recourir à une matrice de variance robuste de type sandwich, [26], ou d'utiliser le bootstrap. [42]

La difficulté de la pondération réside dans la gestion des poids les plus extrêmes, représentant les sujets les plus atypiques. Par exemple, quand un individu a un très faible SP, il aura un poids important, et particulièrement lorsque l'on estime l'ATE. Ces sujets représentent souvent les individus traités contrairement à la prédiction de l'être, et pourront avoir une influence importante sur les résultats, tout en ajoutant de la variance sur l'estimation de l'effet. C'est pourquoi, chacun doit rester vigilant quand il utilise la méthode de la pondération sur les SP à la méthode choisie pour gérer les poids les plus extrêmes.

1.9.2 G-computation

De manière assez concomitante à la proposition de l'utilisation des SP par Rosenbaum et Rubin, Robins propose une méthode alternative d'estimation causale basée sur une formule générale (*g-formula*) pour estimer un effet causal. [20]

Cette approche a pour objectif de faire de l'inférence causale dans le cadre de confusion dépendante du temps. Elle est basée sur une succession de prédictions contrefactuelles de l'évènement, suivie d'une étape de standardisation. Dans notre cadre, avec l'absence de confusion dépendante du temps et avec une seule allocation du traitement,

l'utilisation de la *g-formula* permet d'obtenir l'estimation de la probabilité des deux évènements potentiels sous les deux modalités de traitement : $P(Y^a = 1)$ où $a \in \{0,1\}$.

Robins propose une forme paramétrique issue de la *g-formula*, qui est dénommée *g-computation* (GC), parfois aussi appelée *g-standardisation*.

Nous pouvons découper son utilisation en quatre étapes [22] :

1. Modéliser la survenue de l'évènement chez tous les individus et en fonction des facteurs de confusion, aboutissant au modèle prédisant la survenue de l'évènement (*outcome model*) : $Q(A,L)$;
2. Obtenir à partir de la population initiale, deux copies de populations contrefactuelles ne différant seulement que sur le statut de traitement : une population où tous les sujets auront $A = 1$, et l'autre population où tous les sujets auront $A = 0$;
3. Appliquer l'*outcome model*, $Q(A = a, L)$ dans chaque copie contrefactuelle de la population pour obtenir les probabilités individuelles de faire l'évènement ;
4. Calculer les deux moyennes des probabilités individuelles de survenue de l'évènement, sous $A = 1$ et sous $A = 0$.

Dans ce cadre, la variance sera estimée par bootstrap ou par simulation paramétrique. Afin d'obtenir les effets ATT ou ATU, il suffit de restreindre l'échantillon aux individus traités ou non-traités, respectivement.

1.9.3 Estimateurs doublement robustes

Les méthodes basées sur les estimateurs doublement robustes (EDR) combinent les deux stratégies précédentes : l'utilisation d'un *treatment model* (celui permettant d'estimer le SP) avec un *outcome model* (méthodologie à la base de la GC). Alors que l'utilisation du SP nécessite une bonne spécification du modèle prédisant l'allocation au traitement, et alors que le recours à la GC nécessite une bonne spécification du modèle prédisant la survenue de l'évènement, le principe des EDR est d'obtenir une estimation non biaisée si au moins un des deux modèles est bien spécifié. [43]

Pour aboutir à un EDR, deux méthodes ont été décrites en fonction de leur combinaison : soit via l'utilisation de la GC au sein de l'IPTW, ou recourir à l'IPTW pour optimiser l'estimation de la GC.

La première méthode décrite est l'Augmented-IPW (AIPW) qui consiste en l'utilisation de la fonction de prédiction de l'évènement $Q(A,L)$ pour améliorer l'estimation obtenue par IPTW. [23]

L'autre méthodologie, décrite en 2006 par van der Laan et Rubin est la Targeted Maximum Likelihood Estimator (TMLE). [44] L'objectif est d'obtenir un EDR en utilisant le SP pour réduire la confusion résiduelle au sein de la GC. Les performances de la TMLE apparaissent être meilleures que celle de l'AIPW. [45]

Afin d'obtenir l'estimation de la TMLE, différentes étapes doivent être suivies :

1. Estimer le modèle de l'évènement (*outcome model*) : $Q(A,L)$;
2. Estimer l'allocation au traitement (*treatment model*) : $g(L)$;
3. Optimisation du modèle de l'évènement en utilisant $g(L)$ pour obtenir $Q^*(A,L)$;
   (a) Nécessité d'obtenir deux fonctions contrefactuelles de $g(L)$: H^a ($a \in \{0,1\}$)

(b) Estimation des paramètres de fluctuation du modèle optimisé

$$\text{logit}[Q^*(A,L)] = \text{logit}[Q(A,L)] + \delta^0 \times H^0 + \delta^1 \times H^1$$

4. Estimation de l'effet causal comme dans les dernières étapes de la GC avec obtention de moyennes des probabilités individuelles.

Une des particularités de la TMLE réside dans l'utilisation des *clever covariates*, correspondant aux fonctions H^a, et ressemblant aux poids utilisés dans l'IPTW; mais aussi dans le recours aux paramètres de fluctuation, matérialisant l'association des *clever covariates*. Que ce soit l'AIPW ou la TMLE, la variance peut être obtenue mathématiquement par une fonction d'influence, [46] ou par bootstrap.

Derrière cet intérêt pour les EDR, et notamment la TMLE, il faudra se méfier de la situation où les deux modèles (*treatment model* et *outcome model*) sont mal spécifiés, ce qui pourra impacter la performance de l'estimation. [47] De plus, l'estimation de la variance par les fonctions d'influence peuvent être altérées dans les contextes de faible taille d'échantillon, de valeurs extrêmes de la prévalence du traitement ou même dans le cadre d'une survenue rare de l'évènement. Ces situations peuvent augmenter l'erreur de type I et la couverture par l'intervalle de confiance. [43]

1.9.4 Méthodes avec troncature

Nous avons vu avec l'IPTW que la difficulté de la pondération réside dans la gestion des poids les plus extrêmes, représentant les sujets les plus atypiques. Une solution est de recourir à la *troncature*[2] (attribuer une même valeur de SP à tous les individus ayant une valeur au-delà d'un seuil) ou au *rognage*[3] (suppression des SP dépassant un certain seuil). Ces méthodes ont pour objectif de limiter la contribution des observations les plus extrêmes. [36, 48] La *troncature* permet de garder la globalité de la population d'étude, alors que la méthode de *rognage* aura l'inconvénient de réduire la population, et pourra modifier l'estimand théorique. Dans le cadre du recours au *rognage*, il sera de la responsabilité de l'analyste de s'assurer au préalable que la violation de la positivité est aléatoire, ou dans le cas contraire, de modifier sa population d'étude.

Pour la gestion des sujets les plus extrêmes, ces méthodes vont permettre de diminuer la variance de l'estimation, mais au prix d'une augmentation du biais. [28, 49] Ainsi, la grande difficulté réside dans le choix du seuil pour obtenir le meilleur rapport entre le gain de variance et la perte sur le biais. Des alternatives algorithmiques sont proposées pour adapter ce choix en fonction des données, dans le but d'optimiser la gestion de ces sujets extrêmes dans les situations d'infraction aléatoire de l'hypothèse de positivité. Ces méthodes basées sur les données ont été proposées pour l'IPTW, [50] ou même pour la TMLE. [49]

1.9.5 Évaluation du déséquilibre entre les groupes

Nous avons vu que pour la stratification, l'appariement ou même la pondération sur les SP, il est nécessaire de vérifier l'équilibre des facteurs de confusion entre les groupes.

2. traduction de *truncation*
3. traduction de *trimming*

Le paramètre le plus souvent utilisé est la différence standardisée entre les groupes, en considérant le plus souvent une différence devant être inférieure à 10%. [51, 52] Plus la différence standardisée est importante, plus le déséquilibre entre les groupes est marqué. Ainsi, lorsque les différences standardisées sont importantes, le chercheur doit revoir sa modélisation. A noter que l'utilisation des valeurs de p ne permettent pas d'évaluer de manière fiable entre les groupes. [53]

Qu'importe le paramètre choisi, plusieurs itérations et modifications dans la sélection des variables introduites dans la modélisation pourront être nécessaires pour obtenir un équilibre acceptable, réduisant la confusion des covariables mesurées, tout en améliorant la confiance dans les résultats. [54] A partir de cette idée d'itération, des alternatives basées sur le ML ont émergées. [55–57] Ces techniques ont l'objectif d'optimiser un paramètre, tel que la différence standardisée moyenne entre toutes les covariables ou une performance globale prédictive. Il faudra cependant s'assurer d'utiliser des validations croisées et un intervalle de confiance adéquate pour éviter le surajustement. [58]

Nous pouvons néanmoins discuter de l'intérêt d'utiliser un paramètre tel que la différence standardisée moyenne reflétant le déséquilibre global, car cette valeur ne prendra pas en compte l'effet propre du biais de confusion que pourraient avoir certaines variables associées à l'évènement. Nous pourrions observer une valeur acceptable de différence standardisée moyenne avec un déséquilibre faible sur plusieurs facteurs de confusion ayant peu d'impact, mais un déséquilibre prononcé sur un seul facteur de confusion fortement associé à l'évènement. Il apparait donc nécessaire de considérer le degré d'association avec l'évènement. [35, 59, 60] Une solution proposée est de recourir à une mesure de balance qui pondère chaque différence standardisée selon sa valeur pronostique pour délivrer une valeur globale du déséquilibre reflétant l'impact des variables pronostiques. [61]

Après avoir fait un tour d'horizon de l'état de l'art de l'inférence causale à partir de données observationnelles et les méthodes utilisées pour y parvenir, je vous propose de nous intéresser aux applications dans le domaine des soins critiques.

Chapitre 2

Application aux soins critiques

Sommaire

2.1 Inférence causale dans les soins critiques 20
2.1.1 Les limites des études contrôlées randomisées 20
2.1.2 Données observationnelles et soins critiques 20
2.1.3 Exemple du projet MIMIC 21
2.1.4 Autres bases de données en soins critiques 22
2.2 Présentation de la cohorte AtlanRéa 23
2.2.1 Présentation de la cohorte 23
2.2.2 Données collectées 24
2.2.3 Description de la cohorte (extraction de janvier 2021) 26

2.1 Inférence causale dans les soins critiques

2.1.1 Les limites des études contrôlées randomisées

Alors que les essais randomisés ont fondamentalement changé la pratique médicale, et que la randomisation apparait être le *gold standard* pour estimer sans biais l'effet d'une intervention, certaines limites de ces études remettent en question leur hégémonie, et notamment dans le domaine des soins critiques. [62, 63]

Une des premières limite à discuter est la simplification de l'intervention qui est imposée par le processus de randomisation, avec le plus souvent la comparaison de deux bras interventionnels, alors que la pratique clinique suppose souvent plusieurs modalités de traitement.

Deuxièmement, le calcul de la taille de population est souvent beaucoup trop optimiste, menant à ne mettre en évidence que des effets larges. Ces essais, et notamment dans le domaine des soins critiques, sont souvent négatifs et ne permettant pas d'aboutir à une perspective clinique. Cela peut même mener à se poser la question éthique des essais randomisés sous-dimensionnés qui n'ont pas permis d'évaluer certaines interventions avec la puissance suffisante. Cela peut nous emmener à une situation de confusion, qu'on appelle "l'effet pendule" (*pendulum effect*), alternant des essais randomisés positifs et négatifs s'intéressant à la même intervention thérapeutique.

Par ailleurs, les essais randomisés dans le domaine des soins critiques se focalisent souvent sur la mortalité comme critère de jugement principal, [64] ce qui impose des populations de grande taille, au risque d'accepter une puissance statistique prévue plus faible. Au-delà, nous pouvons aussi discuter de l'intérêt de s'intéresser spécifiquement à la mortalité, qui est un critère objectif mais ne capturant pas à la fois les effets bénéfiques mais aussi ceux indésirables d'une stratégie (*e.g.* un traitement qui diminuerait la mortalité mais en augmentant la probabilité d'être en état végétatif).

D'autre part, de nombreuses pathologies en soins critiques sont étiquetées sous forme syndromique (*e.g.* un choc septique), entrainant que les patients ont des présentations cliniques très hétérogènes, souvent entremêlées. Ces situations rendent difficile l'identification de sous-groupe spécifique de patients, mais surtout peuvent diluer l'effet de l'intervention avec des sous-groupes de patients pouvant être répondeurs et d'autres non-répondeurs (ou voire même avec un effet délétère). Ces situations ont ammené des essais accumulant les critères d'inclusions et d'exclusions, avec des pourcentages de patients inclus vraiment faible par rapport à ceux admis dans le service. Avec des critères aussi stricts, de nombreux essais ne pouvaient pas représenter la population réellement éligible, et souffraient donc d'un problème de généralisation de leurs résultats.

Depuis plus d'une quinzaine d'années, c'est dans ce contexte que le recours aux données observationnelles en soins critiques émergent comme une alternative sérieuse.

2.1.2 Données observationnelles et soins critiques

Dans le domaine de la réanimation et plus largement des soins critiques, il existe une augmentation des cohortes ouvertes de patients, aboutissant à l'augmentation du recueil prospectif de données observationnelles. L'accélération de l'accessibilité à ces données ouvre une opportunité pour des analyses d'inférence causale. En effet, depuis plusieurs

années, différentes publications dans le cadre des soins critiques, font la pédagogie de l'inférence causale, et détaillent l'utilisation que les cliniciens peuvent avoir des méthodes statistiques adaptées à l'inférence causale sur des données observationnelles. [65, 66]

Á l'heure actuelle, avec le développement des algorithmes issus de l'apprentissage automatisé et l'explosion de la quantité des données de santé dans l'ère du *big data*, émerge une médecine du futur qui sera plus personnalisée. L'objectif est *in fine* d'attribuer le bon traitement, au bon patient, à la bonne dose et au bon moment. Avec cette accumulation de données observationnelles, les méthodes d'inférence causale auront un rôle déterminant dans les aides médicales à la décision qui seront proposées aux praticiens.

Les unités d'hospitalisation de soins critiques font partie des unités hospitalières productrices d'une quantité importante de données de santé, avec des patients dont la surveillance des paramètres vitaux (*e.g.* la pression artérielle, la fréquence cardiaque, la saturation percutanée en oxygène, etc.) est continue, mais aussi avec des réalisations itératives de bilan biologique ou d'examens paracliniques. Dans ce cadre, les soins critiques sont un exemple concret de service hospitalier à l'origine d'accumulation de données de santé.

Avec cette volonté ambitieuse de vouloir mettre au point des aides décisionnelles accessibles au lit du patient, des quantités importantes de données sont nécessaires pour accéder à une prédiction fiable. Les avancés récentes dans le domaine de l'informatique permettent la création, le stockage et l'accès à de grands jeux de données médicales avec une finesse dans la granularité des données [1] (*big data*).

2.1.3 Exemple du projet MIMIC

Par exemple, un projet datant de 2001 à Boston (Beth Israel Deaconess Medical Center) a entraîné la création d'une large base de données monocentrique regroupant de nombreuses variables de patients hospitalisés en soins critiques. Cette base de donnée, intitulée MIMIC pour *Medical Information Mart for Intensive Care*, regroupe différentes données : des valeurs issues des courbes de paramètres vitaux, des détails cliniques extraits des dossiers médicaux, et même les données de mortalité issu du registre national (Figure 5).

Ces données sont assemblées, protégées, cryptées dans un répertoire central, et ont été anonymisées pour permettre un accès public. Pour donner un ordre d'idée, MIMIC III est composée de 53 423 admissions hospitalières distinctes en soins critiques pour des patients adultes (plus de 16 ans) hospitalisés entre 2001 et 2012, avec 7870 admissions de nouveau-nés entre 2001 et 2008. [67]

La dernière mise à jour (MIMIC-IV) est accessible après validation de l'accès sur le site suivant : `https://mimic.mit.edu` Par cette accessibilité facilitée aux bases de données MIMIC, cette base de données constituent celle qui a été la plus étudiée et analysée à l'heure actuelle. L'accord d'utilisation des données de MIMIC engage les analystes au partage publique des codes des différents projets issus de MIMIC. Cette politique de partage du savoir a permis le développement rapide de concepts et l'accélération de la recherche.

1. La granularité désigne le niveau de détail. Plus la granularité est fine, plus la donnée est détaillée.

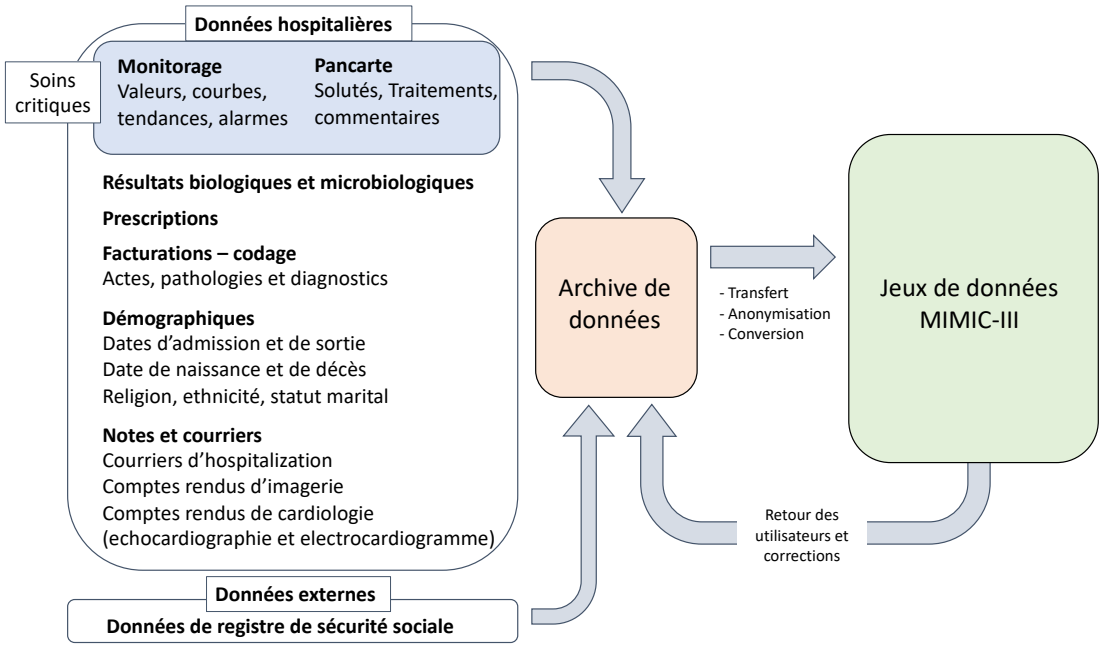


FIGURE 5 – Représentation du processus de gestion des données pour créer la base de données MIMIC (issue et traduite de Johnson *et al.*, 2016

2.1.4 Autres bases de données en soins critiques

Dans le monde Depuis le projet MIMIC, différents projets ont vu le jour dans le monde. Une autre initiative américaine, en partenariat avec Philips® et son institut de recherche, a permis la constitution d'une grande base de données collaborative : *eICU Collaborative Research Database* (eICU-CRD). [68] Cette base de données multicentrique est maintenant composée de plus de 200 000 admissions en soins critiques issues de différents établissements hospitaliers aux États-Unis (plus de 300 services de soins critiques). Cet aspect multicentrique dans la constitution de la base est un exemple dans l'inter-opérabilité des données, qui a été permis par l'archivage des données dans la plateforme de Philips® pour les soins intensifs. L'accès à la base eICU-CRD est aussi disponible sur le site suivant : `https://eicu-crd.mit.edu`. En parallèle, des projets spécifiques récents ont aussi permis la constitution de grandes cohortes observationnelles par l'aggrégagtion de plusieurs jeux de données déjà constitués, avec notamment l'initiative validant le score qSOFA dans le choc septique, [69] ou celle de la classification de Berlin dans le syndrome de détresse respiratoire aïgu. [70]

D'autres projets de cohorte collaborative ont vus le jour dans le monde. Nous pouvons citer le registre des sociétés de réanimation d'Autralie et de Nouvelle-Zélande. [71] Et en Europe, le registre national hollandais regroupe plus de 100 services de soins critiques. [72]. Si on recherche une base de données européenne avec des variables de granularité plus fine, il faut s'intéresser à la base de données *HIRID*, qui est une base accessible (`https://hirid.intensivecare.ai/`) et regroupant plus de 33 000 admissions de patients dans une unité de soins critiques dans l'hôpital universitaire de Bern (Suisse). La base de données est composée de plus de 680 variables , avec un recueil toutes les 2 minutes pour certains des paramètres physiologiques. Une base de données collaborative de données longitudinales s'est aussi constituée à partir de 5 unités de réanimation en Angleterre. [73]

Dans le contexte récent de la pandémie COVID-19, nous pouvons aussi prendre l'exemple

de la constitution de la cohorte COVID-ICU qui a été constituée par la collaboration de 138 hopitaux en France, Belgique et Suisse, et ayant permis la constitution d'une cohorte de 4643 patients entre le 25 février et le 4 mai 2020. Le recueil de données était fait manuellement avec le remplissage d'un cahier de recueil formalisé (CRF) électronique. [74]

En France

Si l'on s'intéresse aux bases de données de soins critiques en France, une des premières initiatives remonte à la constitution de la cohorte prospective FROG-ICU qui visait à évaluer le devenir de patients admis en réanimation, et ayant été ventilé ou sous un agent vasopresseur pendant plus de 24 heures. Cette cohorte constituée par des patients de plus de 24 unités françaises et une unité belge a permis la constitution d'une population de 2087 patients entre 2011 et 2013. [75]

Depuis, plusieurs cohortes prospectives ont vu le jour dans le domaine des soins critiques en France. Peu d'entre elles se constituent sous la forme de cohortes/registres ouverts et remplis en permanence au fil du temps. Nous pouvons cependant citer l'initiative de la TraumaBase (https://www.traumabase.eu/fr_FR) qui a pour mission de recueillir prospectivement les données épidémiologiques, pré et intrahospitalières de patients polytraumatisés graves. Ce registre collaboratif réunit plus d'une vingtaine de centres sur toute la France, et avec plus de 3000 patients déjà inclus. Une autre initiative est portée par le réseau AtlanRéa avec la constitution d'une cohorte au sein de l'inter-région ouest.

A noter, à l'heure actuelle et à notre connaissance, il n'existe pas de cohorte nationale collaborative française en soins critiques se constituant prospectivement par le recueil automatique de données obtenues dans le dossier médical électronique.

2.2 Présentation de la cohorte AtlanRéa

2.2.1 Présentation de la cohorte

Le réseau AtlanRéa est un réseau régional des réanimations du grand Ouest dont l'objet est la promotion de la recherche médicale et scientifique au sein de l'inter-région Ouest de la France (https://www.atlanrea.org). Actuellement, le réseau regroupe les réanimations chirurgicales ou polyvalentes des centres hospitaliers suivants (Figure 6) :

Dans cet objectif de promotion de la recherche médicale, le réseau a mis en place un recueil prospectif multicentrique des patients hospitalisés au sein des réanimations de ces différents centres. L'objectif est d'inclure et de colliger les données de patients âgés de plus de 15 ans, et nécessitant une admission en réanimation pour une prise en charge pour une lésion traumatique (enregistrement clinicaltrials.gov, NCT02426255) ou une atteinte cérébrale non-traumatique, "cérébrolésion non-traumatique" (enregistrement clinicaltrials.gov, NCT02714387). La commission nationale informatique et Liberté (CNIL) a autorisé la collection de ces données anonymisées (décision DR-2013-047).

La cohorte est ouverte depuis mars 2013, avec une ouverture progressive des inclusions dans les différents centres au cours du temps, et plus de 4000 patients sont inclus à ce jour.

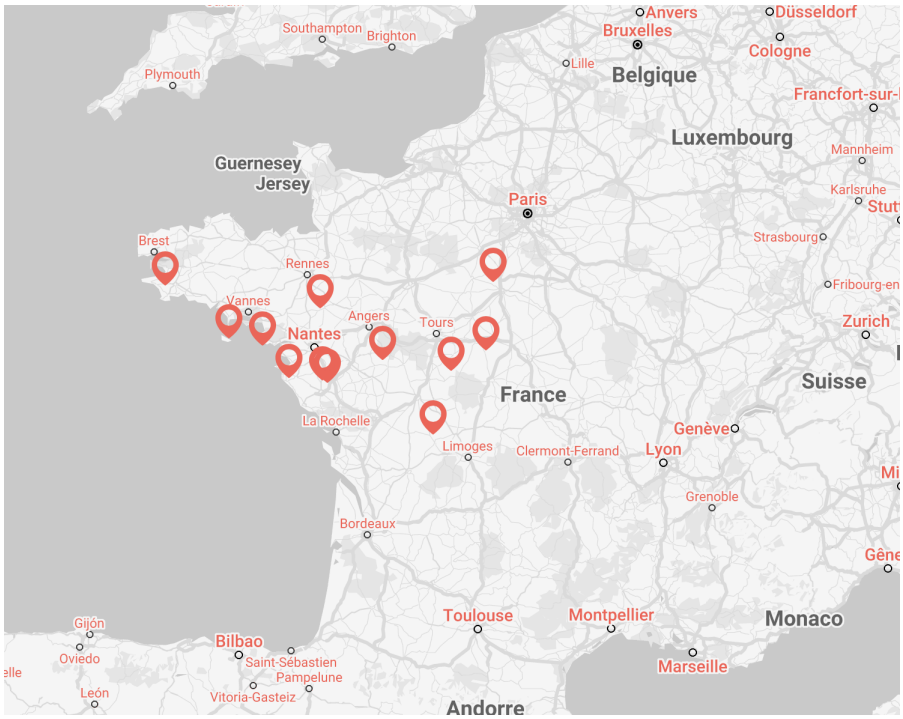


FIGURE 6 – Répartition géographique des centres du réseau AtlanRéa

2.2.2 Données collectées

Le recueil des données pour la cohorte AtlanRéa s'organise autour d'un tronc commun, puis des données spécifiques sont renseignées pour les deux sous-populations particulières : celle des patients polytraumatisés, et celle des patients avec une lésion cérébrale non-traumatique.

Le détail du CRF de la cohorte est présenté en Annexe B.

Variables pour le tronc commun

Concernant les éléments du tronc commun, différentes variables sont relevées :

— des données démographiques,

— des éléments biologiques à l'admission,

— des scores de gravité (score IGS II et score SOFA),

— des détails de prise en charge d'infection au cours du séjour,

— des données détaillant les défaillances d'organes ainsi que le devenir des patients.

Ainsi, différents critères de jugement peuvent théoriquement être évalués, comme par exemple : la durée cumulée de ventilation mécanique ou celle de ventilation mécanique non-invasive, la survenue d'une trachéotomie, les durées cumulées de cathétérisme artériel et veineux, la durée de séjour en réanimation et la mortalité en réanimation. Des renseignements de limitations thérapeutiques au cours du séjour sont aussi colligés.

Enfin, le suivi est assuré par une visite à 90 jours de l'admission par l'évaluation du devenir fonctionnel sur l'échelle Glasgow Outcome Scale (GOS).

Variables pour les patients polytraumatisés

Pour les patients polytraumatisés, en plus des données du tronc commun, d'autres paramètres sont enregistrés :

— des variables détaillant le contexte du traumatisme,

— le bilan lésionnel et la prise en charge avant l'admission en réanimation,

— des données sur la prise en charge chirurgicale,

— des caractéristiques de neuro-réanimation,

— des variables concernant le suivi sérologique et les valeurs de pression intra-crânienne,

— des paramètres précisant la gestion de la ventilation mécanique, ainsi que l'examen clinique à la première tentative d'extubation.

Enfin, des critères particuliers de devenir sont aussi enregistrés avec le détail du score de Glasgow à la sortie des soins critiques, ainsi que le statut de décès à 6 mois.

Variables pour les patients avec une lésion cérébrale non-traumatique

Pour les patients avec une lésion cérébrale non-traumatique, différentes données sont aussi détaillées :

— le contexte de la pathologie cérébrale et les détails de la prise en charge thérapeutique avant l'admission en soins critiques,

— des données sur la prise en charge chirurgicale,

— les paramètres de neuro-réanimation,

— le suivi sérologique et les valeurs de pression intra-crânienne,

— les complications spécifiques à l'hémorragie méningée,

— les détails sur la gestion de la ventilation mécanique, ainsi que l'examen clinique à la première tentative d'extubation.

De même, le score de Glasgow à la sortie de la réanimation ainsi que le statut à 6 mois sont colligés.

Méthode de collection

Les données sont renseignées dans un CRF électronique (Ennov Clinical®) permettant la mise en commun des données recueillies dans chaque centre. Actuellement, aucune méthode de recueil automatique n'a été mise en place pour collecter directement les informations dans les dossiers médicaux électroniques. Chaque donnée est renseignée par des personnels médicaux ou des personnels de recherche, à partir des informations disponibles dans le dossier médical. Les visites du devenir à distance (à 90 jours et à 6 mois) sont réalisées par appel téléphonique, et chaque centre est respondable du recueil de ces données.

2.2.3 Description de la cohorte (extraction de janvier 2021)

Les paramètres décrivant la cohorte sont issus de l'extraction de données datant du 15 janvier 2021. Á cette date, 3203 patients étaient inclus dans la cohorte, avec 1338 patients ayant une lésion cérébrale non-traumatique et 1865 patients polytraumatisés. Nous avons fait le choix de détailler les variables principales et le plus souvent rencontrées dans la littérature de soins critiques, en précisant les valeurs dans chaque sous-groupe (patients avec une cérébrolésion non-traumatique et patients polytraumatisés).

Concernant la répartition des inclusions par centre, elle est représentée dans le tableau 2.1. Nous remarquons que les centres hospitaliers d'Angers, de Nantes, de Poitiers et de Rennes représentent plus de 80% des patients inclus. Au delà des capacités de recrutement pour chaque service, les autres centres hospitaliers ont débuté les inclusions secondairement, expliquant les différences observées.

TABLEAU 2.1 – Répartition des inclusions dans la cohorte AtlanRéa, en détaillant la population de patients avec une lésion cérébale non-traumatique et ceux polytraumatisés (export de la base de janvier 2021).

	Population globale (n=3203)	Cérébrolésion non-traumatique (n=1338)	Polytraumatisme (n=1865)
	n (%)	n (%)	n (%)
ANGERS	683 (21.3)	259 (19.4)	424 (22.7)
BREST	276 (8.6)	121 (9.0)	155 (8.3)
NANTES	939 (29.3)	327 (24.4)	612 (32.8)
POITIERS	705 (22.0)	315 (23.5)	390 (20.9)
RENNES	595 (18.6)	314 (23.5)	390 (20.9)
SAINT-BRIEUC	2 (0.1)	1 (0.1)	1 (0.1)
SAINT-NAZAIRE	1 (0.0)	1 (0.0)	0 (0.0)
TOURS	2 (0.1)	0 (0.0)	2 (0.1)

Le tableau 2.2 présente les paramètres des patients à l'admission en réanimation.

La cohorte est composée de 66% d'hommes avec un âge moyen de 50 ans. Ces patients présentaient des caractéristiques typiques de réanimation avec 4% d'entre eux ayant eu un arrêt cardio-respiratoire avant l'admission, et 18% d'entre eux ayant eu besoin d'une transfusion d'au moins un produit sanguin labile avant l'admission. De même, les scores de gravité clinique reflètent la sévérité dans la présentation des patients avec un score IGS II moyen de 47, et un score SOFA moyen de 7.6.

Concernant les pathologies chroniques, la seule différence notable pouvant différencier la population d'AtlanRéa de la population générale est un taux de consommation chronique d'alcool proche de 20%.

Le sous-groupe des patients polytraumatisés est plus jeune (âge moyen de 45 ans versus 57 ans), avec une prédominance d'hommes (78% versus 50%). Les patients polytraumatisés avaient plus souvent des épisodes d'hypotension artérielle avant leur admission

(38% versus 19) avec un recours plus important à la transfusion (26% versus 6%). Alors que le score IGS II est sensiblement proche entre les deux sous-groupes, le score SOFA est plus sévère dans le groupe des patients polytraumatisés.

TABLEAU 2.2 – Descriptif des patients à l'admission en réanimation dans la cohorte AtlanRéa, en comparant les patients avec une lésion cérébale non-traumatique ("cérébrolésion") à ceux polytraumatisés (export de la base de janvier 2021)

	Population globale (n=3203)		Cérébrolésion (n=1338)		Polytraumatisme (n=1865)		*p-value*
	NA	n (%)	NA	n (%)	NA	n (%)	
Patient masculin	24	2101 (66.1)	9	666 (50.1)	15	1435 (77.5)	<0.001
Cardiopathie chronique	93	64 (2.1)	42	33 (2.5)	51	31 (1.7)	0.105
Néphropathie chronique	58	59 (1.9)	21	33 (2.5)	37	26 (1.4)	0.027
Pneumopathie chronique	93	113 (3.6)	36	62 (4.8)	57	51 (2.8)	0.004
Diabète	65	225 (7.2)	21	105 (8.0)	44	120 (6.6)	0.138
Alcoolisme chronique	283	564 (19.3)	93	197 (15.8)	190	367 (21.9)	<0.001
Néoplasie	59	128 (4.1)	20	85 (6.4)	39	43 (2.4)	<0.001
Tabagisme actif	517	802 (29.9)	161	344 (29.2)	356	458 (30.4)	0.528
Immunodépression	75	62 (2.0)	20	50 (3.8)	55	12 (0.7)	<0.001
Score de Glasgow ≥ 9	56	871 (27.7)	20	305 (23.1)	36	566 (30.9)	<0.001
PAS ≤ 90 mmHg ‡	140	913 (29.8)	71	240 (18.9)	69	673 (37.5)	<0.001
Arrêt cardio-respiratoire ‡	32	137 (4.3)	16	36 (2.7)	16	101 (5.5)	0.001
Transfusion de PSL ‡	43	558 (17.7)	21	75 (5.7)	22	483 (26.2)	<0.001
	NA	m ± e-t	NA	m ± e-t	NA	m ± e-t	
Age (années)	2	50.2 ± 18.7	1	57.1 ± 13.8	1	45.3 ± 20.1	<0.001
IMC (Kg/m^2)	444	25.4 ± 5.1	176	26.0 ± 5.5	268	25.0 ± 4.7	<0.001
Hémoglobinémie (g/dl)	57	11.5 ± 2.4	29	12.3 ± 2.1	28	10.9 ± 2.4	<0.001
Leucocytes (G/l)	79	15.8 ± 6.4	48	15.1 ± 5.8	31	16.4 ± 6.8	<0.001
TP (%)	231	75.1 ± 19.6	142	81.2 ± 17.8	89	71.0 ± 19.6	<0.001
Plaquettes (G/l)	79	193.8 ± 82.3	45	220.9 ± 81.0	34	174.7 ± 77.6	<0.001
Fibrinogenémie (g/l)	894	2.7 ± 1.3	556	3.4 ± 1.2	338	2.4 ± 1.1	<0.001
Lactatémie ($mmol/l$)	582	2.6 ± 2.2	333	2.0 ± 1.7	249	2.9 ± 2.3	<0.001
pH	101	7.3 ± 0.1	39	7.4 ± 0.1	62	7.3 ± 0.1	<0.001
PaO_2/FiO_2	172	313 ± 191	72	319 ± 186	100	309 ± 194	0.131
Créatininémie ($mmol/l$)	38	78.4 ± 46.0	17	69.5 ± 45.7	21	84.7 ± 45.3	<0.001
Protidémie (g/l)	99	57.0 ± 11.1	31	61.6 ± 9.2	68	53.6 ± 11.2	<0.001
Glycémie (g/l)	354	8.5 ± 3.0	148	8.6 ± 2.8	206	8.5 ± 3.1	0.120
Urémie ($mmol/l$)	172	5.6 ± 3.4	98	5.5 ± 3.5	74	5.7 ± 3.3	0.163
Calcémie ($mmol/l$)	444	2.0 ± 0.2	229	2.1 ± 0.2	215	2.0 ± 0.2	<0.001
Score IGS II	333	46.5 ± 14.0	178	47.5 ± 12.6	155	45.8 ± 14.9	0.001
Score SOFA	1171	7.6 ± 3.5	677	6.1 ± 3.3	484	7.9 ± 3.6	<0.001

Abréviations : FiO_2, fraction inspirée en dioxygène ; IGS, Indice de Gravité Simplifié ; IMC, Indice de Masse Corporelle ; PaO_2, Pression partielle artérielle en dioxygène ; PSL, Produit Sanguin Labile ; SOFA, *Sequential Organ Failure Assessment*

m, moyenne ; NA, Not Available (données manquantes) ; e-t, écart-type.

‡ Survenue de l'évènement avant l'admission en réanimation.

Plusieurs paramètres de devenir sont aussi résumés dans le tableau 2.3.

Nous pouvons relever que 25.7% des patients inclus dans la cohorte sont décédés au cours de leur séjour en soins critiques. Plus de 40% d'entre eux ont présenté au moins un épisode de pneumopathie infectieuse au cours de leur séjour.

Concernant le suivi à 90 jours de l'admission en réanimation, même si un certain nombre de perdus de vue (28% de données manquantes) limite le niveau d'interpréta-

tion de la variable, il est intéressant de relever que 38% des patients sont décédés, que 21% d'entre eux ont bien récupéré, et que les patients restants ont un handicap modéré, sévère, ou même un état végétatif.

TABLEAU 2.3 – Devenir des patients inclus dans la cohorte AtlanRéa, en comparant les patients avec une lésion cérébale non-traumatique ("cérébrolésion") à ceux polytraumatisés (export de la base de janvier 2021)

	Population globale (n=3203)		Cérébrolésion (n=1338)		Polytraumatisme (n=1865)		*p-value*
	NA	n (%)	NA	n (%)	NA	n (%)	
Pneumopathie infectieuse	120 ‡	1365 (44.3)	60 ‡	563 (44.0)	60 ‡	802 (44.4)	0.834
SDRA	166 ‡	581 (19.1)	53 ‡	237 (18.4)	113 ‡	344 (19.6)	0.409
Décès en réanimation	56	810 (25.7)	20	411 (31.2)	36	399 (21.8)	<0.001
GOS J90	909		336		573		<0.001
Bonne récupération		478 (20.8)		165 (16.5)		313 (24.2)	
Handicap modéré		377 (16.4)		135 (13.5)		242 (18.7)	
Handicap sévère		534 (23.3)		241 (24.1)		293 (22.7)	
État végétatif		31 (1.4)		13 (1.3)		18 (1.4)	
Décès		874 (38.1)		448 (44.7)		426 (33.0)	

Abréviations : GOS J90, *Glasgow Outcome Scale* mesuré à 90 jours de l'admission en réanimation ; SDRA, Syndrome de Détresse Respiratoire Aiguë.

‡ Données manquantes ou survenue de l'évènement avant l'admission en réanimation.

La suite du manuscrit présente les deux travaux d'analyse de cette thèse. Tout d'abord, le premier objectif de ce travail de thèse était d'étudier l'impact de l'utilisation des barbituriques dans une population de patients traumatisés craniens. Pour cela, nous avons utiliser les données observationnelles de la cohorte AtlanRéa, et nous avons appliqué la méthode de l'IPTW. Le chapitre 3 présente la contribution scientifique de cette analyse.

Le deuxième objectif de cette thèse était d'évaluer différentes méthodes statistiques d'inférence causale dans le cadre de situations de violation de l'hypothèse de positivité. La contribution de ce travail est présentée dans le chapitre 4.

Chapitre 3

Evaluation de l'impact de l'utilisation précoce des barbituriques chez les patients traumatisés graves

3.1 Manuscrit

PLOS ONE

RESEARCH ARTICLE

Early use of barbiturates is associated with increased mortality in traumatic brain injury patients from a propensity score-based analysis of a prospective cohort

Maxime Léger[1,2]*, Denis Frasca[2,3], Antoine Roquilly[4], Philippe Seguin[5], Raphaël Cinotti[4], Claire Dahyot-Fizelier[3], Karim Asehnoune[4], Florent Le Borgne[2,6], Thomas Gaillard[1], Yohann Foucher[2,7], Sigismond Lasocki[1], for AtlanRéa group¶

1 Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire d'Angers, Angers, France, **2** INSERM UMR 1246—SPHERE, Nantes University, Tours University, Nantes, France, **3** Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire de Poitiers, Poitiers, France, **4** Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire de Nantes, Nantes, France, **5** Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire de Rennes, Rennes, France, **6** IDBC-A2COM, Pacé, France, **7** Centre Hospitalier Universitaire de Nantes, Nantes, France

¶ The complete membership of the author group can be found in the Acknowledgments
* maxime.leger@chu-angers.fr

Check for updates

OPEN ACCESS

Citation: Léger M, Frasca D, Roquilly A, Seguin P, Cinotti R, Dahyot-Fizelier C, et al. (2022) Early use of barbiturates is associated with increased mortality in traumatic brain injury patients from a propensity score-based analysis of a prospective cohort. PLoS ONE 17(5): e0268013. https://doi.org/10.1371/journal.pone.0268013

Editor: Aurel Popa-Wagner, Essen University Medical School, GERMANY

Received: November 30, 2020

Accepted: April 20, 2022

Published: May 4, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0268013

Copyright: © 2022 Léger et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Abstract

Barbiturates are proposed as a second/third line treatment for intracranial hypertension in traumatic brain injury (TBI) patients, but the literature remains uncertain regarding their benefit/risk balance. We aimed to evaluate the impact of barbiturates therapy in TBI patients with early intracranial hypertension on the intensive care unit (ICU) survival, the occurrence of ventilator-associated pneumonia (VAP), and the patient's functional status at three months. We used the French AtlanREA prospective cohort of trauma patients. Using a propensity score-based methodology (inverse probability of treatment weighting), we compared patients having received barbiturates within the first 24 hours of admission (barbiturates group) and those who did not (control group). We used cause-specific Cox models for ICU survival and risk of VAP, and logistic regression for the 3-month Glasgow Outcome Scale (GOS) evaluation. Among the 1396 patients with severe trauma, 383 had intracranial hypertension on admission and were analyzed. Among them, 96 (25.1%) received barbiturates. The early use of barbiturates was significantly associated with increased ICU mortality (HR = 1.85, 95%CI 1.03–3.33). However, barbiturates treatment was not significantly associated with VAP (HR = 1.02, 95%CI 0.75–1.41) or 3-month GOS (OR = 1.67, 95%CI 0.84–3.33). Regarding the absence of relevant clinical trials, our results suggest that each early prescription of barbiturates requires a careful assessment of the benefit/risk ratio.

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Monitoring intracranial pressure (ICP) is recommended for the care of patients with severe traumatic brain injury (TBI) [1]. Indeed, even in brief periods, intracranial hypertension is associated with poorer outcomes [2–4]. Many therapeutic options are available in case of elevated ICP, including sedation, osmotherapy, maintenance of high cerebral perfusion pressures, external ventricular drainage, craniectomy, etc. A stepwise implementation of these treatments is usually proposed, and barbiturates are one of these options [1].

Indeed, barbiturates have been recommended to treat high and refractory ICP since the early 80s [5,6]. They are still suggested as a second or third line of treatment in US guidelines [7], or in the recent Seattle International Severe Traumatic Brain Injury Consensus Conference [1], when the increase in ICP is refractory to other medical or surgical alternatives in TBI patients. Nonetheless, the body of evidence for the use of barbiturates in this context is still of low quality. Since the first randomized clinical trial by Eisenberg et al. in 1988, evaluating high doses of barbiturates for intractable ICP elevation in patients with a Glasgow Coma Scale (GCS) of 4–8 [8], few data are available. The most recent Cochrane review, including only seven trials with a total of 341 patients, concluded that there was no evidence for the use of barbiturates therapy in patients with an acute severe head injury [9]. In this analysis, there was no significant effect of barbiturates on mortality or recovery at the end of follow-up.

This lack of proven efficacy is even more worrying in the light of the known adverse effects of barbiturates. Beyond the induced hypotension described since the first uses of this therapy [10], barbiturates could induce immunosuppression [11], promoting the occurrence of ventilator-associated pneumonia (VAP) [12,13]. However, nowadays, barbiturates appear to be still widely used in TBI patients [14].

In this context of widespread use and absence of a well-established benefit-risk balance, further data is needed. The purpose of this study was to evaluate in real-life settings the impact of early barbiturates infusion on death, neurological outcome, and occurrence of VAP in patients admitted to Intensive Care Units (ICU) with severe TBI and intracranial hypertension.

Materials and methods

Study population

We used the prospective multicentric AtlanREA cohort (www.atlanrea.org) of trauma patients (NCT02426255), hospitalized in ICU in the western region of France. Data confidentiality was ensured following the recommendations of the French commission for data protection (Commission Nationale Informatique et Liberté, CNIL decisions DR-2013-047). The study was approved by the Ethics Committee of the French society of anesthesia and intensive care (SFAR), which waived patients' consent for this study according to French law on Bioethics [15]. All patients or their relatives were informed and agreed to the data collection. For this study, all methods were performed in accordance with relevant guidelines and regulations.

All patients in the database aged 15 years or more and hospitalized between March 2013 and March 2018 in one of the four university hospitals (Angers, Nantes, Poitiers, and Rennes) were eligible if they had TBI (defined as a brain lesion on the initial CT-scan) requiring admission to ICU with orotracheal intubation and had intracranial hypertension within the first 24 hours of admission. Intracranial hypertension was defined as an ICP >20 mmHg [16], or the need for a second/third-line therapy for clinical intracranial hypertension (i.e., osmotherapy, infusion of barbiturates, external ventricular drainage, urgent neurosurgery) because of the presence of a focal sign on admission before ICP monitoring.

Patients who received barbiturates (at least a bolus of 250 mg) within the first 24-hours of admission were considered in the barbiturates group, while those who did not were in the control group. The barbiturates group thus only included early administration of barbiturates (within the first 24 hours), in order to build a relevant propensity score with the maximum of variables (we did not have all the data later during the ICU stay).

Available data

Patient characteristics recorded were age, gender, Body Mass Index (BMI), admission Simplified Acute Physiology Score II (SAPS II) [17], worst GCS score (detailing the GCS motor score) during the first 24 hours following ICU admission, and the Injury Severity Score (ISS) [18]. Episodes of unreactive anisocoria or mydriasis, hypoxia, low blood pressure (defined as a systolic blood pressure < 90 mmHg), or blood transfusion were also noted before admission (i.e., during transportation, in the emergency room or the operating room). We also collected comorbidities, including histories of heart failure, renal failure, chronic respiratory disease, diabetes, cancer, and chronic tobacco- or alcohol use. Biological parameters available at ICU admission were hemoglobin, leukocytes, prothrombin time, platelets, fibrinogen, arterial lactate, pH, bicarbonate, the partial pressure of oxygen (PaO2), the fraction of inspired oxygen (fiO2), serum creatinine, protein, glucose, urea, and calcium. Computed tomography severity was specified using Marshall's classification (from I to VI, Class I being indicative for less severe lesions) [19].

The use of other specific neurocritical care therapeutics was also collected: osmotherapy, surgical intracerebral hematoma evacuation, external ventricular drainage, lobectomy, or decompressive craniectomy.

Outcomes

The primary outcome was patient survival in ICU. We also studied the time-to-first VAP, defined according to a standardized definition [20], and the 3-month post-admission Glasgow Outcome Scale (GOS) [21], dichotomized into favorable outcome (good recovery or moderate disability) or unfavorable outcome (severe disability, vegetative state or death).

Statistical analysis

We compared the patient's characteristics between the two groups of interest using Chi-square tests for categorical variables and Student t-tests for continuous variables. For outcomes' comparisons, we weighted on the Propensity Scores (PS) to consider possible confounders. The PS was estimated by multivariable logistic regression. Splines on continuous covariates were used to ensure the log-linearity assumptions. Variables significantly associated with the outcome and treatment in univariate regressions were retained ($p<0.2$). In addition, we also studied propensity score models including less variables by using the IMPACT TBI score (composed of the following covariates: age, Glasgow motor score, pupillary reactivity, hypoxia and hypotension status, and Marshall's CT scan score) since this score is a good predictor of TBI severity [22]. We removed in these models the variables already included in the IMPACT TBI score (i.e., SAP ≤ 90 mmHg, and the diagnosis of extradural or subdural hematoma), to avoid collinearities. These sensitivity analyses were carried out for the ICU survival and the 3-month GOS. For all the analyses, we considered a center effect as a covariate in the PS. We applied stabilized weights estimating the average treatment effect in the entire population (ATE) [23,24]. We assessed the goodness-of-fit of the models by checking the positivity assumptions graphically and studying standardized differences.

For times-to-event, cause-specific Cox models were estimated by maximizing the partial weighted likelihood and using a robust estimator for the variance [25]. Hazard

proportionalities were graphically checked by plotting log-minus-log survival curves. The crude cumulative incidence curves were obtained by the Aalen-Johansen estimator to account for competing risks. For the three-month GOS, we used a logistic regression by maximizing the weighted likelihood and using a robust estimator for the variance. Influential values were detected by a Cook distance greater than one in absolute value.

Sensitivity analysis

As stated above, we had to focus our study on the early administration of barbiturates to be able to build a relevant propensity score. However, some patients received barbiturates later during their stay. We conducted a sensitivity analysis by comparing ICU survival among patients who received barbiturates infusion at any time during the stay to those who did not. We used the same statistical approach (i.e., a cause-specific cox model). Such an analysis did not consider the immortal time bias and should be interpreted with caution.

All the statistical analyses were performed using the Plug-Stat software (www.labcom-risca.com) based on the R software (R Core Team, 2017, version 3.4.0). All the candidate variables for calculating propensity scores are those presented in **Table 1** (**S1 Tables** in S1 File detail multivariable logistic models leading to the propensity scores, **S2 Table** in S1 File show the standardized differences, and **S1 Fig** in S1 File resume the propensity scores distributions).

Results

Cohort description

During the study period, 1396 trauma patients were included in the database. Among them 982 patients had severe TBI and 699 required the insertion of an intracranial pressure monitoring sensor. 383 TBI patients had an intracranial hypertension (i.e., ICP >20 mmHg [16] or the need for therapy of intracranial hypertension because of localizing signs on admission) within the first 24-hours of admission and were included in the analysis, as depicted in **Fig 1.** Ninety-six (25.1%) patients were treated with barbiturates within the first 24 hours of admission, resulting in a control group of 287 (74.9%) patients. As illustrated in **Table 1**, patients treated with barbiturates tended to be younger and had a higher proportion of unreactive mydriasis or anisocoria before admission, GCS scores, but the CT severity of the injuries did not differ between the two groups. The proportion of blood transfusion before admission was higher in the barbiturates group, and these patients received more osmotherapy and had higher ICP at admission. We identified additional differences in plasma fibrinogen, plasma bicarbonate, plasma glucose, plasma proteins, and plasma calcium levels between groups. The IMPACT TBI score was close between the two groups, with a mean value of 9.6 in the barbiturates group versus 8.5 in the control group. According to these values, the calculated probability of death at 6 months was 41.8% in the barbiturates group versus 34.6% in the control group (p = 0.5121).

Patient survival in ICU

During the follow-up, 117 (30.5%) patients have died while in the ICU, including 44 (45.8%) patients in the barbiturates group and 73 (25.4%) patients in the control group. The cumulative probabilities of death in the ICU are presented in **Fig 2**. The observed (non-adjusted) cause-specific HR of death in the ICU was 2.13 (95%CI from 1.45 to 3.13) for patients of the barbiturates group versus those of the control group. After weighting on propensity scores, the corresponding confounder-adjusted HR was 1.85 (95%CI from 1.03 to 3.33). When we considered the propensity score model which included the IMPACT TBI score, the confounder-adjusted HR was 1.85 (95%CI from 1.04 to 3.23).

Table 1. Description of the studied population at baseline.

	Overall (n = 383)			Control group (n = 287)			Barbiturates group (n = 96)			p-value
	NA	n	%	NA	n	%	NA	n	%	
Male	0	308	80.4	0	230	80.1	0	78	81.2	0.8124
History of diabetes	5	22	5.8	4	16	5.7	1	6	6.3	0.8115
Chronic alcoholism	31	64	18.2	24	53	20.2	7	11	12.4	0.0994
Active smoking status	61	94	29.2	49	65	27.3	12	29	34.5	0.2113
SAP $\leq$ 90 mmHg	7	105	27.9	6	71	25.3	1	34	35.8	0.0481
Blood transfusion	1	99	25.9	1	65	22.7	0	34	35.4	0.0141
Hypoxia	10	279	74.8	5	206	73.0	5	73	80.2	0.2183
Glasgow score < 8	5	104	26.2	1	74	25.5	4	30	28.3	0.6036
Glasgow motor score	20			12			8			0.3569
6		35	9.6		27	9.8		8	9.1	
5		63	17.4		51	18.5		12	13.6	
4		82	22.6		66	24.0		16	18.2	
3		36	9.9		27	9.8		9	10.2	
2		34	9.4		23	8.4		11	12.5	
1		113	31.1		81	29.5		32	36.4	
Unreactive mydriasis or anisocoria	5	140	36.6	5	96	33.4	0	44	45.8	0.0498
CT scan classification										
Marshall classification	0			0			0			0.2550
Diffuse injury I		10	2.6		10	3.5		0	0.0	
Diffuse injury II		89	23.2		72	25.1		17	17.7	
Diffuse injury III		33	8.6		23	8.0		10	10.4	
Diffuse injury IV		25	6.5		18	6.3		7	7.3	
Evacuated mass lesion V		151	39.5		108	37.6		43	44.8	
Non-evacuated mass VI		75	19.6		56	19.5		19	19.8	
Osmotherapy	2	255	66.9	1	177	61.9	1	78	82.1	0.0003
Evacuation of subdural or extradural hematoma	0	108	28.2	0	74	25.8	0	34	35.4	0.0694
External ventricular drain	0	28	7.3	0	22	7.7	0	6	6.2	0.6447
Evacuation of cerebral hematoma	0	17	4.4	0	10	3.5	0	7	7.3	0.1493
Decompressive craniectomy	0	74	19.3	0	51	17.8	0	23	24.0	0.1837
	NA	mean	SD	NA	mean	SD	NA	mean	SD	
Age (years)	0	40.5	18.7	0	41.7	19.2	0	36.9	16.8	0.0194
BMI (kg.m-2)	48	24.3	4.5	32	24.4	4.5	16	23.8	4.8	0.3545
Intracranial pressure on admission (mm Hg)	17	22.0	15.8	14	19.3	13.3	3	29.8	19.5	0.0001
Haemoglobin (g/dL)	3	11.0	2.5	1	11.1	2.4	2	10.8	2.6	0.3831
Leukocytes (count/mm^3)	4	17.9	7.1	2	17.5	7.1	2	18.9	7.0	0.0909
Prothrombin (%)	15	69.6	18.9	12	70.4	18.5	3	67.4	20.2	0.2120
Platelets (count/mm^3)	4	178.4	70.0	2	180.5	66.8	2	172.0	79.0	0.3489
Fibrinogen (g/L)	70	2.3	1.2	53	2.4	1.2	17	2.1	1.3	0.0582
Arterial Lactate (mmol/L)	52	2.8	2.3	38	2.7	2.3	14	3.1	2.2	0.1055
Arterial pH	4	7.3	0.1	2	7.3	0.1	2	7.3	0.1	0.0659
Bicarbonate (mmol/L)	4	20.9	4.1	2	21.2	4.0	2	20.0	4.4	0.0175
PaO2 (mm Hg)	8	136.6	79.2	5	134.7	75.9	3	142.6	88.5	0.4437
FiO2	9	0.5	0.2	4	0.5	0.2	5	0.5	0.2	0.7814
PaO2/FiO2 ratio	14	321.0	179.4	8	323.6	190.8	6	313.0	138.7	0.5693
Serum Creatinine (mmol/L)	4	78.5	31.4	3	76.6	31.0	1	84.4	32.1	0.0395
Serum proteins (g/L)	9	54.3	11.0	6	54.9	11.1	3	52.4	10.8	0.0585
Serum glucose (mmol/L)	31	8.7	3.4	17	8.5	2.9	14	9.6	4.7	0.0352

(*Continued*)

Table 1. (Continued)

	Overall (n = 383)			Control group (n = 287)			Barbiturates group (n = 96)			p-value
	NA	n	%	NA	n	%	NA	n	%	
Serum urea (mmol/L)	9	4.9	2.1	8	4.8	2.0	1	5.0	2.3	0.6403
Serum calcium (mmol/L)	42	2.0	0.2	32	2.0	0.2	10	1.9	0.2	0.1091
SAPS II score	19	46.0	12.4	13	45.3	11.9	6	47.9	13.6	0.1091
ISS score	2	27.1	14.7	1	27.1	14.5	1	27.1	15.2	0.9842
IMPACT TBI score	38	8.8	4.3	11	8.5	4.2	27	9.6	4.4	0.0654

NA, number of data Not Available; BMI, Body Mass Index; FIO2, Fraction of Inspired Oxygen; ISS score, Injury Severity Score; PaO2, arterial partial Pressure of Oxygen; SAP, Systolic Arterial Pressure; SAPS, Simplified Acute Physiology Score; SD, standard deviation. IMPACT TBI score includes: age, Glasgow motor score, pupillary reactivity, hypoxia and hypotension status, and Marshall's CT scan score and is predictive of mortality in TBI patients.

https://doi.org/10.1371/journal.pone.0268013.t001

Incidence of VAP

Seven patients treated for aspiration pneumonia on admission in the ICU were excluded. Among the remaining 376 patients, 207 (55.1%) developed a least one VAP during their ICU

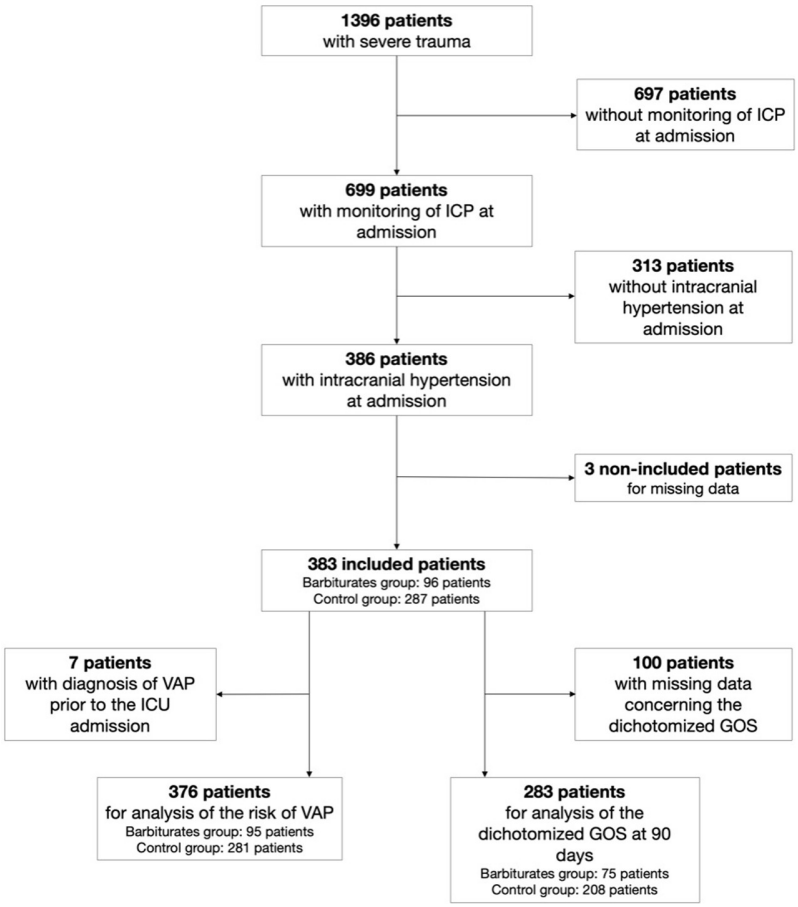


Fig 1. Flow chart of TBI patients of the AtlanREA cohort who had intracranial pressure monitoring and developed intracranial hypertension.

https://doi.org/10.1371/journal.pone.0268013.g001

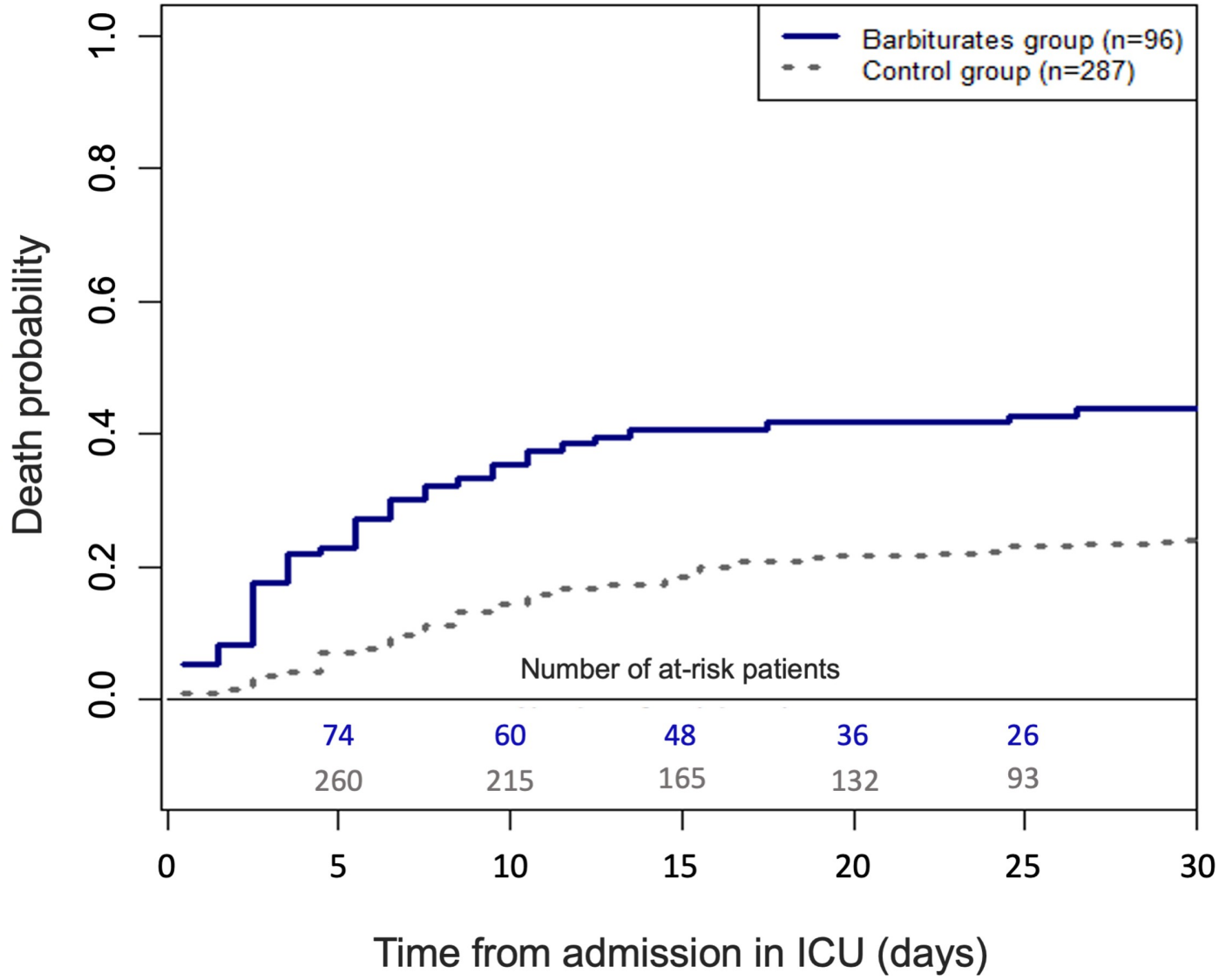


Fig 2. Cumulative incidences curves for death in intensive care unit (estimated by using the Aalen-Johansen estimator with discharge as a competing event, n = 383). The solid line represents the cumulative incidence curve for the barbiturates group (96 patients), while the dotted line corresponds to the control group (287 patients).

https://doi.org/10.1371/journal.pone.0268013.g002

stay. In the barbiturates group, 49 (51.6%) patients developed a VAP compared to 158 (56.2%) in the control group. Fig 3 shows the cumulative incidences of VAP. The observed cause-specific HR of VAP in ICU was 0.96 (95%CI from 0.69 to 1.32) for patients of the barbiturates versus the control group. The corresponding confounder-adjusted HR was 1.02 (95% CI from 0.75 to 1.41).

GOS score at three months

The 3-month GOS was missing for 100 patients, who are excluded from this analysis. Characteristics at the time of admission in ICU of the included and excluded patients are presented in **S3 Table** in S1 File (supporting information). One of the centers was associated with a higher proportion of missing data on the GOS score.

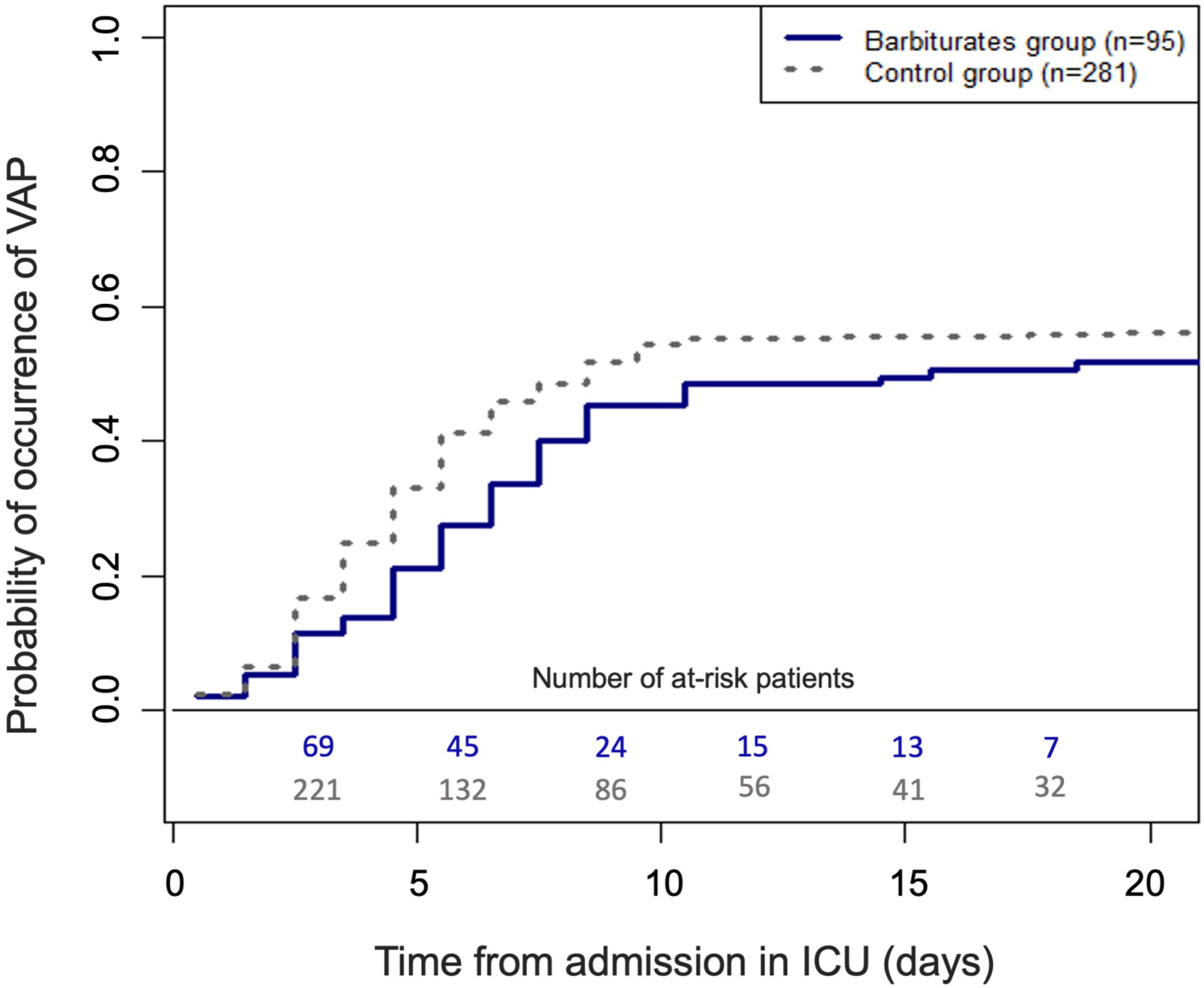


Fig 3. Cumulative incidences curves for ventilator-associated pneumonia (VAP) in intensive care unit (estimated by using the Aalen-Johansen estimator with discharge and death as competing events, n = 376). The solid line represents the cumulative incidence curve for the barbiturates group (95 patients), while the dotted line corresponds to the control group (291 patients).

https://doi.org/10.1371/journal.pone.0268013.g003

The distribution of the 3-month GOS is presented in Fig 4. The overall percentage of patients with unfavorable prognosis was 66.1% (n = 187). The observed percentage was 74.7% in the barbiturates group versus 63.0% in the control group (p = 0.066). The corresponding confounder-adjusted percentages were 74.5% (95% CI from 61.2% to 84.4%) in the barbiturates group versus 63.7% (95% CI from 56.6% to 70.3%) in the control group. The corresponding OR of progressing to an unfavorable outcome at 3-months was 1.67 (95% CI from 0.84 to 3.33) for patients treated with barbiturates versus the others.

When we considered the propensity score model which included the IMPACT TBI score, the corresponding confounder-adjusted OR was 1.96 (95% CI from 0.93 to 4.17) for the barbiturates group versus the control group.

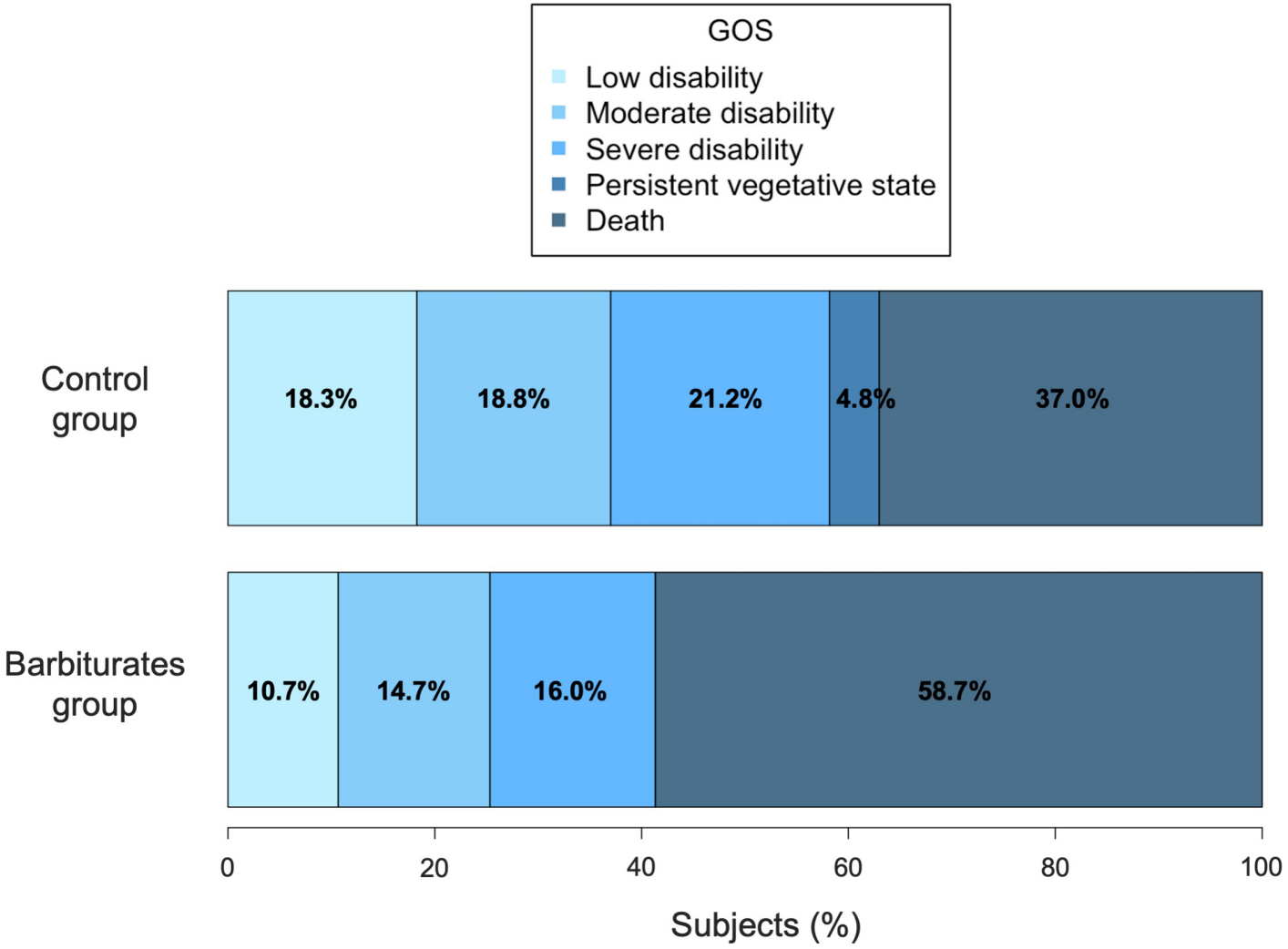


Fig 4. Distribution of the 3-month Glasgow Outcome Scale score according to barbiturates treatment (n = 283). The Glasgow outcome scale is represented from the light grey group (to the left of the bar plot) and corresponding to patients with little deficiency at three months, to the dark grey group (to the right of the bar plot) with a death status at three months.

https://doi.org/10.1371/journal.pone.0268013.g004

Sensitivity analysis

Among the 383 patients, 172 (44.9%) received barbiturates during their ICU stay at any time. During the follow-up, 117 patients died (76 (44%) among those who received barbiturates at any time and 41 (19%) among those who never received barbiturates). The cause-specific HR was 2.56 (95% CI from 1.72 to 3.70) for patients who received barbiturates at any time compared to those who did not. The cause-specific adjusted HR was 2.17 (95% CI from 1.35 to 3.45).

Discussion

In our cohort of 1396 trauma patients, more than a quarter had early intracranial hypertension and around a quarter of these patients received early barbiturates (within 24-hours of admission). These patients had a lower ICU survival without ~~a~~ difference in the incidence of VAP or ~~a~~ in three-month functional outcome (i.e., the GOS at 3 months).

The first lesson of our cohort is that barbiturates appear to be frequently used to control early intracranial hypertension in TBI patients, with almost a quarter of them treated with barbiturates within 24 hours of admission. This proportion is in agreement with the one reported by Majdan et al. [14]. The question of the risk-benefit balance of this therapeutic is thus relevant [1]. Unlike Majdan et al. [14], we choose to select only patients with intracranial hypertension, since only this subgroup of patients may beneficiate from barbiturates, as a third-line therapy [1]. We defined intracranial hypertension as an ICP $\geq$ 20 mmHg, which this is the threshold of the French guidelines for therapeutic management [16]. We could have chosen a higher cut-off (22 or 25 mmHg), as proposed in the recent US guidelines [7]. But in France, treatment of elevated ICP is usually started for ICP $\geq$ 20 mmHg. We have also defined intracranial hypertension as the need for therapeutic intervention (i.e., before the measure of ICP), as it is often the case in real-life practices. Combining these criteria ensures us an exhaustive identification of the most severe patients eligible for barbiturates therapy within 24 hours of the intracranial hypertension episode. Despite these selection criteria, we still obtain a relatively large population (n = 383) compared, for example, to the most recent Cochrane review, in which 341 patients were studied (including 105 patients treated with barbiturates) [9]. There are indeed few trials evaluating barbiturates in severe TBI patients, and none are recent.

We reported a significant association between early barbiturates use and ICU survival. This finding is not in agreement with the result of the Cochrane review [9], and while all the precautions of causal inference have been considered, these findings emerged from an analysis of observational data (contrary to the Cochrane review which includes randomized trials). However, this may not be surprising, since barbiturates may impair cerebral oxygenation and thus may impair outcomes [26,27]. We must also underline that we focus on the early use of barbiturates (within the first 24 hours) when barbiturates are recommended as a third-line of therapeutic [1,7]. And even if our results highlight an adverse effect of the early use of barbiturates on ICU mortality, we cannot examine the later use of this therapeutic alternative in later onset episodes of intracranial hypertension.

However, we have reported that the long-term functional outcome is not significantly different, even if the GOS score tends to be worse in the barbiturates group. This result is in agreement with those obtained by Majdan et al. [14], who did not find any significant effect of barbiturates on the six-month GOS score. Few authors have studied the long-term outcomes of barbiturates-treated TBI patients. In 1985, Ward et al. have reported no significant difference in the one-year GOS score, from a randomized controlled trial of prophylactic pentobarbital versus standard treatment in 53 TBI patients [28]. Marshall et al. have suggested that barbiturates, used as rescue therapy in a cohort of 55 patients, was associated to a favorable functional outcome with 68% of survivors at one year, but without a control group [29]. In a series of mixed cases of 49 patients admitted for head injuries and subarachnoid hemorrhages, who have been treated with barbiturates, the results remained inconclusive on the GOS score at one year [30]. All these data are relatively old, and many practices have improved since. We have studied the outcome of patients treated according to recent recommendations and practices. In our study, the non-significant effect on the three-month GOS score should be interpreted in light of the number of missing data.

Barbiturates have been accused of being immunosuppressive and of promoting VAP. Indeed, the administration of barbiturates may promote reversible bone marrow suppression, inhibit normal monocyte behavior, and disrupt the NF-κB activating cascade [11,31,32]. In 1995, Nadal et al. have described a significant association of barbiturates uses and VAP occurrence in 151 patients with a head injury [12]. Lepelletier et al. have reported an adjusted OR of 2.68 (95% CI: 1.06–6.80) for the occurrence of early-onset ventilator associated-pneumonia during barbiturates therapy, in a cohort of 161 patients with head trauma [13]. The same

observation has been reported by Hamele et al. in a pediatric population [33]. However, in all these studies, the statistical adjustment may be insufficient. Contrary to these observations, our data do not allow us to conclude for a significant association on the occurrence of VAP. We should acknowledge that we only evaluate the association between early barbiturates use and the first episode of VAP, which may be partly related to pulmonary aspiration.

Our study has other limitations. Firstly, although we have analyzed a relatively large number of patients compared to the available literature, the lack of significant effect, particularly for the GOS score, can be attributed to a lack of power. The number of missing data for the 3-month GOS lowered the power of this analysis. Secondly, even if the variables collected at baseline seem to be able to describe the clinical picture of each patient, other confounders could have been omitted. That is why we also use the IMPACT TBI score in our propensity score Models. In the context of the analysis of observational data, no standardized barbiturates prescription protocol was used and cannot be reported. Physicians within the AtlanREA network follow US guidelines, using barbiturates as a second or third line [1,7]. Thirdly, the control group includes 76(19.8%) patients who have received barbiturates later in the course of their ICU stay. This highlights even more that barbiturates are frequently used in patients with intracranial hypertension (with a total of 172(45%) treated patients). The development of time-dependent propensity scores would have allowed to compare treated and untreated patients with the same characteristics at any post-admission time [34,35]. For that purpose, we need a cohort in which all the possible confounders are collected regularly (at least daily). Since we did not have all the data at each time points during the ICU stay, we restricted our purpose to the early use of barbiturates. However, we performed a sensitivity, exploratory analysis comparing the patients who received barbiturates at any time to those who did not and found consistent results, showing no benefit of barbiturates administration. In addition, we were not able to evaluate the dose and duration of barbiturates infusion, which could interfere with the effect of treatment [14]. The usual practice of the centers is to use high-doses barbiturates (boluses of 250–500 mg followed by continuous infusions around 4–8 g/24h) and to adjust them according to the measured ICP. Finally, it is a challenge to draw conclusions when all the factors leading the clinician to prescribe barbiturates cannot be controlled. Only a randomized controlled study may help to control these limits and to clarify the benefit of barbiturates in severe TBI patients.

Conclusions

In our cohort of TBI patients with intracranial hypertension on admission, early use of barbiturates (within the first 24-hours) was associated with a lower ICU survival, but not to an increased incidence of VAP or poorer three-month functional outcome. In the absence of relevant clinical trials demonstrating the clinical benefit of barbiturates, each prescription requires a careful assessment of the benefit/risk ratio.

Supporting information

S1 File. Supplementary tables and figures.
(DOCX)

S1 Dataset. Anonymized dataset.
(CSV)

Acknowledgments

We thank the participating hospital centers of the AtlanREA Group: Angers, Nantes, Rennes, and Poitiers.

AtlanRéa group consortium

Soizic Gergaud,[1] Pierre Lemarié,[1]Thierry Bénard,[3] Rémy Bellier,[3] Dominique Demeure Dit-Latte,[4] Pierre-Joachim Mahe,[4] Yoann Launey,[5] Thomas Lebouvier,[5] Audrey Tawa,[5]

Author Contributions

Conceptualization: Maxime Léger, Denis Frasca, Florent Le Borgne, Sigismond Lasocki.

Data curation: Maxime Léger, Denis Frasca, Antoine Roquilly, Philippe Seguin, Raphaël Cinotti, Claire Dahyot-Fizelier, Karim Asehnoune, Thomas Gaillard.

Formal analysis: Maxime Léger, Sigismond Lasocki.

Investigation: Maxime Léger, Raphaël Cinotti.

Methodology: Maxime Léger, Denis Frasca, Florent Le Borgne, Yohann Foucher, Sigismond Lasocki.

Project administration: Yohann Foucher, Sigismond Lasocki.

Software: Florent Le Borgne.

Supervision: Yohann Foucher, Sigismond Lasocki.

Validation: Antoine Roquilly, Philippe Seguin, Raphaël Cinotti, Claire Dahyot-Fizelier, Karim Asehnoune, Thomas Gaillard, Yohann Foucher, Sigismond Lasocki.

Writing – original draft: Maxime Léger, Florent Le Borgne, Thomas Gaillard.

Writing – review & editing: Antoine Roquilly, Philippe Seguin, Raphaël Cinotti, Claire Dahyot-Fizelier, Karim Asehnoune, Yohann Foucher, Sigismond Lasocki.

References

1. Hawryluk GWJ, Aguilera S, Buki A, Bulger E, Citerio G, Cooper JD, et al. A management algorithm for patients with intracranial pressure monitoring: the Seattle International Severe Traumatic Brain Injury Consensus Conference (SIBICC). Intensive Care Med. 2019; 45:1783–1794. https://doi.org/10.1007/s00134-019-05805-9 PMID: 31659383
2. Badri S, Chen J, Barber J, Temkin NR, Dikmen SS, Chesnut RM, et al. Mortality and long-term functional outcome associated with intracranial pressure after traumatic brain injury. Intensive Care Med. 2012; 38(11):1800–9. https://doi.org/10.1007/s00134-012-2655-4 PMID: 23011528
3. Balestreri M, Czosnyka M, Hutchinson P, Steiner LA, Hiler M, Smielewski P, et al. Impact of Intracranial Pressure and Cerebral Perfusion Pressure on Severe Disability and Mortality After Head Injury. Neurocrit Care.2006; 4(1):008–13. https://doi.org/10.1385/NCC:4:1:008 PMID: 16498188
4. Karamanos E, Teixeira PG, Sivrikoz E, Varga S, Chouliaras K, Okoye O, et al. Intracranial pressure versus cerebral perfusion pressure as a marker of outcomes in severe head injury: a prospective evaluation. Am J Surg. 2014; 208(3):363–71. https://doi.org/10.1016/j.amjsurg.2013.10.026 PMID: 24524863
5. Shapiro HM, Wyte SR, Loeser J. Barbiturate-augmented hypothermia for reduction of persistent intracranial hypertension. J Neurosurg. 1974; 40(1):90–100. https://doi.org/10.3171/jns.1974.40.1.0090 PMID: 4808489
6. Bricolo AP, Glick RP. Barbiturate effects on acute experimental intracranial hypertension. J Neurosurg. 1981; 55(3):397–406. https://doi.org/10.3171/jns.1981.55.3.0397 PMID: 7264731
7. Carney N, Totten AM, O'Reilly C, Ullman JS, Hawryluk GWJ, Bell MJ, et al. Guidelines for the Management of Severe Traumatic Brain Injury, Fourth Edition. Neurosurgery. 2017; 80:6–15. https://doi.org/10.1227/NEU.0000000000001432 PMID: 27654000
8. Eisenberg HM, Frankowski RF, Contant CF, Marshall LF, Walker MD. High-dose barbiturate control of elevated intracranial pressure in patients with severe head injury. J Neurosurg. 1988; 69(1):15–23. https://doi.org/10.3171/jns.1988.69.1.0015 PMID: 3288723
9. Roberts I, Sydenham E. Barbiturates for acute traumatic brain injury. Cochrane Injuries Group, editor. Cochrane Database Syst Rev. 2012. https://doi.org/10.1002/14651858.CD000033.pub2 PMID: 23235573

10. Traeger SM, Henning RJ, Dobkin W, Giannotta S, Weil MH, Weiss M. Hemodynamic effects of pentobarbital therapy for intracranial hypertension. Crit Care Med. 1983; 11(9):697–701. https://doi.org/10.1097/00003246-198309000-00005 PMID: 6884049

11. Stover JF, Stocker R. Barbiturate coma may promote reversible bone marrow suppression in patients with severe isolated traumatic brain injury. Eur J Clin Pharmacol. 1998; 54(7):529–534. https://doi.org/10.1007/s002280050508 PMID: 9832294

12. Nadal P, Nicolás JM, Font C, Vilella A, Nogué S. Pneumonia in ventilated head trauma patients: the role of thiopental therapy. Eur J Emerg Med. 1995; 2(1):14–6. https://doi.org/10.1097/00063110-199503000-00004 PMID: 9422174

13. Lepelletier D, Roquilly A, Demeure dit latte D, Mahe PJ, Loutrel O, Champin P, et al. Retrospective Analysis of the Risk Factors and Pathogens Associated With Early-onset Ventilator-associated Pneumonia in Surgical-ICU Head-trauma Patients. J Neurosurg Anesthesiol.2010; 22(1):32–713. https://doi.org/10.1097/ANA.0b013e3181bdf52f PMID: 20027012

14. Majdan M, Mauritz W, Wilbacher I, Brazinova A, Rusnak M, Leitgeb J. Barbiturates Use and Its Effects in Patients with Severe Traumatic Brain Injury in Five European Countries. J Neurotrauma. 2013; 30 (1):23–9. https://doi.org/10.1089/neu.2012.2554 PMID: 22950895

15. Toulouse E, Masseguin C, Lafont B, McGurk G, Harbonn A, Roberts J, et al. French legal approach to clinical research. Anaesth Crit Care Pain Med. 2018; 37(6):607–14. https://doi.org/10.1016/j.accpm.2018.10.013 PMID: 30580775

16. Geeraerts T, Velly L, Abdennour L, Asehnoune K, Audibert G, Bouzat P, et al. Management of severe traumatic brain injury (first 24 hours). Anaesth Crit Care Pain Med. 2018; 37(2):171–86. https://doi.org/10.1016/j.accpm.2017.12.001 PMID: 29288841

17. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. JAMA. 1993; 270(24):2957–63. https://doi.org/10.1001/jama.270.24.2957 PMID: 8254858

18. Baker SP, O'Neill B, Haddon W, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. J Trauma. 1974; 14(3):187–96. PMID: 4814394

19. Marshall LF, Marshall SB, Klauber MR, Van Berkum Clark M, Eisenberg H, Jane JA, et al. The diagnosis of head injury requires a classification based on computed axial tomography. J Neurotrauma. 1992; 9 Suppl 1:S287–292.

20. Roquilly A, Feuillet F, Seguin P, Lasocki S, Cinotti R, Launey Y, et al. Empiric antimicrobial therapy for ventilator-associated pneumonia after brain injury. Eur Respir J. 2016; 47(4):1219–28. https://doi.org/10.1183/13993003.01314-2015 PMID: 26743488

21. Jennett B, Bond M. Assessment of outcome after severe brain damage. Lancet Lond Engl. 1975;1; 1 (7905):480–4. https://doi.org/10.1016/s0140-6736(75)92830-5 PMID: 46957

22. Steyerberg EW, et al. Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. PLoS Med. 2008;5; 5(8):e165. https://doi.org/10.1371/journal.pmed.0050165 PMID: 18684008

23. Robins JM, Hernan MA, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. Epidemiology. 2000; 11(5):550–60. https://doi.org/10.1097/00001648-200009000-00011 PMID: 10955408

24. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat. Med. 2014; 33:1242–1258. https://doi.org/10.1002/sim.5984 PMID: 24122911

25. Lin DY, Wei LJ. The Robust Inference for the Cox Proportional Hazards Model. J Am Stat Assoc. 1989; 84(408):1074–8.

26. Cruz J. Adverse effects of pentobarbital on cerebral venous oxygenation of comatose patients with acute traumatic brain swelling: relationship to outcome. J Neurosurg. 1996; 85(5):758–61. https://doi.org/10.3171/jns.1996.85.5.0758 PMID: 8893711

27. Chen HI, Malhotra NR, Oddo M, Heuer GG, Levine JM, Leroux PD. Barbiturate Infusion for Intractable Intracranial Hypertension and Its Effect on Brain Oxygenation. Neurosurgery. 2008; 63:880–887. https://doi.org/10.1227/01.NEU.0000327882.10629.06 PMID: 19005378

28. Ward JD, Becker DP, Miller JD, Choi SC, Marmarou A, Wood C, et al. Failure of prophylactic barbiturate coma in the treatment of severe head injury. J Neurosurg. 1985; 62(3):383–8. https://doi.org/10.3171/jns.1985.62.3.0383 PMID: 3882899

29. Marshall GT, James RF, Landman MP, O'Neill PJ, Cotton BA, Hansen EN, et al. Pentobarbital Coma For Refractory Intra-Cranial Hypertension After Severe Traumatic Brain Injury: Mortality Predictions and One-Year Outcomes in 55 Patients. J Trauma Inj Infect Crit Care. 2010; 69(2):275–83. https://doi.org/10.1097/TA.0b013e3181de74c7 PMID: 20699736

30. Dereeper E, Berré J, Vandesteene A, Lefranc F, Vincent J-L. Barbiturate coma for intracranial hypertension: clinical observations. J Crit Care. 2002; 17(1):58–62.28. https://doi.org/10.1053/jcrc.2002.33032 PMID: 12040550

31. Wheeler DW, Thompson AJ, Corletto F, Reckless J, Loke JCT, Lapaque N, et al. Anaesthetic Impairment of Immune Function Is Mediated via GABAA Receptors. PLoS ONE. 2011;24; 6(2):e17152. https://doi.org/10.1371/journal.pone.0017152 PMID: 21390329

32. Loop T, Humar M, Pischke S, Hoetzel A, Schmidt R, Pahl HL, et al. Thiopental Inhibits Tumor Necrosis Factor α–induced Activation of Nuclear Factor κB through Suppression of IκB Kinase Activity. Anesthesiology. 2003; 99(2):360–7. https://doi.org/10.1097/00000542-200308000-00017 PMID: 12883408

33. Hamele M, Stockmann C, Cirulis M, Riva-Cambrin J, Metzger R, Bennett TD, et al. Ventilator-Associated Pneumonia in Pediatric Traumatic Brain Injury. J Neurotrauma. 2016; 33(9):832–9. https://doi.org/10.1089/neu.2015.4004 PMID: 26203702

34. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology. 2000; 11: 561–570. https://doi.org/10.1097/00001648-200009000-00012 PMID: 10955409

35. Lu B. Propensity score matching with time-dependent covariates. Biometrics. 2005; 721–8. https://doi.org/10.1111/j.1541-0420.2005.00356.x PMID: 16135023

3.2 Éléments supplémentaires

Early use of barbiturates is associated with increased mortality in traumatic brain injury patients from a propensity score-based analysis of a prospective cohort.

Maxime Léger, MD [1,2*], mxmleger@gmail.com
Denis Frasca, MD,PhD [2,3], denis.frasca@gmail.com
Antoine Roquilly, MD,PhD [4], antoine.roquilly@chu-nantes.fr
Philippe Seguin, MD,PhD [5], philippe.seguin@chu-rennes.fr
Raphaël Cinotti, MD, PhD [4], raphael.cinotti@chu-nantes.fr
Claire Dahyot-Fizelier, MD, PhD [3], claire.dahyot-fizelier@chu-poitiers.fr
Karim Asehnoune, MD, PhD [4], Karim.ASEHNOUNE@chu-nantes.fr
Florent Le Borgne, PhD [2,6], fleborgne@idbc.fr
Thomas Gaillard, MD [1], ThGaillard@chu-angers.fr
Yohann Foucher, PhD [2,7], yohann.foucher@univ-nantes.fr
Sigismond Lasocki, MD, PhD [1], sigismond@lasocki.com
for AtlanREA group, atlanreagroup@gmail.com

[1] Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire d'Angers, Angers, France.

[2] INSERM UMR 1246 - SPHERE, Nantes University, Tours University, Nantes, France.

[3] Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire de Poitiers, Poitiers, France.

[4] Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire de Nantes, Nantes, France.

[5] Département d'Anesthésie Réanimation, Centre Hospitalier Universitaire de Rennes, Rennes, France.

[6] IDBC-A2COM, Pacé, France.

[7] Centre Hospitalier Universitaire de Nantes, Nantes, France.

Corresponding author : Dr Maxime Léger – Département d'Anesthésie Réanimation, CHU d'Angers, 4 rue Larrey 49100 Angers, France – Tel: +33 241353635 – Fax: +33 241353967 – mxmleger@gmail.com

Twitter: @mxmleger

Supplemental Digital Content

Table S1. Results of the multivariable logistic models leading to the propensity score.

To take into account possible confounding variables, we used the method of weighting according to inverse probability (Inverse Probability Weighting - IPW) of the Propensity Score (PS). The PS was estimated by a multivariable logistic regression with splines on continuous covariates to ensure the log-linearity assumption. Stabilized weights have been used in order to obtain a pseudo dataset with similar sample size than the original one and to estimate the average effect of the exposure (Average Treatment effect in the Entire population - ATE). To provide a relative measure of the effect, a logistic regression was estimated maximizing the weighted likelihood and using a robust estimator for the variance. The logistic regression leading to the propensity score is presented on table S2. Influential values were detected by a Cook distance greater than 1 in absolute value. Model was adjusted on center (effect was included in the propensity score).
For each outcome, the variables integrated into the model were those significantly associated with both the treatment (initiation of barbiturates) and the analyzed outcome. The significance threshold was defined in univariate analysis with a p-value < 0.2.

Table S1.a. Results of the multivariable logistic models leading to the propensity score regarding the survival analysis in ICU (n= 300).

	OR	95% CI	p-value
Osmotherapy	0.43	[0.22 ; 0.84]	0.0128
SAP ≤ 90 mmHg before admission	0.87	[0.45 ; 1.67]	0.6764
Evacuation of subdural or extradural hematoma	0.64	[0.34 ; 1.20]	0.1621
Blood transfusion before admission	0.42	[0.21 ; 0.86]	0.0166
Arterial pH	1.09	[0.59 ; 2.02]	0.7872
Serum proteins (g/L)	1.07	[0.53 ; 2.15]	0.8572
Serum glucose (mmol/l)	1.20	[0.67 ; 2.17]	0.5384
Serum calcium (mmol/l)	0.68	[0.35 ; 1.31]	0.2518

CI, Confidence Interval; OR, Odds Ratio; SAP, Systolic Arterial Pressure

Table S1.b. Results of the second multivariable logistic models leading to the propensity score regarding the survival analysis in ICU (n= 279).

	OR	95% CI	p-value
Osmotherapy	0.46	[0.23 ; 0.93]	0.0310
Blood transfusion before admission	0.49	[0.23 ; 1.05]	0.0680
Serum glucose ≥ 8 mmol/l	1.09	[0.57 ; 2.08]	0.7850
IMPACT TBI score ≥ 9	0.84	[0.44 ; 1.58]	0.5860
Arterial pH			0.8631
spline - 1	2.86	[0.16 ; 50.10]	
spline - 2	1.07	[0.13 ; 9.16]	
spline - 3	3.03	[0.01 ; 1471.37]	
spline - 4	1.62	[0.10 ; 25.41]	
Serum proteins (g/L)			0.9461
spline - 1	1.54	[0.14 ; 17.13]	
spline - 2	1.74	[0.12 ; 25.66]	
spline - 3	15.47	[0.01 ; 26536.32]	
spline - 4	12.67	[0.00 ; 1148159.12]	
Serum calcium (mmol/L)			0.2573
spline - 1	0.00	[0.00 ; 0.85]	

	OR	95% CI	p-value
spline - 2	0.00	[0.00 ; 2.14]	
spline - 3	0.00	[0.00 ; 0.74]	
spline - 4	0.03	[0.00 ; 2682.15]	

CI, confidence interval; OR, odds ratio.

Table S1.c. Results of the multivariable logistic model leading to the propensity score regarding the analysis of the occurrence of ventilator-associated pneumonia (n= 372).

	OR	95% CI	p-value
Decompressive craniectomy	0.74	[0.40 ; 1.36]	0.3292
Age (years)			0.1658
spline - 1	0.34	[0.02 ; 5.32]	
spline - 2	2.39	[0.19 ; 29.71]	
spline - 3	2.12	[0.15 ; 29.25]	
Serum Creatinine (mmol/L)	1.39	[0.84 ; 2.31]	0.2009

CI, Confidence Interval; OR, Odds Ratio.

Table S1.d. Results of the multivariable logistic model leading to the propensity score regarding the analysis of the occurrence of dichotomized GOS (n= 251).

	OR	95% CI	p-value
Osmotherapy	0.35	[0.17 ; 0.71]	0.0038
Blood transfusion before admission	0.57	[0.28 ; 1.17]	0.1269
Platelets (count/mm^3)			0.7933
spline - 1	0.37	[0.00 ; 72.16]	
spline - 2	3.61	[0.20 ; 64.14]	
spline - 3	0.26	[0.01 ; 10.88]	
Arterial pH			0.7095
spline - 1	1.28	[0.00 ; 649.00]	
spline - 2	2.10	[0.07 ; 58.96]	
spline - 3	0.49	[0.01 ; 20.87]	
Serum calcium (mmol/L)	0.64	[0.34 ; 1.22]	0.1780

CI, confidence interval; OR, odds ratio.

Table S1.e. Results of the second multivariable logistic models leading to the propensity score regarding the analysis of the occurrence of dichotomized GOS (n= 230).

	OR	95% CI	p-value
Osmotherapy	0.34	[0.16 ; 0.71]	0.0042
Blood transfusion before admission	0.45	[0.21 ; 0.98]	0.0440
Arterial pH ≥7.3	0.58	[0.29 ; 1.16]	0.1255
Serum calcium ≥1.9 mmol/L	2.17	[1.08 ; 4.37]	0.0295
IMPACT TBI score ≥10	0.97	[0.50 ; 1.86]	0.9189
Platelets (count/mm^3)			0.3807
spline - 1	0.13	[0.00 ; 4.51]	
spline - 2	0.17	[0.01 ; 2.40]	
spline - 3	0.00	[0.00 ; 3.58]	
spline - 4	0.24	[0.01 ; 4.29]	

CI, confidence interval; OR, odds ratio.

Table S1.f. Results of the multivariable logistic models leading to the propensity score regarding the survival analysis in ICU for the sensitivity analysis: all patients who were treated with barbiturates at any time during their stay versus the control group (n= 279).

	OR	95% CI	p-value
Osmotherapy	0.53	[0.30 ; 0.91]	0.0220
SAP ≤ 90 mmHg before admission	0.78	[0.43 ; 1.42]	0.4199
Blood transfusion before admission	0.87	[0.46 ; 1.62]	0.6517
Plasma bicarbonate ≥21 mmol/L	1.01	[0.60 ; 1.71]	0.9620
Fibrinogen (g/L)			0.2506
spline - 1	1.20	[0.25 ; 5.75]	
spline - 2	6.15	[0.96 ; 39.56]	
spline - 3	1.39	[0.03 ; 58.77]	
spline - 4	0.49	[0.01 ; 16.78]	
Serum calcium (mmol/L)			0.3839
spline - 1	0.28	[0.01 ; 6.23]	
spline - 2	0.85	[0.07 ; 10.66]	
spline - 3	0.55	[0.00 ; 909.91]	
spline - 4	22.83	[0.01 ; 43742.44]	
ISS score			0.9142
spline - 1	0.77	[0.27 ; 2.23]	
spline - 2	1.29	[0.32 ; 5.17]	
spline - 3	0.86	[0.08 ; 8.87]	
spline - 4	0.36	[0.04 ; 3.12]	

CI, confidence interval; ICU, OR, odds ratio, SAP, Systolic Arterial Pressure

Table S2. The PS-adjusted samples.

Table S2.a. The PS-adjusted sample for the survival analysis (n= 302, weighted sample).

	Whole sample		Barbiturates group		Control group		Standardized differences (%)
	n	%	n	%	n	%	
Osmotherapy	197.6	65.3	46.5	62.2	151.1	66.3	8.6
SAP ≤ 90 mmHg before admission	79.1	26.2	17.3	23.1	61.9	27.2	9.4
Evacuation of subdural or extradural hematoma	77.0	25.4	16.6	22.2	60.4	26.5	10.0
Blood transfusion before admission	85.7	28.3	20.3	27.1	65.4	28.7	3.6
	m	**SD**	**m**	**SD**	**m**	**SD**	
Arterial pH	7.3	0.1	7.3	0.1	7.3	0.1	11.6
Serum proteins (g/L)	53.7	10.9	52.8	9.3	54.0	11.4	11.1
Serum glucose (mmol/L)	8.6	3.0	8.9	3.5	8.4	2.8	15.4
Serum calcium (mmol/L)	2.0	0.2	2.0	0.2	2.0	0.2	5.0

m, weighted mean; SAP, Systolic Arterial Pressure; sd, weighted standard deviation;

Table S2.b. The PS-adjusted sample for the survival analysis with the second model (TBI IMPACT score) (n= 278, weighted sample).

	Whole sample		Barbiturates group		Control group		Standardized differences (%)
	n	%	n	%	n	%	
Osmotherapy	184.3	66.4	43.5	69.7	140.8	65.4	9.2
Blood transfusion before admission	77.8	28.0	16.1	25.9	61.7	28.7	6.3
Serum glucose ≥8 mmol/l	156.1	56.2	37.4	59.9	118.7	55.1	9.7
IMPACT TBI score ≥9	137.2	49.4	28.8	46.2	108.4	50.4	8.4
	m	**SD**	**m**	**SD**	**m**	**SD**	
Arterial pH	7.3	0.1	7.3	0.1	7.3	0.1	11.8
Serum proteins (g/L)	53.7	10.9	53.2	10.0	53.8	11.2	6.4
Serum calcium (mmol/L)	2.0	0.2	2.0	0.2	2.0	0.2	5.4

m, mean; SD, standard deviation.

Table S2.c. The PS-adjusted sample for analysis of VAP occurrences (n= 374; weighted sample).

	Whole sample		Barbiturates group		Control group		Standardized differences (%)
	n	%	n	%	n	%	
Decompressive craniectomy	70.3	18.8	16.9	17.5	53.4	19.3	4.7
	m	**SD**	**m**	**SD**	**m**	**SD**	
Age (years)	40.6	18.9	40.8	19.2	40.5	18.8	1.6
Serum Creatinine (mmol/L)	77.7	30.6	81.1	32.6	76.6	29.8	14.7

m, weighted mean; SD, standard deviation.

Table S2.d. The PS-adjusted sample for analysis of dichotomized Glasgow Outcome Scale score at 3 months (n= 250; weighted sample).

	Whole sample		Barbiturates group		Control group		Standardized differences (%)
	n	%	n	%	n	%	
Osmotherapy	169.0	67.7	43.6	70.4	125.3	66.8	7.7
Blood Transfusion before admission	77.7	31.1	20.3	32.8	57.3	30.6	4.8
	m	**SD**	**m**	**SD**	**m**	**SD**	
Platelets (count/mm^3)	175.3	72.3	172.2	73.4	176.3	72.1	5.6
Arterial pH	7.3	0.1	7.3	0.1	7.3	0.1	3.1
Serum calcium (mmol/L)	1.9	0.2	1.9	0.2	2.0	0.2	14.5

m, weighted mean; SD, standard deviation.

Table S2.e. The PS-adjusted sample for analysis of dichotomized Glasgow Outcome Scale score at 3 months with the second model (TBI IMPACT score) (n= 230; weighted sample).

	Whole sample		Barbiturates group		Control group		Standardized differences (%)
	n	%	n	%	n	%	
Osmotherapy	151.9	66.0	38.2	65.4	113.6	66.3	1.9
Blood transfusion before admission	70.9	30.8	18.3	31.3	52.6	30.7	1.4
Arterial pH ≥7.3	113.7	49.4	27.3	46.6	86.4	50.4	7.4
Serum calcium ≥1.9 mmol/L	136.3	59.3	33.5	57.2	102.8	60.0	5.6
IMPACT TBI score ≥10	108.0	46.9	28.9	49.5	79.0	46.1	6.8
	m	**SD**	**m**	**SD**	**m**	**SD**	
Platelets (count/mm^3)	176.1	70.4	173.3	69.0	177.0	71.2	5.3

m, mean; SD, standard deviation.

Table S2.f. The PS-adjusted sample for the survival analysis of the sensitivity analysis: all patients who were treated with barbiturates at any time during their stay versus the control group (n= 279; weighted sample).

	Whole sample		Barbiturates group		Control group		Standardized differences (%)
	n	%	n	%	n	%	
Osmotherapy	187.6	67.4	82.1	67.5	105.5	67.3	0.4
SAP ≤ 90 mmHg before admission	78.3	28.1	33.6	27.6	44.7	28.5	1.9
Blood transfusion before admission	88.7	31.9	39.0	32.0	49.8	31.7	0.7
Plasma bicarbonate ≥21 mmol/L	139.0	49.9	60.6	49.8	78.4	50.0	0.3
	m	**SD**	**m**	**SD**	**m**	**SD**	
Fibrinogen (g/L)	2.2	1.1	2.2	1.1	2.2	1.1	1.7
Serum calcium (mmol/L)	2.0	0.2	2.0	0.2	2.0	0.2	0.7

	Whole sample		Barbiturates group		Control group		Standardized differences (%)
	n	%	n	%	n	%	
ISS score	29.4	14.7	29.5	14.8	29.3	14.7	1.8

m, mean; SAP, Systolic Arterial Pressure ; SD, standard deviation.

Table S3. Description of the analyzed patients for the dichotomized Glasgow outcome score at 90 days versus excluded patients for missing data.

	Overall (n=383)			Analyzed patients for dichotomized Glasgow outcome score at 90 days (n=283)			Excluded patients (n=100)			p-value
	NA	n	%	NA	n	%	NA	n	%	
Center	0			0			0			0.0012
A		87	22.7		69	24.4		18	18.0	
B		171	44.6		135	47.7		36	36.0	
C		69	18.0		49	17.3		20	20.0	
D		56	14.7		30	10.6		26	26.0	
Male patient	0	308	80.4	0	225	79.5	0	83	83.0	0.4490
History of diabetes	5	22	5.8	2	16	5.7	3	6	6.2	0.0649
Chronic alcoholism	31	64	18.2	45	68	14.5	11	30	21.1	0.0608
Active smoking status	61	94	29.2	49	66	28.2	12	28	31.8	0.5251
SAP ≤ 90 mmHg before admission	7	105	27.9	5	82	29.5	2	23	23.5	0.2528
Hypoxia	10	279	74.8	6	211	76.2	4	68	70.8	0.3670
Blood transfusion before admission	1	99	25.9	0	80	28.3	1	19	19.2	0.0761
Glasgow score < 8	5	104	27.2	3	70	24.7	2	34	34.0	0.0909
Glasgow motor score	20			11			9			0.0814
6		35	9.6		28	10.3		7	7.7	
5		63	17.4		51	18.8		12	13.2	
4		82	22.6		53	19.5		29	31.9	
3		36	9.9		29	10.6		7	7.7	
2		34	9.4		24	8.8		10	11.0	
1		113	31.1		87	32.0		26	28.6	
Unreactive mydriasis or anisocoria	5	140	36.5	2	114	40.1	3	26	26.0	0.0122
CT scan classification										
Marshall classification										
Diffuse injury I	0			0			0			0.1255
Diffuse injury II		10	2.6		5	1.8		5	5.0	
Diffuse injury III		89	23.2		64	22.6		25	25.0	
Diffuse injury IV		33	8.6		24	8.5		9	9.0	
Evacuated mass lesion V		25	6.5		22	7.8		3	3.0	
Non-evacuated mass lesion VI		151	39.5		107	37.8		44	44.0	
		75	19.6		61	21.5		14	14.0	
Osmotherapy at admission	2	255	66.9	2	188	66.9	0	67	67.0	0.9860
Evacuation of subdural or extradural hematoma at admission	0	108	28.2	0	72	25.4	0	36	36.0	0.0437
External ventricular drain at admission	0	28	7.3	0	23	8.1	0	5	5.0	0.3018

Evacuation of cerebral hematoma or lobectomy at admission	0	17	4.4	0	8	2.8	0	9	9.0	0.0196
Decompressive craniectomy at admission	0	74	19.3	0	53	18.7	0	21	21.0	0.6208
	NA	**m**	**sd**	**NA**	**m**	**sd**	**NA**	**m**	**sd**	
Age (years)	0	40.5	18.7	0	41.1	19.4	0	38.9	16.6	0.2673
BMI (kg.m-2)	48	24.3	4.5	39	24.4	4.6	9	23.8	4.5	0.2710
Intracranial pressure at admission (mm Hg)	17	22.0	15.8	14	22.8	15.7	3	19.7	16.0	0.0974
Hemoglobin (g/dL)	3	11.0	2.5	1	11.0	2.5	2	11.2	2.3	0.4672
Leukocytes (count/mm^3)	4	17.9	7.1	1	18.2	7.3	3	16.9	6.3	0.0967
Prothrombin (%)	15	69.6	18.9	9	68.3	19.2	6	73.4	17.8	0.0191
Platelets (count/mm^3)	4	178.4	70.0	1	177.1	70.4	3	182.3	69.2	0.0232
Fibrinogen (g/L)	70	2.3	1.2	38	2.2	1.1	32	2.7	1.6	0.0232
Arterial lactate (mmol/L)	52	2.8	2.3	32	2.9	2.5	20	2.4	1.4	0.0079
Arterial pH	4	7.3	0.1	2	7.3	0.1	2	7.3	0.1	0.0005
Bicarbonate (mmol/L)	4	20.9	4.1	3	20.4	4.0	1	22.2	4.3	0.0005
PaO2/FiO2 ratio	14	321.0	179.4	8	313.3	179.2	6	343.5	179.0	0.0069
Serum creatinine (mmol/L)	4	78.5	31.4	1	80.7	33.3	3	72.2	23.9	0.0069
Serum proteins (g/L)	9	54.3	11.0	4	53.4	11.3	5	57.1	9.6	0.0022
Serum glucose (mmol/L)	31	8.7	3.4	17	9.0	3.6	14	8.0	2.8	0.0105
Serum urea (mmol/L)	9	4.9	2.1	5	5.0	2.2	4	4.5	1.7	0.0215
Serum calcium (mmol/L)	42	2.0	0.2	27	1.9	0.2	15	2.0	0.2	0.0180
SAPS II score	19	46.0	12.4	12	47.3	12.2	7	42.2	12.2	0.0006
ISS score	2	27.1	14.7	2	27.5	15.4	0	25.9	12.5	0.2953
IMPACT TBI score	38	8.8	4.3	23	9.1	4.3	15	7.6	4.1	0.0041

p-values are obtained using Chi-square test for categorical variables and using Student t-test for continuous variables.

BMI, Body Mass Index; FIO2, Fraction of Inspired Oxygen; GCS, Glasgow Coma Score; ISS, Injury Severity Score; NA, Not Available; PaO2, arterial partial Pressure of Oxygen; SAP, Systolic Arterial Pressure; SD, standard deviation; SAPS, Simplified Acute Physiology Score.

Figure S1. Propensity score distributions

Figure S1.a. Propensity score distribution for survival analysis according to the exposure groups of interest.

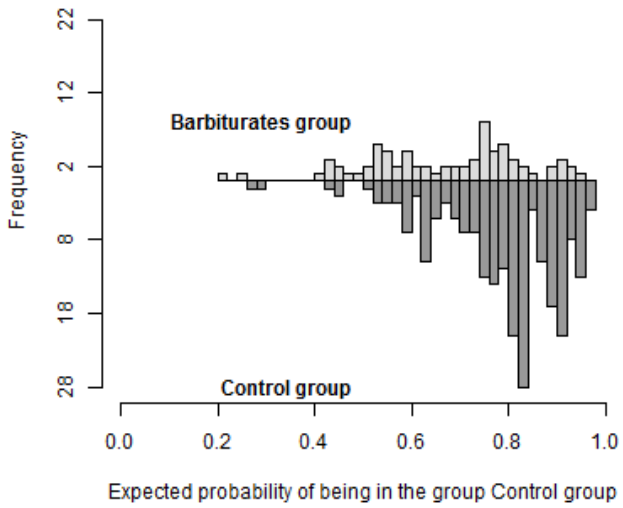


Figure S1.b. Propensity score distribution for survival analysis with the second model (TBI IMPACT score) according to the exposure groups of interest.

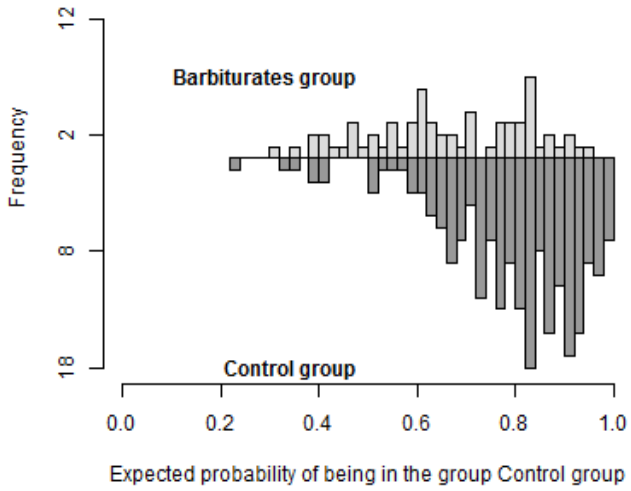


Figure S1.c. Propensity score distribution for incidence analysis of ventilator-associated pneumonia according to the exposure groups of interest.

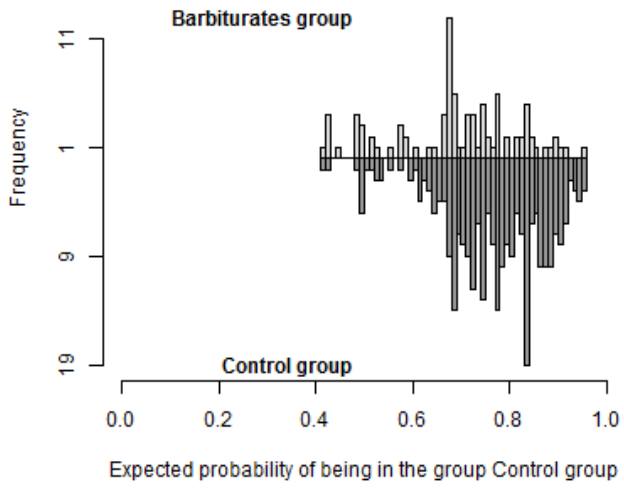


Figure S1.d. Propensity score distribution for analysis of dichotomized Glasgow Outcome Scale score at 3 months according to the exposure groups of interest.

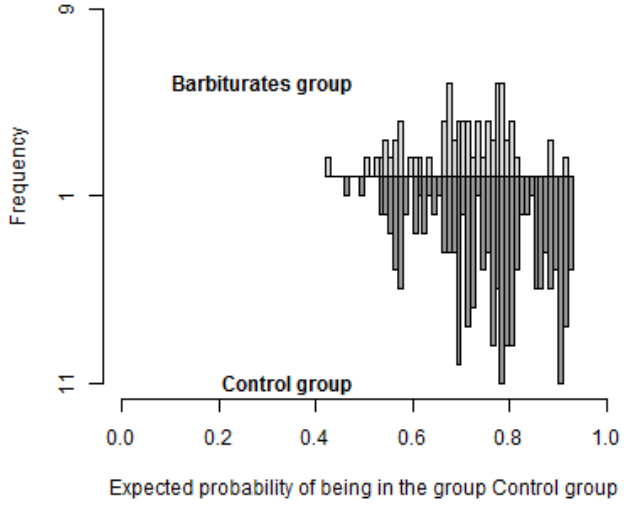


Figure S1.e. Propensity score distribution for analysis of dichotomized Glasgow Outcome Scale score at 3 months with the second model (TBI IMPACT score) according to the exposure groups of interest.

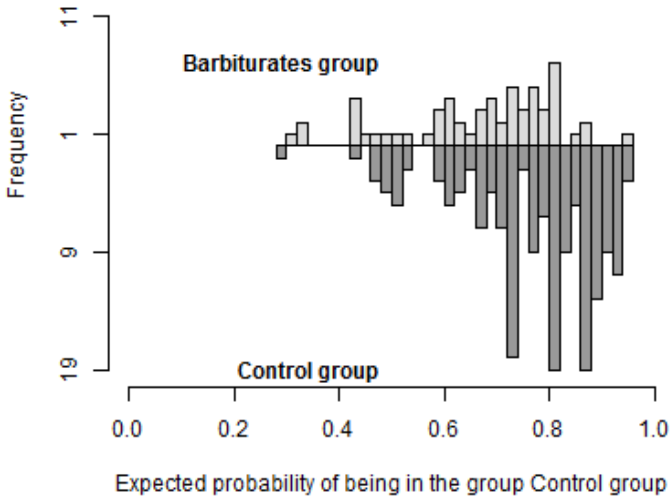
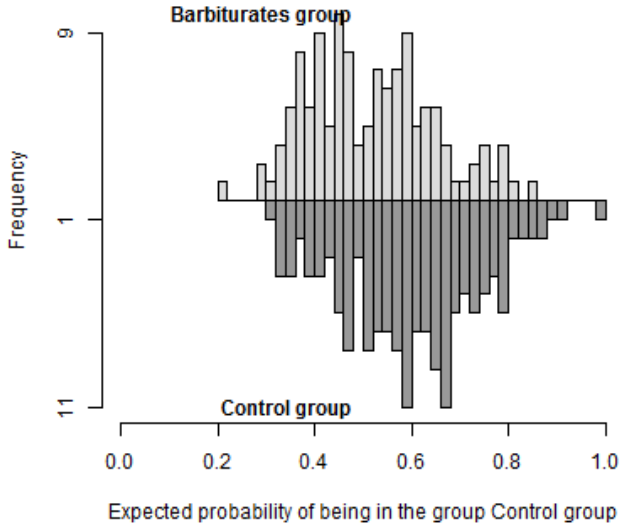


Figure S1.f. Propensity score distribution for survival analysis according to the exposure groups of interest, for the sensitivity analysis: all patients who were treated with barbiturates at any time during their stay versus the control group.



Chapitre 4

Comparaison de méthodes d'estimation causale dans des situations d'infraction aléatoire de la positivité

4.1 Manuscrit

Les principales mesures de performance des études de simulation, adaptées de Morris *et al.*, [76] sont résumées dans l'Annexe A de ce manuscrit.

Check for updates

Received: 28 October 2020 | Revised: 7 September 2021 | Accepted: 24 October 2021

DOI: 10.1002/bimj.202000323

Biometrical Journal

RESEARCH ARTICLE

Causal inference in case of near-violation of positivity: comparison of methods

Maxime Léger[1,2] | **Arthur Chatton**[1,3] | **Florent Le Borgne**[1,3] | **Romain Pirracchio**[4] | **Sigismond Lasocki**[2] | **Yohann Foucher**[1,5]

[1] INSERM UMR 1246 - SPHERE, Université de Nantes, Université de Tours, Nantes, France

[2] Département d'Anesthésie-Réanimation, Centre Hospitalier Universitaire d'Angers, Angers, France

[3] IDBC-A2COM, Nantes, France

[4] Department of Anesthesia and Perioperative Care, University of California, San Francisco, CA, USA

[5] Centre Hospitalier Universitaire de Nantes, Nantes, France

Correspondence
Yohann Foucher, Centre Hospitalier Universitaire de Nantes, Nantes, France.
Email: Yohann.Foucher@univ-nantes.fr

Funding information
French National Research Agency, Grant/Award Number: ANR-16-LCV1-0003-01

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. Results in this article were not fully reproducible due to insufficient quality of code and documentation.

Abstract

In causal studies, the near-violation of the positivity may occur by chance, because of sample-to-sample fluctuation despite the theoretical veracity of the positivity assumption in the population. It may mostly happen when the exposure prevalence is low or when the sample size is small. We aimed to compare the robustness of g-computation (GC), inverse probability weighting (IPW), truncated IPW, targeted maximum likelihood estimation (TMLE), and truncated TMLE in this situation, using simulations and one real application. We also tested different extrapolation situations for the sub-group with a positivity violation. The results illustrated that the near-violation of the positivity impacted all methods. We demonstrated the robustness of GC and TMLE-based methods. Truncation helped in limiting the bias in near-violation situations, but at the cost of bias in normal conditions. The application illustrated the variability of the results between the methods and the importance of choosing the most appropriate one. In conclusion, compared to propensity score-based methods, methods based on outcome regression should be preferred when suspecting near-violation of the positivity assumption.

KEYWORDS

causal inference, doubly robust estimators, g-computation, positivity, propensity score, real-world evidence, simulations

1 | INTRODUCTION

There is growing interest in causal methods (Hernán & Robins, 2020), notably the propensity score (PS)-based methods (Austin, 2011; Williamson et al., 2012). The PS is related to the exposure prediction. One can distinguish four different approaches: matching, stratification, conditional adjustment, and inverse probability weighting (IPW) (Robins et al., 2000; Rosenbaum & Rubin, 1983). IPW and matching on PS estimate marginal effects, while stratification and conditioning estimate conditional effects. In the settings of nonlinear link functions, marginal and conditional estimates may differ due to the noncollapsibility issues. IPW and matching emerge as preferable methods for estimating marginal effects (Austin,

 © 2022 Wiley-VCH GmbH

2013). However, both IPW and matching suffer efficiency limitations: IPW due to extreme weights and matching due to nonmatching subjects resulting in loss of information. Despite the problems of the most extreme subjects, IPW emerges as a preferable options in terms of both bias and precision (Abdia et al., 2017; Hajage et al., 2016; Le Borgne et al., 2016; Lendle et al., 2013). An alternative is g-computation (GC), also known as parametric g-formula or (g-)standardisation (Robins, 1986; Snowden et al., 2011; Vansteelandt & Keiding, 2011). The latter estimator relies on an outcome model rather than an exposure model like for PS-based methods. Some estimators combine GC and PS to create doubly robust estimators (DREs) aiming to minimize the impact of model misspecification on consistency (Bang & Robins, 2005; Neugebauer & van der Laan, 2005). One of the most studied doubly robust methods is the targeted maximum likelihood estimation (TMLE) (van der Laan & Rubin, 2006).

1.1 | Positivity violation

Regardless of the type of estimator, in order to conclude causally, one has to make several assumptions: consistency, conditional exchangeability, and positivity. Positivity is met if, for any combination of the covariates, there is a nonnull probability of being exposed or unexposed. Positivity violations can occur in two situations: (i) theoretical violation: we know that there are patients with a null probability of being exposed or unexposed, for example, if certain patients present a contraindication to receiving a treatment of interest; (ii) near or practical violation, sampling variability may result in subjects having a null probability of being exposed or unexposed for certain combinations of covariate values. This may be particularly frequent for cases of low exposure prevalence or small sample sizes (Westreich & Cole, 2010).

The theoretical violation is a consequence of a conceptual problem in the study design and calls for restricting the studied population (Petersen et al., 2012; Westreich & Cole, 2010), that is, excluding patients with a theoretical null probability of being exposed or unexposed (for instance, patients with a counter-indication for one of the studied treatments). In contrast, in case of near-violation, the target population is well defined. In this situation, the goal is to select an estimator that does not suffer from the near-violation. For IPW, one can empirically set threshold values for truncating (Cole & Hernán, 2008) or trimming the PS (Crump et al., 2009). These approaches aim to limit the maximum contribution of extreme observations. Truncation has the advantage of preserving clinical equipoise in the target population, whereas excluding certain subjects would result in a trimmed population that would change the estimand.

1.2 | Extrapolation issue

For the methods based on the outcome regression, the problem is an extrapolation of the outcome prediction for the patients affected by the near-violation, rather than using actual observations in the data (Neugebauer & van der Laan, 2006; van der Laan & Robins, 2003).

Let (Y, A, Z) denote the binary outcome ($Y = 1$ for events and 0 otherwise), the binary exposure ($A = 1$ for exposed individuals and 0 otherwise), and the p baseline covariates $(Z_1, \dots, Z_p)$. Let us define $f(Z_1|A)$, the density function of the quantitative covariate Z_1 conditional to A, Z_1 being a true confounder that causes both the exposure status A and the outcome Y. As illustrated in Figure 1, consider a near-violation of the positivity for $Z_1 > \alpha$ and an effect of a theoretical increase in the conditional probability of the outcome under $A = 1$ for larger values of Z_1.

Because of the lack of information when $Z_1 > \alpha$ due to the near-violation of the positivity assumption, the estimation of the exposure effect relies on extrapolating the observed effect, that is, when $Z_1 \leq \alpha$. Even when the outcome model is adequately specified in the region supported by data, the model may be inadequate for the region suffering from positivity near-violation.

The causal inferences will depend on the formulation of nontestable hypotheses.

One can note that this illustration (with Z_1 as a quantitative confounder can be extended for a binary confounder. Consider that Z_1 represents the gender. If there is no information regarding the outcomes among exposed women, one cannot properly infer the average exposure effect in the target population.

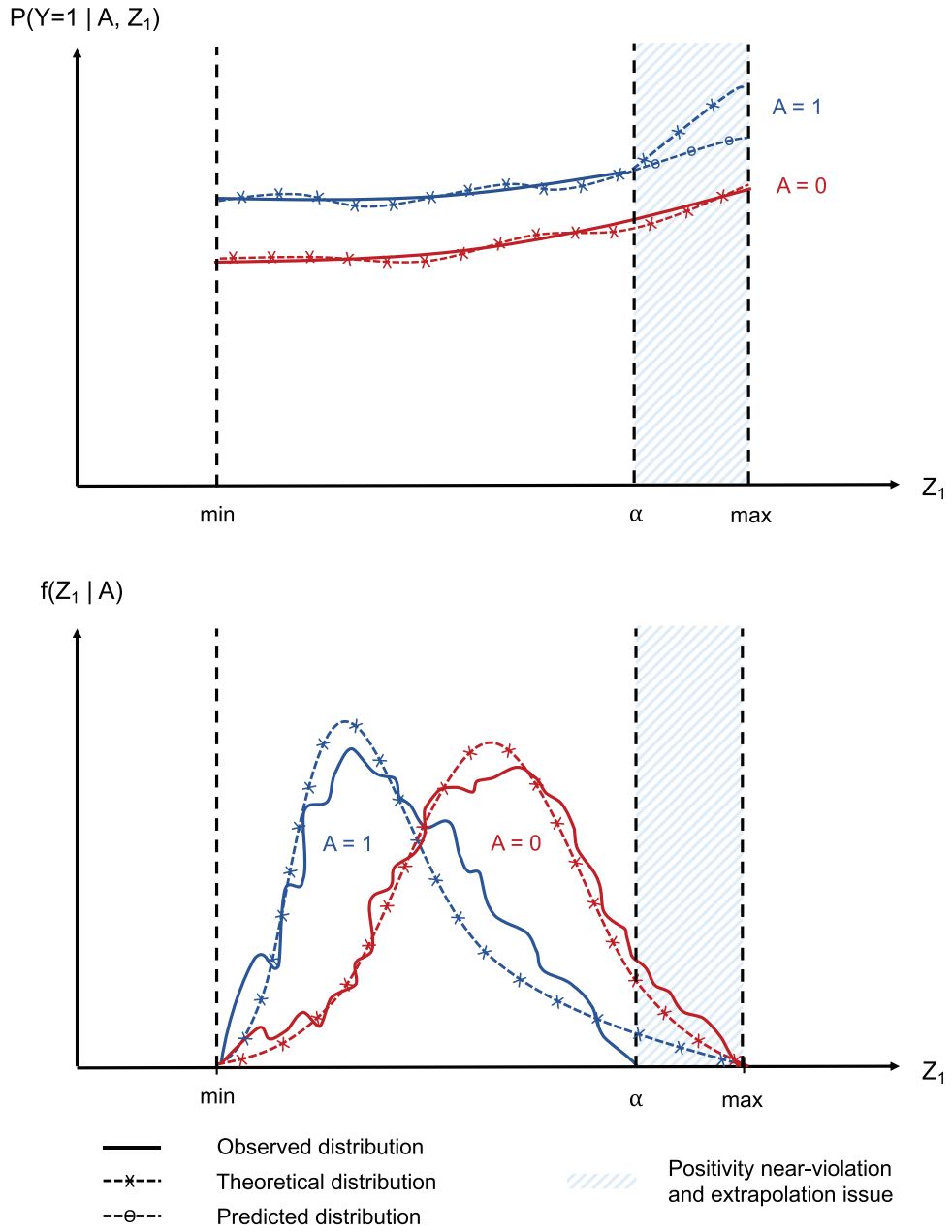


FIGURE 1 Representative illustration of the extrapolation issue occurring with a positivity near-violation. The left y-axis represents the conditional distribution function of the covariate Z_1 according to the exposure status. The right y-axis represents the conditional probability of the outcome

1.3 | Framework

The literature does not provide a clear answer as to the most reliable method in cases of positivity near-violation. Indeed, even though several studies have compared the previous methods in the context of positivity violation (Lendle et al., 2013; Moore et al., 2012; Petersen et al., 2012), suggesting better stability and reduced bias for GC and DRE, they did not investigate the extrapolation issue.

In the situation of positivity near-violation, Petersen et al. (2012) introduced the problem of extrapolation. Nevertheless, they did not study its impact.

In this context, we performed a simulation-based study to evaluate the robustness of IPW, truncated IPW, GC, TMLE, and truncated TMLE in the situation of the extrapolation issue and positivity near-violation. We also evaluated one application from a real dataset. This study is structured as follows: Section 2 outlines the methods used, Section 3 presents

the design and the results of the simulation study, in Section 4, we apply the developed application to a real dataset, and finally we discuss the results and provide recommendations.

2 | METHODS

2.1 | Setting and notations

Consider a resulting sample of size n in which one can observe the realizations of these random variables (y, a, z). Define $\pi_a = P(Y = 1|do(A = a))$ as the expected proportions of event if the entire population is exposed ($do(A = 1)$) or unexposed ($do(A = 0)$) (Pearl et al., 2016). The average exposure effect on the entire population is defined as $\Delta = \pi_1 - \pi_0$. The corresponding marginal causal odds ratio is expressed as $\text{OR} = (\pi_1/(1 - \pi_1))/(\pi_0/(1 - \pi_0))$.

2.2 | Inverse Probability Weighting

Formally, the PS for a subject i ($i = 1, \dots, n$) is $p_i = P(A = 1|z_i)$, that is, the probability that a subject is exposed according to her/his observed characteristics z_i (Rosenbaum & Rubin, 1983). The PS is often estimated from logistic regression, but other models or algorithms can be used such as random forest, boosting, or super learner (Austin, 2012; Pirracchio & Carone, 2018). IPW results in weighting the contribution of each subject i by $\omega_i = A_i P(A_i = 1)/p_i + (1 - A_i)P(A_i = 0)/(1 - p_i)$, where $P(A_i = 1)$ and $P(A_i = 0)$ denote the marginal probability of exposure and its complementary. The use of such stabilized weights are preferred to optimize the variance estimation (Robins et al., 2000; Xu et al., 2010). Based on ω_i, the maximization of the weighted likelihood of the logistic regression with Y as the outcome and A as the unique explanatory variables allows us to obtain $\hat{\pi}_0^{IPW}$, $\hat{\pi}_1^{IPW}$, and $\widehat{\text{OR}}^{IPW}$.

2.3 | Truncated IPW

The weights ω_i can largely inflate for a subject i concerned by positivity near-violation. The usual approach is to truncate the lowest and the highest p_i estimations by the 10th and 90th percentiles, respectively (Cole & Hernán, 2008). We also analyzed alternative thresholds, including the 5th and 95th percentiles, as well as the 2.5th and 97.5th percentiles of the estimated PS. We obtained truncated stabilized weights, and the estimations $\hat{\pi}_0^{T-IPW}$, $\hat{\pi}_1^{T-IPW}$, and $\widehat{\text{OR}}^{T-IPW}$.

2.4 | G-computation

GC is based on the outcome regression, frequently called the Q-model (Snowden et al., 2011). The logistic regression is often used when Y is binary. Other models or algorithms can constitute alternatives (Austin, 2012). Consider the following Q-model: $\text{logit}\{P(Y = 1|A, Z)\} = \gamma A + \beta Z$. Once fitted, one can compute for each subject i the two expected probabilities of events if she/he is exposed or unexposed, that is, $\hat{P}(Y_i = 1|do(A_i = 1), z_i)$ and $\hat{P}(Y_i = 1|do(A_i = 0), z_i)$, respectively (Snowden et al., 2011). One can then obtain $\hat{\pi}_a^{GC} = n^{-1} \sum_i \hat{P}(Y_i = 1|do(A_i = a), z_i)$ for $a = 0, 1$; $\hat{\Delta}^{GC} = \hat{\pi}_0^{GC} - \hat{\pi}_1^{GC}$ and $\widehat{\text{OR}}^{GC} = (\hat{\pi}_1^{GC}/(1 - \hat{\pi}_1^{GC}))/(\hat{\pi}_0^{GC}/(1 - \hat{\pi}_0^{GC}))$. This method is implemented in the `RISCA` package, in R (Foucher et al., 2019).

2.5 | Targeted Maximum Likelihood Estimation

The first step is to fit the Q-model and estimate the two expected probabilities of events $\hat{\pi}_1^{GC}$ and $\hat{\pi}_0^{GC}$. The additional "targeting" step involves the estimation of p_i, which is then used to update the initial estimates obtained by the Q-model. This step aims to compute first: the clever covariates $H(1, Z) = A/(\text{expit}(\hat{p}_i))$ and $H(0, Z) = (1 - A)/(1 - \text{expit}(\hat{p}_i))$, where $\text{expit}(\dot{)}$ represents the inverse logit function ($\frac{\exp(\dot{)}}{1+\exp(\dot{)}}$), and second: a vector fluctuation parameter $\hat{\epsilon} = (\hat{\epsilon}_0, \hat{\epsilon}_1)$ estimated

through a maximum likelihood procedure. The fluctuation parameter is computed using an outcome model where the logit of the initial prediction of the Q-model is an offset in an intercept-free logistic regression with the clever covariates as explanatory variables (Luque-Fernandez et al., 2018). Therefore, we can generate updated estimates of the set of potential outcomes (Y_1^* and Y_0^*) by incorporating information from the mechanisms to reduce potential biases. We generate $\text{logit}(\hat{Y}_1^*) = \text{logit}(\hat{Y}_1) + \hat{\epsilon} \times H_1$ and $\text{logit}(\hat{Y}_0^*) = \text{logit}(\hat{Y}_0) + \hat{\epsilon} \times H_0$ (Schuler & Rose, 2017). In the presence of residual confounders, the PS provides additional information to improve the initial estimates. It results in the estimations $\hat{\pi}_0^{TMLE}$ and $\hat{\pi}_1^{TMLE}$, that is, the updated values of $\hat{\pi}_0^{GC}$ and $\hat{\pi}_0^{GC}$, respectively. This method is implemented in the `tmle` package, in R (Gruber & van der Laan, 2012).

2.6 | TMLE with truncated PS

As for IPW, the TMLE can use truncated PS in its second stage. The usual method is the truncation of the lowest and highest values of p_i by 0.1 and 0.9, respectively. We also analyzed other alternative truncation levels: 0.05/0.95 and 0.025/0.975. One can then obtain $\hat{\pi}_1^{T-TMLE}$, $\hat{\pi}_0^{T-TMLE}$ and $\widehat{\text{OR}}^{T-TMLE}$. We used the `gbounds` arguments in the `tmle` function of the `tmle` package in R (Gruber & van der Laan, 2012).

2.7 | Variance estimators

For each method, the variance was obtained from the usual and well-validated method. For IPW, we used a robust sandwich-type variance estimator (Robins et al., 2000), with the `sandwich` package in R (Zeileis, 2006). For GC, we generated 1000 bootstrapped samples. This method is implemented in the `RISCA` package in R (Foucher et al., 2019). For TMLE, we used the efficient curve based variance estimator, implemented in the `tmle` package in R (Gruber & van der Laan, 2012).

To improve the comparability of the results, we additionally used the bootstrap for IPW and TMLE-based methods.

3 | SIMULATION STUDY

3.1 | Data generation

Figure S1 (Supplementary Material) represents the directed acyclic graph of the simulations. We first independently generated covariates $Z = (Z_1, \ldots, Z_9)$: six binary covariates using Bernoulli distributions with different probabilities (0.1 for Z_1, 0.4 for Z_2, 0.7 for Z_4, 0.5 for Z_5, 0.3 for Z_7, and 0.8 for Z_8), and three continuous covariates using a Gaussian distribution with mean at 0 and standard deviation at 1. We generated the exposure A according to a Bernoulli distribution with probability obtained from a logistic model with the following linear predictor: $\alpha_0 + \alpha Z_1 + \alpha Z_2 + \alpha Z_4 + \alpha Z_6 + \alpha Z_7 + \alpha Z_8$, α being the regression coefficients associated with the covariates as detailed in Table S1, and α_0 was set to 1.05 or -0.45 to simulate a prevalence of exposed patients at 80% or 50%, respectively. This design allows us to expect situations of positivity near-violation (Figures S2 and S3), especially for Z_1 which was generated with a 10% prevalence. A prevalence of 50% improved the PS distribution overlap between exposed and nonexposed subjects, and reducing the risk of positivity near-violation. Furthermore, because the near-violation is more susceptible for small samples, we studied several sample sizes: $n = 100, 200, 500$, and 1000.

We randomly generated the outcome from a Bernoulli distribution with probability obtained from a logistic model with the following linear predictor: $-0.8 + \beta_A A + \beta_Z Z_1 + \beta_Z Z_2 + \beta_Z Z_3 + \beta_Z Z_4 + \beta_Z Z_5 + \beta_Z Z_6 + \beta_{A,Z_1} A * Z_1$, where (β_A, β_Z) were the regression coefficients of A and Z, respectively. To create an extrapolation issue as illustrated in Figure 1, we considered an interaction between A and Z_1 in the outcome-generating model to obtain a poorly calibrated model in the area where Z_1 violated the positivity assumption. The values of β_A and β_Z are presented in Table S1. The regression coefficient β_{A,Z_1} of the interaction ranged from $0.0 \cdot \beta_A$ to $2.0 \cdot \beta_A$, according to the intensity of the extrapolation issues : 0.0 for no issue, 0.3 for low issues, 0.9 for moderate issues, and 2.0 for high issues.

For each of the 32 scenarios (four sample sizes, two exposures, and four extrapolation scenarios), we generated 1000 datasets. Among the generated datasets for a 50% exposure prevalence, the near-violation of the positivity assumption

(no unexposed subjects with $Z_1 = 1$) concerned 0.0% of the datasets for $n = 1000$ or 500 subjects, 1.3% for $n = 200$ subjects, and 14.1% for $n = 100$ subjects. For an 80% exposure prevalence, this near-violation concerned 0.2% of the datasets for $n = 1000$, 7.2% for $n = 500$ subjects, 31.8% for $n = 200$ subjects, and 58.2% for $n = 100$ subjects.

3.2 | Estimations

We used correctly specified exposure and outcome models to study the impact of positivity near-violation and the extrapolation issue. The interaction between Z_1 and A was introduced in both the models for data generation and the models estimated in each simulated dataset. Even if the outcome model was theoretically well specified, its estimation could result in poor calibrated predictions where there was no data support in the near-violation area.

The interaction between Z_1 and A was introduced in both the models for data generation and the models estimated in each simulated dataset. Even if the outcome model is theoretically well specified, its estimation may result in poor calibrated predictions where there is no data support in the near-violation area.

We estimated the true values of π_1 and π_0 by averaging the values obtained from a univariate logistic model (the exposure as the only covariate), fitted from datasets generated as above, except that the exposure A was simulated independently of the covariates Z (Gayat et al., 2012).

To ensure comparability between methods, we decided to set the same strategy of variables' selection. Our set of covariates corresponded to all the outcome causes, theoretically defined by the simulation design (Figure S1), that is, Z_1, Z_2, Z_3, Z_4, Z_5, and Z_6 (Chatton et al., 2020). We did not study data-adaptive methods to optimize our set of covariates (for instance, the collaborative targeted maximum likelihood estimation (van der Laan & Gruber, 2010)), or even a data-adaptive choice of the truncated PS threshold (Bembom & van der Laan, 2008).

The main estimand was the log(OR). We reported several associated criteria: the mean absolute bias (MAB) ($MAB = E(\log(\widehat{\text{OR}})) - \log(\text{OR}))$, the variance estimation ratio (VER) by the ratio of estimated model standard deviation to empirical standard deviation ($VER = (\sqrt{E[\widehat{\text{Var}}(\log(\widehat{\text{OR}}))]}/\sqrt{\text{Var}(\log(\widehat{\text{OR}}))})$, the mean square error (MSE) ($MSE = E[(\log(\widehat{\text{OR}}) - \log(\text{OR}))^2]$) , the coverage rate of the 95% confidence interval (95%CI), and the statistical power. We also reported the mean bias of the probability of an event under the two counterfactual treatments as well as their difference (Δ). We computed the Monte Carlo standard errors for each metric (Morris et al., 2019). We performed all of the analyses using the `R` software package (R Core Team, 2014).

3.3 | Results

The results are presented in Figures 2–4 for an 80% exposure prevalence. For the methods with truncation, we report in this subsection the results obtained by using the 10th and 90th percentiles, which were associated with the lower MSE values. We also performed the analyses for the 5th and 95th percentiles, and the 2.5th and 97.5th percentiles. These additional results are detailed as Supplementary Information in Tables S2 and S3 for an exposure prevalence at 80%, and in Tables S6 and S7 for an exposure prevalence at 50% (with Figures S6–8). The bootstrap-based results for an 80% exposure prevalence are presented in Tables S4, S5, with Figures S4 and S5. The results under the null hypothesis are presented in Tables S8–9 with Figures S9–11 for a prevalence of 80%, and in Tables S10–11 with Figures S12–14 for a prevalence of 50%. The standard Monte Carlo errors were negligible and are not presented in the results.

3.3.1 | Mean bias

The truncated IPW estimator was biased in almost every situation. For the other methods, the bias increased as the near-violation of positivity was accentuated, that is, when the sample size decreased (Figure 2). This increase was more significant for IPW estimators. For instance, for the scenario without an extrapolation issue, the MAB was 0.065 for a sample size of 100 subjects, versus −0.002 for 1000 subjects.

The extrapolation issue increased the MAB for methods based on the outcome modeling (GC, TMLE, and truncated TMLE), but only when the level was high. For instance, for 200 subjects without extrapolation issue, the MAB for GC and truncated TMLE were −0.006 and 0.000, respectively, versus −0.038 and −0.032 with high extrapolation issue. Even

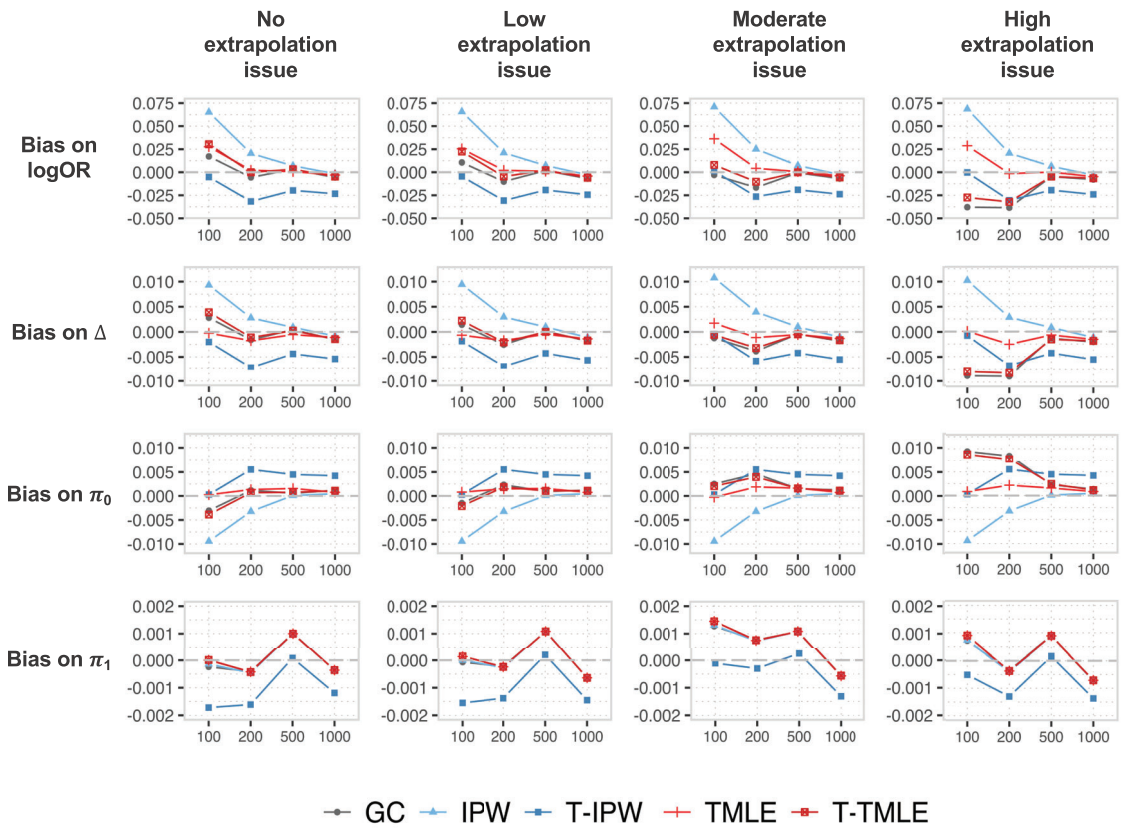


FIGURE 2 The mean absolute bias (y-axis) according to different sample size (from 100 to 1000, x-axis) and extrapolation issue. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10th and 90th percentiles); TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9); π_1, the expected proportions of event if the entire population is exposed; π_0, the expected proportions of event if the entire population is unexposed; Δ, the corresponding difference (π_1 - π_0); OR, the corresponding odds-ratio

when its level was high, the extrapolation issue had minor consequences when the sample size was equal to or higher than 500 subjects. The TMLE seemed to be the most robust method across all scenarios, especially for small sample sizes ($n = 200$).

For a prevalence of exposure of 50% where the positivity near-violation was lower, the MABs were lower for all of the methods, with comparable results in terms of bias. Even in the most extreme situations (100 subjects with high extrapolation issue), the methods remained robust.

3.3.2 | Variance

As illustrated in Figure 3, the decreases in the variance associated with the sample size was comparable across all methods. The extrapolation issue did not affect the variance estimation. However, GC was associated with larger variance when the sample size was smaller. The estimated standard deviation for GC was 1.167 for 100 subjects, 0.399 for 200 subjects, 0.244 for 500 subjects, and 0.177 for 1000 subjects. GC was the only method based on bootstrapping, which can explain this result. Therefore and for comparability sake, we subsequently used bootstrapping for the other methods (Tables S4, S5 and Figure S4). In this situation, variance was similar among all methods for 100 subjects.

Note that regardless of the method used for variance estimation, the standard deviations were similar when the prevalence was 50%. For 100 subjects without extrapolation issue, we estimated a standard deviation at 0.447 for GC, 0.458 for the IPW, 0.433 for the truncated IPW, 0.406 for the TMLE, and 0.405 for the truncated TMLE.

The VER was lower for the TMLE-based methods. This over-optimistic estimation of the variance was partially corrected for the largest sample sizes. More precisely, the VER for TMLE were 0.715 for 100 subjects, 0.798 for 200 subjects, 0.839 for

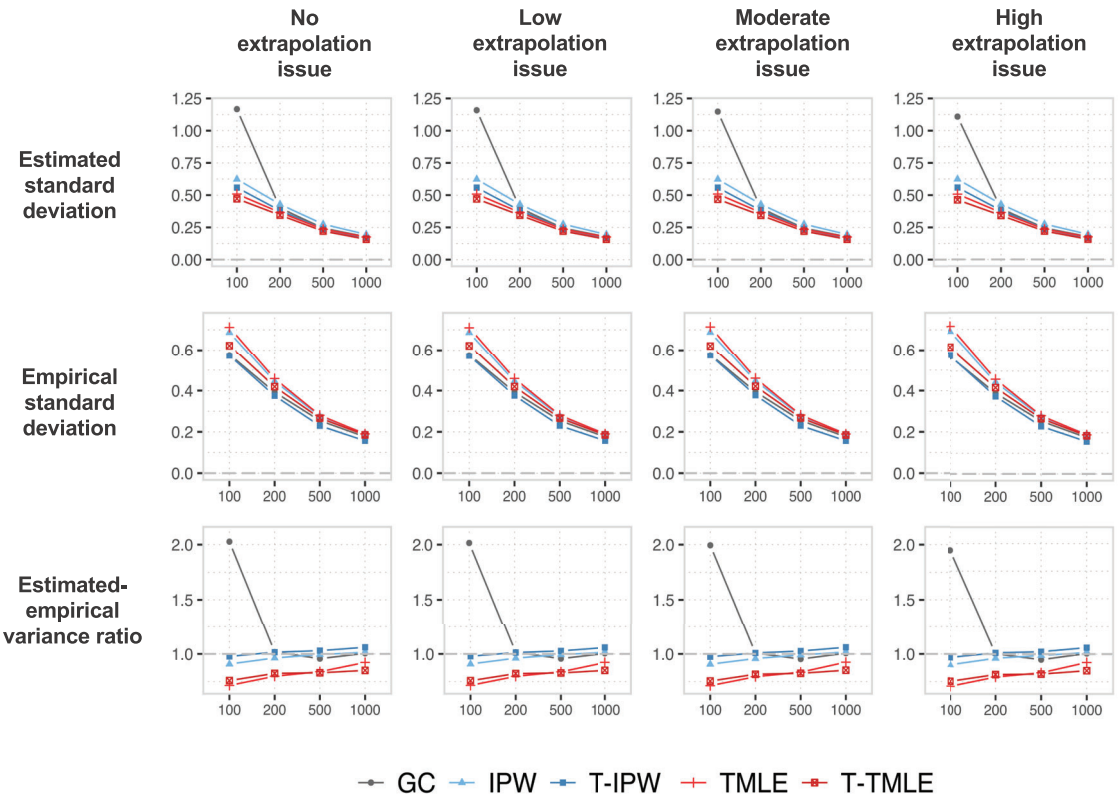


FIGURE 3 Graphical representation of the evolution of accuracy (empirical standard deviation, estimated standard deviation, and the variance estimation ratio) according to different sample size (from 100 to 1000, x-axis) and extrapolation issue. The target parameter was log(OR). Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10th and 90th percentiles); TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9)

500 subjects, and 0.923 for 1000 subjects. The use of bootstrapping corrected this over-optimistic estimation (Figure S4). Note that truncated TMLE was associated with lower variances (Tables S2 and S3).

3.3.3 | MSE, coverage and power

As illustrated in Figure 4, we observed an increase in the MSE values with the level of the positivity near-violation, in agreement with the previously reported increase in the MAB values. Nevertheless, the MSE was not significantly affected by the problem of extrapolation. The MSE was lower for GC and truncated methods in the most extreme situation. For instance, for 100 subjects, MSE values were 0.331, 0.507, 0.388, 0.475, and 0.326 for GC, TMLE, truncated TMLE, IPW, and truncated IPW, respectively. The lowest MSE was always obtained with the truncated IPW. The second method was GC. Truncated IPW and GC were the two methods with the best bias-variance trade-off. Note that when the prevalence was 50% (Tables S6 and S7), the MSE for the different methods was similar. However, truncated IPW remained the method with the lowest MSE.

As presented in Figure 4, IPW-based methods and GC resulted in nominal coverage values regardless of the sample sizes. TMLE and truncated TMLE underestimated the variance, resulting in coverage issues. For TMLE-based methods, the underestimated variance results in anti-conservative confidence intervals. More precisely, for scenarios without extrapolation issues, the coverage value of TMLE was 84.6% for 100 subjects, 88.4% for 200 subjects, 88.6% for 500 subjects, and 91.4% for 1000 subjects. The use of bootstrapping allowed to correct this underestimation. However, as reported in Tables S4 and S5, we obtained values greater than 95%, regardless of the extrapolation issue: 97.1% for 100 subjects, 96.4% for 200 subjects, 95.7% for 500 subjects, and 94.7% for 1000 subjects. The previous results under the alternative hypothesis remained consistent under the null hypothesis. The type I error rate was close to the nominal 5% value at for all methods, except for the TMLE-based methods (variance estimation with efficient curves), with values close to 10% throughout the scenarios.

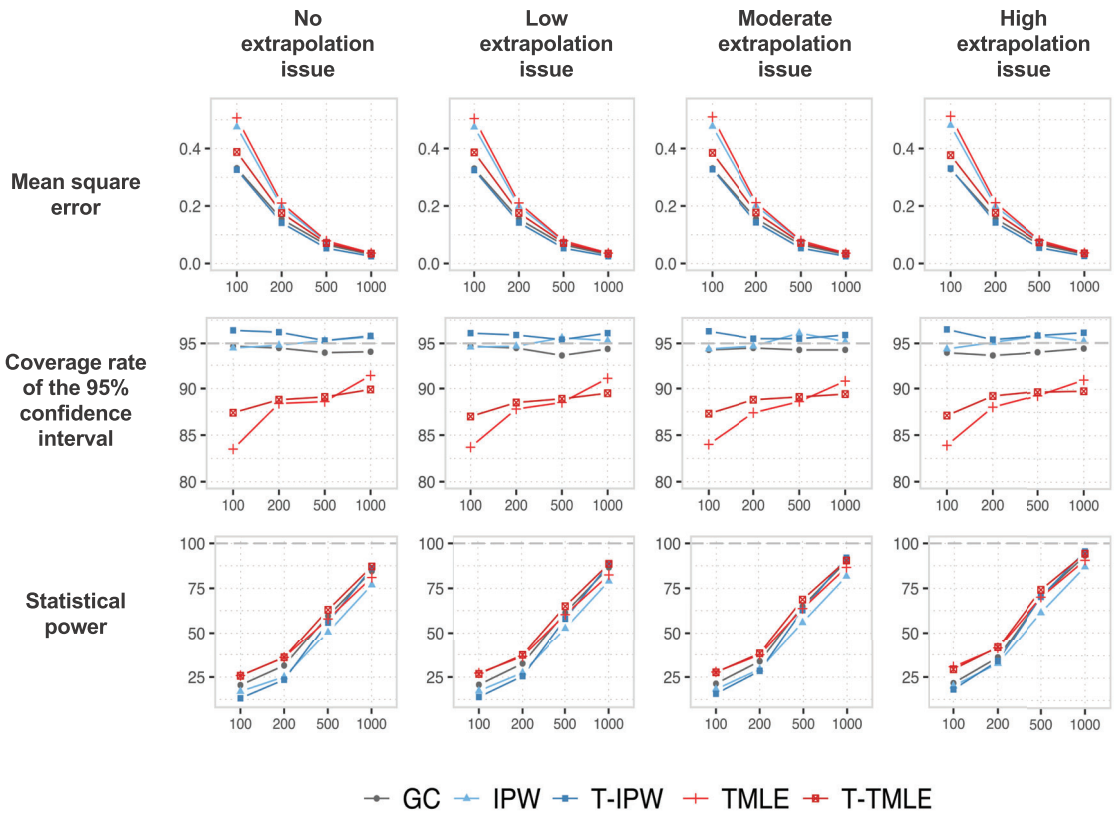


FIGURE 4 The mean square error, the coverage of the 95% confidence interval and the statistical power according to different sample size (from 100 to 1000, x-axis), and extrapolation issue. The target parameter was log(OR). Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10th and 90th percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9)

The progressive increase in the extrapolation issue allowed a slight increase in the statistical power for all methods, regardless of the magnitude of the positivity near-violation. In contrast, the statistical power was strongly impacted by the size of the population for all methods, with values around 20% per 100 subjects compared to values around 90% per 1000 subjects. The IPW presented the lowest values, while the truncated TMLE was the highest statistical power method. These results were in agreement with the over-optimistic estimation of the variance for TMLE-based methods. The power of GC was close to the truncated TMLE. The use of truncated methods improved the statistical power. For example, for 200 subjects without extrapolation issues, the powers were 36.0% for truncated TMLE, 35.9% for TMLE, 31.4% for GC, 25.1% for IPW, and 23.5% for truncated IPW. For 1000 subjects, the powers were 87.1% for truncated TMLE, 81.0% for TMLE, 85.6% for GC, 76.9% for IPW, and 86.3% for truncated IPW.

4 | APPLICATION: EFFECT OF BARBITURATES IN PATIENTS WITH INTRACRANIAL HYPERTENSION

We compared the five methods on a real dataset, in situations that could suggest a near-violation of the positivity assumption. We studied barbiturate prescription for the treatment for refractory intracranial hypertension during the first 24 h post-admission, and its relationship to, in-hospital mortality.

4.1 | Methods

We included 1584 patients from the AtlanREA cohort (www.atlanrea.org, CNIL DR-2013-047). These patients were admitted to an intensive care unit (ICU) in France's western region between March 2013 and February 2018, and were monitored for intracranial pressure.

For covariates selection, to be consistent with the simulations, we selected the covariates causing the outcome (Chatton et al., 2020). For this purpose, as proposed by VanderWeele and Shpitser (2011), we asked experts which covariates caused the outcome (i.e., a history of head trauma, use of osmotherapy, type of brain injury, age, SAPS II score, signs of intracranial hypertension on admission, lactate, and creatinine levels on admission). We did not test interactions. We applied B-spline transformations to continuous covariates when the log-linearity assumption did not hold. For IPW-based approaches, we additionally checked the balance between the two weighted groups with standardized differences. We performed complete case analyses.

4.2 | Description of the cohort

Among the 1584 patients, 1119 had no missing data on the outcome or covariates. One hundred and twenty-seven (127) patients were in the treated group versus 992 control patients (no barbiturate during the first 24 h post-admission).

We performed a comparison of analyzed patients versus patients excluded due to missing data and the results are shown in Table S12. Excluded patients were mainly less severe (higher Glasgow scores and lower SAPS II scores), with a higher proportion of women, and a different distribution of hospital care centers. Table S13 provides a comparison between the control and barbiturate-treated groups.

Sixty-six patients in the group administered barbiturates died in ICU compared to 256 in the control group. One can note that only six patients in the treated group (4.7%) were over 70 years old versus 126 (12.7%) in the control group (Figure S15). The age ranged was from 19 to 90 years old in the control group versus 19 to 76 years old in the treated group. One can explain a near-violation of the positivity because of two main reasons. First, elderly patients have a lower probability of receiving last-line treatment for intracranial hypertension because of therapy limitations (Calland et al., 2012). Second, the treatment prevalence was small, resulting in only 127 patients with barbiturates and the possible sample-to-sample fluctuation.

4.3 | Marginal effects estimates

In situations where the age-related near-violation of the positivity concerned 10% of the sample, we first performed an analyses of the overall sample. Next, we restricted the inclusion of patients to those younger than 70 years old. Figure S16 confirms that the patient age, for which we described the positivity violation, was associated with in-hospital mortality. The results are presented in Table 1 and plotted in Figure S17.

By observing the entire sample results, one can notice significant differences between the different methods. The most extreme effects were obtained with the truncated methods, while the previous simulation-based results highlighted their higher bias. More precisely, the truncated IPW (10th and 90th percentiles) had the highest OR (2.909, 95%CI from 1.990 to 4.254), while the truncated TMLE (bounds at 0.1 and 0.9) had the lowest OR (1.043, 95%CI from 0.814 to 1.338). The IPW and the TMLE were the two methods with the highest variance (0.362 and 0.299, respectively). The techniques with the lowest variances were the truncated approaches (0.127 for truncated TMLE and 0.194 for truncated IPW). Only the methods based on the TMLE have a 95% CI for the OR incorporating the value 1.

By comparing the results obtained from the entire sample with those reduced to patients under 70 years old, one can note relative stability in the estimates achieved by the five methods. Nevertheless, the estimations did not vary in the same direction: a slight increase between the estimations performed on the entire sample versus those in the subgroup for the GC, IPW, and truncated IPW, and a modest decrease in values for the TMLE-based methods. The methods with the closest results between the entire cohort and the sub-sample were based on the outcome model (TMLE, truncated TMLE, and GC). We reported a more considerable difference for IPW and truncated IPW. For instance, the OR obtained with truncated TMLE varied from 1.043 (95%CI from 0.814 to 1.338) to 1.047 (95%CI from 0.791 to 1.396), whereas the values obtained with IPW ranged from 2.158 (95%CI from 1.060 to 4.390) to 2.237 (95%CI from 1.082 to 4.624). Population restriction leads to an increase in variance, especially for TMLE-based methods (Figure S17).

The conclusions that can be drawn from the 95%CI did not change between the overall population and the restricted population. However, one can note that only the TMLE-based methods resulted in nonsignificant statistical effects, that is, rendering the study statistically "inconclusive," in contrast to the results obtained by the other methods.

TABLE 1 Results obtained by using g-computation (GC), inverse probability weighting (IPW), truncated IPW, targeted maximum likelihood estimator (TMLE), and truncated TMLE for estimating barbiturates effects

AtlanREA cohort: the barbiturates effect							
	π_1	π_0	**Δ**	**log(OR)**	**SD**	**OR**	**95%CI OR**
Whole sample							
GC	0.396	0.272	0.124	0.560	0.194	1.750	1.207 - 2.524
IPW	0.447	0.273	0.174	0.769	0.362	2.158	1.060 - 4.390
Truncated IPW [10–90%]	0.515	0.267	0.248	1.068	0.194	2.909	1.990 - 4.254
Truncated IPW [5–95%]	0.470	0.271	0.199	0.872	0.236	2.391	1.506 - 3.795
Truncated IPW [2.5–97.5%]	0.467	0.272	0.195	0.853	0.245	2.347	1.453 - 3.792
TMLE	0.320	0.289	0.031	0.146	0.299	1.158	0.645 - 2.079
Truncated TMLE [0.1–0.9]	0.298	0.288	0.010	0.043	0.127	1.043	0.814 - 1.338
Truncated TMLE [0.05–0.95]	0.311	0.289	0.022	0.107	0.156	1.112	0.820 - 1.509
Truncated TMLE [0.025–0.975]	0.311	0.289	0.022	0.108	0.219	1.114	0.725 - 1.711
Restricted sample							
GC	0.370	0.243	0.127	0.606	0.196	1.833	1.259 - 2.670
IPW	0.418	0.243	0.175	0.805	0.371	2.237	1.082 - 4.624
Truncated IPW [10–90%]	0.499	0.238	0.261	1.160	0.203	3.188	2.142 - 4.746
Truncated IPW [5–95%]	0.447	0.242	0.205	0.932	0.242	2.539	1.579 - 4.082
Truncated IPW [2.5–97.5%]	0.447	0.243	0.204	0.902	0.250	2.464	1.509 - 4.025
TMLE	0.280	0.263	0.017	0.090	0.344	1.094	0.558 - 2.145
Truncated TMLE [0.1–0.9]	0.272	0.263	0.009	0.046	0.143	1.047	0.791 - 1.386
Truncated TMLE [0.05–0.95]	0.274	0.263	0.011	0.056	0.183	1.058	0.739 - 1.515
Truncated TMLE [0.025–0.975]	0.275	0.263	0.012	0.066	0.269	1.068	0.630 - 1.809

Abbreviations: π_1, the expected proportions of event if the entire population is exposed; π_0, the expected proportions of event if the entire population is unexposed; Δ, the risk difference (π_1 - π_0); log(OR), the logarithm of the odds ratio; SD, the standard deviation for the logarithm of the odds ratio; OR, the corresponding odds-ratio ; 95%CI, 95% confidence interval of the odds-ratio.

5 | DISCUSSION

The results of the simulations illustrated that the near-violation of the positivity assumption could impact the bias and precision of the five methods. In terms of MAB, one can conclude that methods based on the outcome modeling showed the best results. The addition of an extrapolation issue altered the MAB for these methods, but in a magnitude similar to the one observed for the IPW-based approaches. Whilst the truncated methods introduced bias, they reduced the variance estimation, as previously described by Moore et al. (2012). Methods with the best balance between variance and bias were truncated IPW and GC. TMLE-based methods were associated with an over-optimistic estimate of the variance, resulting in lower coverage than the nominal value. We did not observe this issue when the prevalence of exposure was 50%, that is, reducing the positivity near-violation. Although the TMLE is a doubly robust estimator consistent when at least one nuisance model is well-specified, the variance estimation can be challenging. Petersen et al. (2014) and Lendle et al. (2017) reported the potential inflation of the type I error and poor coverage in the presence of positivity near-violations. Our results confirm their findings and the potential of a bootstrap-based approach as an alternative. We performed additional simulations with the average exposure effect as the estimand (instead of the logOR), our results were consistent (data not shown).

Whilst the simulations illustrated important differences between each methods performance, the "real-dataset" application emphasized the importance of the method chosen. Indeed, the clinical conclusion varied according to the specific method. In agreement with the simulations, the variances of truncated methods were smaller, but this benefit has to be counterbalanced with the risk of bias (Cole & Hernán, 2008; Ju et al., 2019). The main concern lies in their optimal cut-off choice, giving us the best bias-variance trade-off. We have studied consensual thresholds, defined either by a bound value of PS (for TMLE) or by the value of a percentile of the weights (for IPW). An alternative would lie in establishing an algorithm seeking the best bias-variance trade-off, which would be guided by the data. This solution has recently been studied to choose data-driven PS truncation thresholds adapted to IPW (Bembom & van der Laan, 2008) or to TMLE (Ju et al.,

2019), with promising results for positivity violation situations. Another solution may also lie in the use of modern methods such as limited overlap, matching and entropy weightings to reduce the influence of the most extreme observations and focus on the data area with the most overlap, therefore capturing the processing effect for which we have the most information (Zhou et al., 2020). These techniques enable us to estimate an average treatment effect on the population overlap (Li et al., 2018).

Causal inference in observational studies relies directly on the assumption that all participants are eligible to be exposed (or unexposed). Our results confirmed the importance of this assumption since all the methods compared were affected in terms of bias and/or variance. This assumption's violation is more identifiable by using PS-based approaches as it consists of regressing the exposure probability. In contrast, GC involves outcome modeling, and this violation can remain unidentified (Kang & Schafer, 2007). For IPW, subjects who have a low likelihood of exposure but who are exposed, results in extreme weights with unstable estimations and high variances (Kang & Schafer, 2007). The inflated variance and the associated extreme weighting obtained in this way can alert investigators. Unfortunately, the situations at risk of extrapolation are not directly identifiable, and only the violations of positivity can be revealed.

The near-violation of the positivity represents an obstacle to causal inferences only when it concerns true confounders, that is, those associated with both the exposure and the outcome (Westreich & Cole, 2010). In contrast, imbalance of variables was only associated with exposure, also called instrumental variables, and will have no impact on the bias.

Several authors have previously documented different techniques for detecting restrictions on the positivity assumption in the context of PS analysis (Austin & Stuart, 2015; Cole & Hernán, 2008). The first approach is to study the distribution of the exposure regimen for each covariate, but this can become tedious when dealing with many covariates. One can also use standardized differences (Austin & Stuart, 2015). Another possibility is to compare a groups weights distributions, or even to focus on the distribution of PS. Histogram of the PS distribution by exposure group is an example of interesting representation. In practice, one can assess the positivity assumption by searching for a lack of sufficient overlap of the PS distributions between the exposure groups. However, while useful to diagnose potential positivity violations, these techniques do not provide any quantitative estimate of the estimator bias due to positivity near-violation. Petersen et al. (2012) proposed a parametric bootstrap approach to provide an optimistic bias estimate specifically targeted for positivity violations and near-violations.

Our study has several limitations. First, we only considered TMLE-based methods, while other DRE approaches exist, such as the augmented inverse probability of treatment weighting (A-IPTW) (Glynn & Quinn, 2010). We focused on TMLE because of its better stability compared to A-IPTW (Luque-Fernandez et al., 2018; Neugebauer & van der Laan, 2005; Porter et al., 2011). Second, we did not study the different methods for the construction of the model, as this would have multiplied the number of possible approaches to compare. For instance, an alternative to reduce the variance of TMLE is the collaborative TMLE (C-TMLE), which uses a sequential selection of covariates estimating PS (Lendle et al., 2013; Pirracchio et al., 2018; Porter et al., 2011). Machine learning techniques were also proposed for GC (Austin, 2012), or for PS-based methods (Pirracchio et al., 2015). The improvement of the methods we studied by machine learning techniques is an interesting perspective of our work, especially because it can help to reduce the problem of extrapolation. Third, our simulation-based study was not associated with theoretical justification, and it does not demonstrate which method is the best in all situations. Even though our results are in agreement with the current literature, additional studies are required, such as incorporating the extrapolation issue for patients with a higher susceptibility of positivity near-violation.

To conclude, our study illustrates that all the causal methods were sensitive to the near-positivity violation. Nevertheless, we reported the methods' robustness based on the outcome model (GC and TMLE), even with an extrapolation issue. The truncated method, whilst attractive in terms of variance reduction, should be used with caution due to the associated risk of increased bias. G-computation appears to present the best compromise when considering its ability to reduce the bias and its statistical power.

ACKNOWLEDGMENTS

The authors would like to thank the members of AtlanREA Group for their involvement in the study, the physicians who helped recruit patients, and all patients who participated in this study. We also thank the clinical research associates who participated in the data collection. The analyses and interpretation of these data are the responsibility of the authors. This work was partially supported by a public grant overseen by the French National Research Agency (ANR) to create the Common Laboratory RISCA (www.labcom-risca.com, reference: ANR-16-LCV1-0003-01). The funder had no role in study design, analyses, and interpretation of data, writing the report, or the decision to submit the report for publication.

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. Results in this article were not fully reproducible due to insufficient quality of code and documentation.

ORCID

Maxime Léger https://orcid.org/0000-0002-5447-6595

REFERENCES

Abdia, Y., Kulasekera, K. B., Datta, S., Boakye, M., & Kong, M. (2017). Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study. *Biometrical Journal*, *59*, 967–985.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res*, *46*, 399–424.

Austin, P. C. (2012). Using ensemble-based methods for directly estimating causal effects: An investigation of tree-based G-computation. *Multivariate Behavioral Research*, *47*, 115–135.

Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine*, *32*, 2837–2849.

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Statistics in Medicine*, *34*, 3661–3679.

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics*, *61*, 962–973.

Bembom, O., & van der Laan, M. J. (2008). Data-adaptive selection of the truncation level for inverse-probability-of-treatment-weighted estimators. U.C. Berkeley Division of Biostatistics Working Paper Series.

Calland, J. F., Ingraham, A. M., Martin, N., Marshall, G. T., Schulman, C. I., Stapleton, T., Barraco, R. D., & Eastern Association for the Surgery of Trauma (2012). Evaluation and management of geriatric trauma: An eastern association for the surgery of trauma practice management guideline. *Journal of Trauma and Acute Care Surgery*, *73*, S345–350.

Chatton, A., Le Borgne, F., Leyrat, C., Gillaizeau, F., Rousseau, C., Barbin, L., Laplaud, D., Léger, M., Giraudeau, B., & Foucher, Y. (2020). G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: A comparative simulation study. *Scientific Reports*, *10*, 9219.

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, *168*, 656–664.

Crump, R. K., Hotz, V. J., Imbens, G. W., & Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, *96*, 187–199.

Foucher, Y., Borgne, F. L., Dantan, E., Gillaizeau, F., Chatton, A., & Combescure, C. (2019). RISCA: Causal inference and prediction in cohort-based analyses. https://CRAN.R-project.org/package=RISCA.

Gayat, E., Resche-Rigon, M., Mary, J.-Y., & Porcher, R. (2012). Propensity score applied to survival data analysis through proportional hazards models: A Monte Carlo study. *Pharmaceutical Statistics*, *11*, 222–229.

Glynn, A. N., & Quinn, K. M. (2010). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis*, *18*, 36–56.

Gruber, S., & van der Laan, M. J. (2012). TMLE: An R package for targeted maximum likelihood estimation. *Journal of Statistical Software*, *51*, 1–35. http://www.jstatsoft.org/v51/i13/. https://doi.org/10.18637/jss.v051.i13.

Hajage, D., Tubach, F., Steg, P. G., Bhatt, D. L., & De Rycke, Y. (2016). On the use of propensity scores in case of rare exposure. *BMC Medical Research Methodology*, *16*, 32.

Hernán, M. A., & Robins, J. M. (2020). *Causal Inference: What if?*. Chapman & Hall/CRC. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/.

Ju, C., Schwab, J., & van der Laan, M. J. (2019). On adaptive propensity score truncation in causal inference. *Statistical Methods in Medical Research*, *28*, 1741–1760.

Kang, J. D. Y., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science*, *22*, 523–539.

Le Borgne, F., Giraudeau, B., Querard, A. H., Giral, M., & Foucher, Y. (2016). Comparisons of the performance of different statistical tests for time-to-event analysis with confounding factors: Practical illustrations in kidney transplantation. *Statistics in Medicine*, *35*, 1103–1116.

Lendle, S. D., Fireman, B., & van der Laan, M. J. (2013). Targeted maximum likelihood estimation in safety analysis. *Journal of Clinical Epidemiology*, *66*, S91–98.

Lendle, S. D., Schwab, J., Petersen, M. L., & van der Laan, M. J. (2017). LTMLE: An R package implementing targeted minimum loss-based estimation for longitudinal data. *Journal of Statistical Software*, *81*, 1–21. https://www.jstatsoft.org/index.php/jss/article/view/v081i01.

Li, F., Morgan, K. L., & Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, *113*, 390–400.

Luque-Fernandez, M. A., Schomaker, M., Rachet, B., & Schnitzer, M. E. (2018). Targeted maximum likelihood estimation for a binary treatment: A tutorial. *Statistics in Medicine*, *37*, 2530–2546.

Moore, K. L., Neugebauer, R., van der Laan, M. J., & Tager, I. B. (2012). Causal inference in epidemiological studies with strong confounding. *Statistics in Medicine*, *31*, 1380–1404.

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*, 2074–2102.

Neugebauer, R., & van der Laan, M. J. (2005). Why prefer double robust estimators in causal inference? *Journal of Statistical Planning and Inference*, *129*, 405–426.

Neugebauer, R., & van der Laan, M. J. (2006). G-computation estimation for causal inference with complex longitudinal data. *Computational Statistics & Data Analysis*, *51*, 1676–1697.

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons.

Petersen, M. L., Porter, K. E., Gruber, S., Wang, Y., & van der Laan, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Statistical Methods in Medical Research*, *21*, 31–54.

Petersen, M., Schwab, J., Gruber, S., Blaser, N., Schomaker, M., & van der Laan, M. (2014). Targeted maximum likelihood estimation for dynamic and static longitudinal marginal structural working models. *Journal of Causal Inference*, *2*, 147–185.

Pirracchio, R., & Carone, M. (2018). The balance super learner: A robust adaptation of the super learner to improve estimation of the average treatment effect in the treated based on propensity score matching. *Statistical Methods in Medical Research*, *27*, 2504–2518.

Pirracchio, R., Petersen, M. L., & van der Laan, M. (2015). Improving propensity score estimators' robustness to model misspecification using super learner. *American Journal of Epidemiology*, *181*, 108–119.

Pirracchio, R., Yue, J. K., Manley, G. T., van der Laan, M. J., & Hubbard, A. E. (2018). Collaborative targeted maximum likelihood estimation for variable importance measure: Illustration for functional outcome prediction in mild traumatic brain injuries. *Statistical Methods in Medical Research*, *27*, 286–297.

Porter, K. E., Gruber, S., van der Laan, M. J., & Sekhon, J. S. (2011). The relative performance of targeted maximum likelihood estimator. *International Journal of Biostatistics*, *7*, 1557–4679.

R Core Team. (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Robins, J. M. (1986). A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling*, *7*, 1393–1512.

Robins, J. M., Hernán, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, *11*, 550–560.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41–55.

Schuler, M. S., & Rose, S. (2017). Targeted maximum likelihood estimation for causal inference in observational studies. *American Journal of Epidemiology*, *185*, 65–73.

Snowden, J. M., Rose, S., & Mortimer, K. M. (2011). Implementation of G-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology*, *173*, 731–738.

van der Laan, M. J., & Gruber, S. (2010). Collaborative double robust targeted maximum likelihood estimation. *The International Journal of Biostatistics*, *6*, 17–17.

van der Laan, M. J., & Robins, J. M. (2003). *Unified Methods for Censored Longitudinal Data and Causality*. Springer.

van der Laan, M. J., & Rubin, D. B. (2006). Targeted maximum likelihood learning. *The International Journal of Biostatistics*, *2*, https://doi.org/10.2202/1557-4679.1043.

VanderWeele, T. J., & Shpitser, I. (2011). A new criterion for confounder selection. *Biometrics*, *67*, 1406–1413.

Vansteelandt, S., & Keiding, N. (2011). Invited commentary: G-computation-lost in translation? *American Journal of Epidemiology*, *173*, 739–742.

Westreich, D., & Cole, S. R. (2010). Invited commentary: Positivity in practice. *American Journal of Epidemiology*, *171*, 674–677; discussion 678–681.

Williamson, E. J., Morley, R., Lucas, A., & Carpenter, J. (2012). Propensity scores: From naïve enthusiasm to intuitive understanding. *Statistical Methods in Medical Research*, *21*, 273–293.

Xu, S., Ross, C., Raebel, M. A., Shetterly, S., Blanchette, C., & Smith, D. (2010). Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. *Value Health*, *13*, 273–277.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, *16*, 1–16.

Zhou, Y., Matsouaka, R. A., & Thomas, L. (2020). Propensity score weighting under limited overlap and model misspecification. *Statistical Methods in Medical Research*, *29*, 3721–3756.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Léger, M., Chatton, A., Le Borgne, F., Pirracchio, R., Lasocki, S., Foucher, Y. Causal inference in case of near-violation of positivity: comparison of methods. *Biometrical Journal*. 2022;1–15. https://doi.org/10.1002/bimj.202000323

4.2 Éléments supplémentaires

TABLE S1. The values of the regression coefficients used in the simulation study.

Logistic model	Coefficients
Exposure	$\alpha_{Z_1} = \log(5)$
	$\alpha_{Z_2} = \log(1.5)$
	$\alpha_{Z_3} = 0$
	$\alpha_{Z_4} = \log(1.5)$
	$\alpha_{Z_5} = 0$
	$\alpha_{Z_6} = \log(1.5)$
	$\alpha_{Z_7} = \log(1.5)$
	$\alpha_{Z_8} = \log(1.5)$
	$\alpha_{Z_9} = 0$
Outcome	$\beta_A = \log(1.75)$
	$\beta_{Z_1} = \log(1.5)$
	$\beta_{Z_2} = \log(1.5)$
	$\beta_{Z_3} = \log(1.5)$
	$\beta_{Z_4} = \log(1.5)$
	$\beta_{Z_5} = \log(1.5)$
	$\beta_{Z_6} = \log(1.5)$
	$\beta_{Z_7} = 0$
	$\beta_{Z_8} = 0$
	$\beta_{Z_9} = 0$
	$\beta_{Z_10} = 0$
	$\beta_{Z_11} = 0$
	$\beta_{Z_12} = 0$
	$\beta_{A,Z_1} = 0$ for no extrapolation issue
	$\beta_{A,Z_1} = 0.3\beta_A$ for low extrapolation issue
	$\beta_{A,Z_1} = 0.9\beta_A$ for moderate extrapolation issue
	$\beta_{A,Z_1} = 2.0\beta_A$ for high extrapolation issue

TABLE S2. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 80% with a binary outcome, for sample sizes of 1,000 and 500 subjects. The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Power
1000	High	Crude	0.496	0.350	0.606	-0.032	0.001	0.008	0.027	0.161	0.166	1.030	94.90	95.40
1000	High	GC	0.496	0.350	0.606	-0.008	-0.001	0.001	0.031	0.176	0.177	1.006	94.40	93.20
1000	High	IPW	0.496	0.350	0.606	-0.003	-0.001	0.000	0.037	0.192	0.195	1.014	95.20	86.90
1000	High	T-IPW 0.1/0.9	0.496	0.350	0.606	-0.024	-0.001	0.004	0.025	0.157	0.166	1.059	96.10	95.50
1000	High	T-IPW 0.05/0.95	0.496	0.350	0.606	-0.008	-0.001	0.001	0.025	0.159	0.169	1.065	95.90	95.50
1000	High	T-IPW 0.025/0.975	0.496	0.350	0.606	0.001	-0.001	-0.001	0.026	0.162	0.172	1.063	96.10	95.50
1000	High	TMLE	0.496	0.350	0.606	-0.005	-0.001	0.001	0.036	0.190	0.175	0.924	91.00	90.60
1000	High	T-TMLE 0.1/0.9	0.496	0.350	0.606	-0.006	-0.001	0.001	0.034	0.185	0.157	0.850	89.80	94.30
1000	High	T-TMLE 0.05/0.95	0.496	0.350	0.606	-0.006	-0.001	0.001	0.035	0.188	0.168	0.893	90.70	93.10
1000	High	T-TMLE 0.025/0.975	0.496	0.350	0.606	-0.005	-0.001	0.001	0.036	0.189	0.174	0.920	91.00	90.90
1000	Moderate	Crude	0.484	0.350	0.560	-0.040	-0.001	0.008	0.027	0.161	0.166	1.030	94.90	90.30
1000	Moderate	GC	0.484	0.350	0.560	-0.007	-0.001	0.001	0.031	0.176	0.177	1.007	94.30	90.10
1000	Moderate	IPW	0.484	0.350	0.560	-0.002	-0.001	0.000	0.037	0.192	0.195	1.017	95.20	81.70
1000	Moderate	T-IPW 0.1/0.9	0.484	0.350	0.560	-0.024	-0.001	0.004	0.025	0.157	0.166	1.061	95.90	92.00
1000	Moderate	T-IPW 0.05/0.95	0.484	0.350	0.560	-0.008	-0.001	0.001	0.025	0.159	0.169	1.068	95.90	92.40
1000	Moderate	T-IPW 0.025/0.975	0.484	0.350	0.560	0.002	-0.001	-0.001	0.026	0.161	0.172	1.066	95.80	92.40
1000	Moderate	TMLE	0.484	0.350	0.560	-0.004	-0.001	0.001	0.036	0.190	0.175	0.925	90.80	86.60
1000	Moderate	T-TMLE 0.1/0.9	0.484	0.350	0.560	-0.006	-0.001	0.001	0.034	0.185	0.157	0.852	89.40	90.50
1000	Moderate	T-TMLE 0.05/0.95	0.484	0.350	0.560	-0.005	-0.001	0.001	0.035	0.187	0.168	0.896	90.40	88.20
1000	Moderate	T-TMLE 0.025/0.975	0.484	0.350	0.560	-0.004	-0.001	0.001	0.036	0.189	0.174	0.922	90.60	87.10
1000	Low	Crude	0.477	0.350	0.532	-0.045	-0.002	0.008	0.028	0.161	0.166	1.030	95.10	85.00
1000	Low	GC	0.477	0.350	0.532	-0.007	-0.001	0.001	0.031	0.176	0.177	1.005	94.40	86.60
1000	Low	IPW	0.477	0.350	0.532	-0.003	-0.001	0.000	0.037	0.192	0.195	1.014	95.30	79.00
1000	Low	T-IPW 0.1/0.9	0.477	0.350	0.532	-0.024	-0.001	0.004	0.025	0.157	0.166	1.059	96.10	88.40
1000	Low	T-IPW 0.05/0.95	0.477	0.350	0.532	-0.008	-0.001	0.001	0.025	0.159	0.169	1.065	95.90	88.90
1000	Low	T-IPW 0.025/0.975	0.477	0.350	0.532	0.001	-0.001	-0.001	0.026	0.162	0.172	1.063	96.00	88.80
1000	Low	TMLE	0.477	0.350	0.532	-0.004	-0.001	0.001	0.036	0.190	0.175	0.924	91.10	82.40
1000	Low	T-TMLE 0.1/0.9	0.477	0.350	0.532	-0.006	-0.001	0.001	0.034	0.185	0.158	0.851	89.50	88.60
1000	Low	T-TMLE 0.05/0.95	0.477	0.350	0.532	-0.005	-0.001	0.001	0.035	0.188	0.168	0.894	90.70	85.40
1000	Low	T-TMLE 0.025/0.975	0.477	0.350	0.532	-0.005	-0.001	0.001	0.036	0.189	0.174	0.921	90.90	82.90
1000	None	Crude	0.473	0.350	0.516	-0.047	-0.003	0.008	0.028	0.160	0.166	1.033	94.60	83.00
1000	None	GC	0.473	0.350	0.516	-0.006	-0.000	0.001	0.031	0.176	0.177	1.006	94.10	84.60
1000	None	IPW	0.473	0.350	0.516	-0.002	-0.000	0.000	0.037	0.192	0.195	1.014	95.70	76.90
1000	None	T-IPW 0.1/0.9	0.473	0.350	0.516	-0.023	-0.001	0.004	0.025	0.157	0.166	1.060	95.80	86.30
1000	None	T-IPW 0.05/0.95	0.473	0.350	0.516	-0.007	-0.001	0.001	0.025	0.159	0.169	1.067	95.40	88.00
1000	None	T-IPW 0.025/0.975	0.473	0.350	0.516	0.003	-0.000	-0.001	0.026	0.161	0.172	1.065	96.20	87.80
1000	None	TMLE	0.473	0.350	0.516	-0.003	-0.000	0.001	0.036	0.190	0.175	0.923	91.40	81.00
1000	None	T-TMLE 0.1/0.9	0.473	0.350	0.516	-0.005	-0.000	0.001	0.034	0.185	0.158	0.851	89.90	87.10
1000	None	T-TMLE 0.05/0.95	0.473	0.350	0.516	-0.004	-0.000	0.001	0.035	0.188	0.168	0.894	91.00	83.60
1000	None	T-TMLE 0.025/0.975	0.473	0.350	0.516	-0.004	-0.000	0.001	0.036	0.189	0.174	0.920	91.20	81.80
500	High	Crude	0.494	0.350	0.603	-0.027	0.002	0.008	0.058	0.238	0.237	0.993	93.50	71.20
500	High	GC	0.494	0.350	0.603	-0.005	0.001	0.002	0.066	0.257	0.244	0.950	94.00	70.30
500	High	IPW	0.494	0.350	0.603	0.006	0.001	0.000	0.077	0.278	0.275	0.988	95.80	61.30
500	High	T-IPW 0.1/0.9	0.494	0.350	0.603	-0.019	0.000	0.004	0.054	0.231	0.236	1.022	95.80	70.20
500	High	T-IPW 0.05/0.95	0.494	0.350	0.603	-0.002	0.001	0.001	0.055	0.234	0.241	1.029	96.10	71.30
500	High	T-IPW 0.025/0.975	0.494	0.350	0.603	0.008	0.001	-0.001	0.056	0.238	0.245	1.030	96.10	71.60
500	High	TMLE	0.494	0.350	0.603	0.000	0.001	0.002	0.080	0.283	0.235	0.831	89.30	70.00
500	High	T-TMLE 0.1/0.9	0.494	0.350	0.603	-0.005	0.001	0.002	0.072	0.269	0.220	0.819	89.70	74.10
500	High	T-TMLE 0.05/0.95	0.494	0.350	0.603	-0.005	0.001	0.003	0.075	0.273	0.231	0.845	90.40	71.00
500	High	T-TMLE 0.025/0.975	0.494	0.350	0.603	-0.004	0.001	0.002	0.076	0.276	0.234	0.850	90.30	69.90
500	Moderate	Crude	0.483	0.350	0.558	-0.035	0.001	0.008	0.057	0.237	0.237	0.997	93.70	62.90
500	Moderate	GC	0.483	0.350	0.558	-0.001	0.001	0.001	0.065	0.255	0.243	0.954	94.30	65.00
500	Moderate	IPW	0.483	0.350	0.558	0.007	0.001	0.000	0.077	0.278	0.275	0.991	96.10	55.90
500	Moderate	T-IPW 0.1/0.9	0.483	0.350	0.558	-0.019	0.000	0.004	0.053	0.230	0.236	1.026	95.50	62.80
500	Moderate	T-IPW 0.05/0.95	0.483	0.350	0.558	-0.001	0.001	0.001	0.054	0.233	0.241	1.034	95.80	65.10
500	Moderate	T-IPW 0.025/0.975	0.483	0.350	0.558	0.009	0.001	-0.001	0.056	0.236	0.245	1.035	96.40	64.60
500	Moderate	TMLE	0.483	0.350	0.558	0.001	0.001	0.002	0.079	0.282	0.235	0.835	88.60	63.70
500	Moderate	T-TMLE 0.1/0.9	0.483	0.350	0.558	-0.000	0.001	0.002	0.072	0.268	0.221	0.825	89.10	68.70
500	Moderate	T-TMLE 0.05/0.95	0.483	0.350	0.558	-0.000	0.001	0.002	0.074	0.272	0.231	0.850	89.70	65.30
500	Moderate	T-TMLE 0.025/0.975	0.483	0.350	0.558	0.000	0.001	0.002	0.075	0.274	0.235	0.855	89.50	64.20
500	Low	Crude	0.476	0.350	0.530	-0.040	-0.001	0.008	0.058	0.237	0.237	0.999	93.40	58.10
500	Low	GC	0.476	0.350	0.530	0.002	0.001	0.001	0.065	0.254	0.243	0.957	93.70	61.40
500	Low	IPW	0.476	0.350	0.530	0.007	0.001	0.000	0.076	0.277	0.275	0.994	95.60	52.60
500	Low	T-IPW 0.1/0.9	0.476	0.350	0.530	-0.019	0.000	0.004	0.053	0.230	0.236	1.027	95.40	58.00
500	Low	T-IPW 0.05/0.95	0.476	0.350	0.530	-0.001	0.001	0.001	0.054	0.233	0.241	1.036	95.60	59.40
500	Low	T-IPW 0.025/0.975	0.476	0.350	0.530	0.009	0.001	-0.001	0.056	0.236	0.245	1.037	95.90	60.80
500	Low	TMLE	0.476	0.350	0.530	0.001	0.001	0.002	0.079	0.281	0.235	0.838	88.50	60.30
500	Low	T-TMLE 0.1/0.9	0.476	0.350	0.530	0.002	0.001	0.001	0.071	0.267	0.221	0.828	88.90	65.00
500	Low	T-TMLE 0.05/0.95	0.476	0.350	0.530	0.002	0.001	0.001	0.074	0.271	0.231	0.853	89.60	62.20
500	Low	T-TMLE 0.025/0.975	0.476	0.350	0.530	0.002	0.001	0.001	0.075	0.274	0.235	0.858	89.40	60.70
500	None	Crude	0.472	0.350	0.516	-0.043	-0.001	0.008	0.058	0.236	0.237	1.001	93.30	54.70
500	None	GC	0.472	0.350	0.516	0.003	0.001	0.001	0.065	0.255	0.244	0.957	94.00	59.60
500	None	IPW	0.472	0.350	0.516	0.007	0.001	0.000	0.076	0.277	0.275	0.994	95.30	50.50
500	None	T-IPW 0.1/0.9	0.472	0.350	0.516	-0.020	0.000	0.004	0.053	0.230	0.236	1.030	95.30	55.90
500	None	T-IPW 0.05/0.95	0.472	0.350	0.516	-0.002	0.001	0.001	0.054	0.232	0.241	1.038	95.70	57.90
500	None	T-IPW 0.025/0.975	0.472	0.350	0.516	0.009	0.001	-0.001	0.056	0.236	0.245	1.039	96.10	58.10
500	None	TMLE	0.472	0.350	0.516	0.001	0.001	0.002	0.079	0.280	0.235	0.839	88.60	57.90
500	None	T-TMLE 0.1/0.9	0.472	0.350	0.516	0.003	0.001	0.001	0.071	0.267	0.221	0.829	89.10	63.00
500	None	T-TMLE 0.05/0.95	0.472	0.350	0.516	0.003	0.001	0.001	0.073	0.271	0.232	0.855	89.60	59.60
500	None	T-TMLE 0.025/0.975	0.472	0.350	0.516	0.003	0.001	0.001	0.075	0.273	0.235	0.860	89.70	58.50

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S3. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 80% with a binary outcome, for sample sizes of 200 and 100 subjects. The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Power
200	High	Crude	0.496	0.350	0.623	-0.042	0.001	0.010	0.146	0.379	0.404	1.064	94.80	36.50
200	High	GC	0.496	0.350	0.623	-0.038	-0.000	0.008	0.155	0.392	0.393	1.003	93.70	35.90
200	High	IPW	0.496	0.350	0.623	0.021	-0.000	-0.003	0.198	0.445	0.429	0.963	95.10	32.50
200	High	T-IPW 0.1/0.9	0.496	0.350	0.623	-0.030	-0.001	0.006	0.142	0.376	0.381	1.012	95.40	33.90
200	High	T-IPW 0.05/0.95	0.496	0.350	0.623	-0.008	-0.001	0.001	0.148	0.385	0.390	1.014	95.90	34.30
200	High	T-IPW 0.025/0.975	0.496	0.350	0.623	0.005	-0.001	-0.001	0.155	0.394	0.398	1.011	96.20	34.40
200	High	TMLE	0.496	0.350	0.623	-0.001	-0.000	0.002	0.212	0.460	0.365	0.793	88.00	41.20
200	High	T-TMLE 0.1/0.9	0.496	0.350	0.623	-0.032	-0.000	0.008	0.176	0.419	0.341	0.815	89.20	41.60
200	High	T-TMLE 0.05/0.95	0.496	0.350	0.623	-0.029	-0.000	0.007	0.189	0.434	0.358	0.825	89.10	39.00
200	High	T-TMLE 0.025/0.975	0.496	0.350	0.623	-0.024	-0.000	0.007	0.195	0.442	0.361	0.816	88.80	39.80
200	Moderate	Crude	0.484	0.350	0.574	-0.045	0.000	0.010	0.147	0.381	0.404	1.061	94.70	31.10
200	Moderate	GC	0.484	0.350	0.574	-0.017	0.001	0.005	0.156	0.394	0.397	1.008	94.50	33.80
200	Moderate	IPW	0.484	0.350	0.574	0.025	0.001	-0.003	0.200	0.447	0.429	0.958	94.70	29.40
200	Moderate	T-IPW 0.1/0.9	0.484	0.350	0.574	-0.026	-0.000	0.006	0.143	0.377	0.381	1.009	95.50	28.40
200	Moderate	T-IPW 0.05/0.95	0.484	0.350	0.574	-0.004	0.000	0.001	0.149	0.386	0.391	1.011	96.00	29.70
200	Moderate	T-IPW 0.025/0.975	0.484	0.350	0.574	0.009	0.001	-0.001	0.156	0.395	0.398	1.008	96.30	29.90
200	Moderate	TMLE	0.484	0.350	0.574	0.004	0.001	0.002	0.212	0.460	0.365	0.793	87.40	37.40
200	Moderate	T-TMLE 0.1/0.9	0.484	0.350	0.574	-0.010	0.001	0.004	0.177	0.421	0.343	0.816	88.80	38.20
200	Moderate	T-TMLE 0.05/0.95	0.484	0.350	0.574	-0.007	0.001	0.004	0.190	0.436	0.360	0.824	88.60	36.10
200	Moderate	T-TMLE 0.025/0.975	0.484	0.350	0.574	-0.004	0.001	0.003	0.198	0.445	0.363	0.815	88.40	36.80
200	Low	Crude	0.478	0.350	0.550	-0.054	-0.002	0.010	0.147	0.379	0.404	1.065	94.90	27.70
200	Low	GC	0.478	0.350	0.550	-0.010	-0.000	0.002	0.154	0.393	0.399	1.015	94.50	32.50
200	Low	IPW	0.478	0.350	0.550	0.021	-0.000	-0.003	0.199	0.446	0.429	0.962	94.70	27.40
200	Low	T-IPW 0.1/0.9	0.478	0.350	0.550	-0.031	-0.001	0.006	0.142	0.376	0.381	1.013	95.90	25.50
200	Low	T-IPW 0.05/0.95	0.478	0.350	0.550	-0.008	-0.001	0.001	0.148	0.385	0.391	1.016	96.30	26.50
200	Low	T-IPW 0.025/0.975	0.478	0.350	0.550	0.005	-0.000	-0.001	0.155	0.393	0.398	1.013	96.40	27.40
200	Low	TMLE	0.478	0.350	0.550	0.002	-0.000	0.001	0.210	0.459	0.366	0.797	87.80	36.50
200	Low	T-TMLE 0.1/0.9	0.478	0.350	0.550	-0.005	-0.000	0.002	0.176	0.419	0.345	0.822	88.50	37.30
200	Low	T-TMLE 0.05/0.95	0.478	0.350	0.550	-0.002	-0.000	0.002	0.189	0.435	0.361	0.830	88.90	35.40
200	Low	T-TMLE 0.025/0.975	0.478	0.350	0.550	0.000	-0.000	0.001	0.196	0.443	0.364	0.820	88.50	36.00
200	None	Crude	0.474	0.350	0.535	-0.058	-0.003	0.010	0.146	0.378	0.404	1.070	95.20	25.70
200	None	GC	0.474	0.350	0.535	-0.006	-0.000	0.001	0.154	0.392	0.399	1.018	94.50	31.40
200	None	IPW	0.474	0.350	0.535	0.020	-0.000	-0.003	0.198	0.445	0.429	0.963	94.80	25.10
200	None	T-IPW 0.1/0.9	0.474	0.350	0.535	-0.032	-0.002	0.006	0.141	0.375	0.381	1.016	96.20	23.50
200	None	T-IPW 0.05/0.95	0.474	0.350	0.535	-0.009	-0.001	0.001	0.147	0.384	0.391	1.018	96.30	25.30
200	None	T-IPW 0.025/0.975	0.474	0.350	0.535	0.004	-0.001	-0.001	0.154	0.392	0.398	1.015	96.60	26.00
200	None	TMLE	0.474	0.350	0.535	0.002	-0.000	0.001	0.210	0.459	0.366	0.798	88.40	35.90
200	None	T-TMLE 0.1/0.9	0.474	0.350	0.535	-0.000	-0.000	0.001	0.175	0.419	0.345	0.823	88.80	36.00
200	None	T-TMLE 0.05/0.95	0.474	0.350	0.535	0.002	-0.000	0.001	0.189	0.434	0.361	0.832	89.10	34.60
200	None	T-TMLE 0.025/0.975	0.474	0.350	0.535	0.004	-0.000	0.000	0.196	0.443	0.364	0.822	88.90	35.20
100	High	Crude	0.494	0.352	0.627	-0.010	0.002	0.004	0.340	0.583	1.207	2.069	94.20	22.70
100	High	GC	0.494	0.352	0.627	-0.038	0.001	0.009	0.328	0.572	1.109	1.939	94.00	21.60
100	High	IPW	0.494	0.352	0.627	0.069	0.001	-0.009	0.481	0.690	0.624	0.903	94.40	20.00
100	High	T-IPW 0.1/0.9	0.494	0.352	0.627	-0.000	-0.001	0.000	0.330	0.575	0.558	0.970	96.50	18.10
100	High	T-IPW 0.05/0.95	0.494	0.352	0.627	0.027	-0.000	-0.005	0.348	0.590	0.575	0.975	95.90	18.10
100	High	T-IPW 0.025/0.975	0.494	0.352	0.627	0.044	0.000	-0.007	0.367	0.605	0.588	0.972	95.70	18.10
100	High	TMLE	0.494	0.352	0.627	0.029	0.001	0.001	0.512	0.715	0.507	0.708	83.90	30.70
100	High	T-TMLE 0.1/0.9	0.494	0.352	0.627	-0.027	0.001	0.009	0.377	0.614	0.463	0.754	87.10	29.30
100	High	T-TMLE 0.05/0.95	0.494	0.352	0.627	-0.016	0.001	0.008	0.422	0.650	0.487	0.749	86.20	28.20
100	High	T-TMLE 0.025/0.975	0.494	0.352	0.627	-0.006	0.001	0.007	0.464	0.681	0.495	0.727	85.00	30.20
100	Moderate	Crude	0.483	0.352	0.580	-0.016	0.000	0.004	0.336	0.580	1.207	2.081	94.30	20.30
100	Moderate	GC	0.483	0.352	0.580	-0.003	0.001	0.002	0.331	0.575	1.148	1.994	94.30	21.30
100	Moderate	IPW	0.483	0.352	0.580	0.071	0.001	-0.009	0.477	0.687	0.624	0.907	94.40	18.20
100	Moderate	T-IPW 0.1/0.9	0.483	0.352	0.580	0.001	-0.000	0.000	0.328	0.573	0.558	0.975	96.30	15.70
100	Moderate	T-IPW 0.05/0.95	0.483	0.352	0.580	0.029	0.000	-0.005	0.346	0.588	0.575	0.979	96.30	15.50
100	Moderate	T-IPW 0.025/0.975	0.483	0.352	0.580	0.046	0.001	-0.007	0.365	0.603	0.588	0.975	95.70	16.20
100	Moderate	TMLE	0.483	0.352	0.580	0.036	0.001	-0.000	0.510	0.714	0.508	0.712	84.00	27.80
100	Moderate	T-TMLE 0.1/0.9	0.483	0.352	0.580	0.008	0.001	0.002	0.385	0.621	0.468	0.754	87.30	27.60
100	Moderate	T-TMLE 0.05/0.95	0.483	0.352	0.580	0.018	0.001	0.001	0.432	0.658	0.493	0.749	86.50	27.60
100	Moderate	T-TMLE 0.025/0.975	0.483	0.352	0.580	0.023	0.001	0.001	0.470	0.686	0.500	0.729	85.40	28.00
100	Low	Crude	0.476	0.352	0.554	-0.027	-0.002	0.004	0.334	0.577	1.207	2.091	94.40	18.80
100	Low	GC	0.476	0.352	0.554	0.011	-0.000	-0.002	0.330	0.575	1.159	2.015	94.70	20.70
100	Low	IPW	0.476	0.352	0.554	0.066	0.000	-0.009	0.475	0.686	0.624	0.909	94.60	17.30
100	Low	T-IPW 0.1/0.9	0.476	0.352	0.554	-0.005	-0.002	0.000	0.325	0.570	0.558	0.979	96.10	13.90
100	Low	T-IPW 0.05/0.95	0.476	0.352	0.554	0.024	-0.001	-0.005	0.343	0.585	0.575	0.982	96.10	14.20
100	Low	T-IPW 0.025/0.975	0.476	0.352	0.554	0.041	-0.000	-0.007	0.362	0.601	0.588	0.978	96.20	15.00
100	Low	TMLE	0.476	0.352	0.554	0.025	0.000	0.001	0.504	0.710	0.508	0.715	83.70	27.20
100	Low	T-TMLE 0.1/0.9	0.476	0.352	0.554	0.022	0.000	-0.002	0.386	0.622	0.471	0.757	87.00	26.70
100	Low	T-TMLE 0.05/0.95	0.476	0.352	0.554	0.032	0.000	-0.003	0.434	0.658	0.495	0.752	86.30	27.00
100	Low	T-TMLE 0.025/0.975	0.476	0.352	0.554	0.032	0.000	-0.002	0.470	0.685	0.502	0.733	84.90	27.00
100	None	Crude	0.472	0.352	0.539	-0.030	-0.003	0.004	0.334	0.578	1.207	2.089	94.20	17.40
100	None	GC	0.472	0.352	0.539	0.017	-0.000	-0.003	0.331	0.576	1.167	2.027	94.70	20.50
100	None	IPW	0.472	0.352	0.539	0.065	-0.000	-0.009	0.475	0.687	0.624	0.908	94.50	17.00
100	None	T-IPW 0.1/0.9	0.472	0.352	0.539	-0.005	-0.002	0.000	0.326	0.571	0.558	0.978	96.40	13.20
100	None	T-IPW 0.05/0.95	0.472	0.352	0.539	0.023	-0.001	-0.005	0.344	0.586	0.575	0.981	96.00	13.60
100	None	T-IPW 0.025/0.975	0.472	0.352	0.539	0.041	-0.001	-0.007	0.363	0.602	0.588	0.977	95.80	14.00
100	None	TMLE	0.472	0.352	0.539	0.028	0.000	0.000	0.507	0.712	0.508	0.715	83.50	25.80
100	None	T-TMLE 0.1/0.9	0.472	0.352	0.539	0.030	0.000	-0.004	0.388	0.622	0.472	0.759	87.40	25.70
100	None	T-TMLE 0.05/0.95	0.472	0.352	0.539	0.039	0.000	-0.004	0.436	0.659	0.497	0.754	86.70	25.50
100	None	T-TMLE 0.025/0.975	0.472	0.352	0.539	0.039	0.000	-0.003	0.472	0.686	0.504	0.734	85.10	25.90

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S4. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 80% with a binary outcome for sample sizes of 1,000 and 500 subjects. The target parameter was log(OR), and variance was estimated by bootstrapping for each method .

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Power
1000	High	Crude	0.496	0.350	0.606	-0.029	0.001	0.008	0.027	0.161	0.166	1.030	94.90	95.40
1000	High	GC	0.496	0.350	0.606	-0.007	-0.001	0.002	0.031	0.176	0.177	1.006	94.40	93.20
1000	High	IPW	0.496	0.350	0.606	0.002	-0.001	0.000	0.036	0.192	0.190	0.988	93.90	88.90
1000	High	T-IPW 0.1/0.9	0.496	0.350	0.606	-0.021	-0.001	0.004	0.025	0.157	0.162	1.031	95.00	96.10
1000	High	T-IPW 0.05/0.95	0.496	0.350	0.606	-0.005	-0.001	0.001	0.026	0.159	0.164	1.029	95.40	96.60
1000	High	T-IPW 0.025/0.975	0.496	0.350	0.606	0.005	-0.001	-0.001	0.026	0.162	0.166	1.026	95.40	96.60
1000	High	TMLE	0.496	0.350	0.606	-0.001	-0.001	0.001	0.035	0.190	0.188	0.994	94.30	90.90
1000	High	T-TMLE 0.1/0.9	0.496	0.350	0.606	-0.005	-0.001	0.002	0.034	0.185	0.183	0.989	94.30	91.90
1000	High	T-TMLE 0.05/0.95	0.496	0.350	0.606	-0.004	-0.001	0.002	0.035	0.188	0.185	0.985	93.70	90.90
1000	High	T-TMLE 0.025/0.975	0.496	0.350	0.606	-0.004	-0.001	0.002	0.035	0.189	0.186	0.983	93.60	91.00
1000	Moderate	Crude	0.484	0.350	0.560	-0.037	-0.001	0.008	0.027	0.161	0.166	1.030	94.90	90.30
1000	Moderate	GC	0.484	0.350	0.560	-0.005	-0.001	0.002	0.030	0.176	0.177	1.007	94.30	90.10
1000	Moderate	IPW	0.484	0.350	0.560	0.003	-0.001	0.000	0.036	0.192	0.190	0.991	93.80	84.10
1000	Moderate	T-IPW 0.1/0.9	0.484	0.350	0.560	-0.021	-0.001	0.004	0.025	0.157	0.162	1.033	94.80	93.00
1000	Moderate	T-IPW 0.05/0.95	0.484	0.350	0.560	-0.004	-0.001	0.001	0.025	0.159	0.164	1.032	95.00	93.50
1000	Moderate	T-IPW 0.025/0.975	0.484	0.350	0.560	0.006	-0.001	-0.001	0.026	0.161	0.166	1.030	95.00	93.70
1000	Moderate	TMLE	0.484	0.350	0.560	-0.000	-0.001	0.001	0.035	0.189	0.189	0.995	94.20	85.90
1000	Moderate	T-TMLE 0.1/0.9	0.484	0.350	0.560	-0.003	-0.001	0.002	0.033	0.185	0.183	0.991	94.00	87.80
1000	Moderate	T-TMLE 0.05/0.95	0.484	0.350	0.560	-0.002	-0.001	0.001	0.034	0.187	0.185	0.987	93.50	87.10
1000	Moderate	T-TMLE 0.025/0.975	0.484	0.350	0.560	-0.002	-0.001	0.001	0.035	0.189	0.186	0.984	93.80	86.40
1000	Low	Crude	0.477	0.350	0.532	-0.042	-0.002	0.008	0.028	0.161	0.166	1.030	95.10	85.00
1000	Low	GC	0.477	0.350	0.532	-0.004	-0.001	0.002	0.031	0.176	0.177	1.005	94.40	86.60
1000	Low	IPW	0.477	0.350	0.532	0.002	-0.001	0.000	0.036	0.192	0.190	0.990	94.00	80.40
1000	Low	T-IPW 0.1/0.9	0.477	0.350	0.532	-0.022	-0.001	0.004	0.025	0.157	0.162	1.032	94.90	89.30
1000	Low	T-IPW 0.05/0.95	0.477	0.350	0.532	-0.004	-0.001	0.001	0.026	0.159	0.164	1.030	94.80	91.10
1000	Low	T-IPW 0.025/0.975	0.477	0.350	0.532	0.005	-0.001	-0.001	0.026	0.162	0.166	1.027	94.90	90.80
1000	Low	TMLE	0.477	0.350	0.532	-0.001	-0.001	0.001	0.035	0.190	0.189	0.993	94.00	81.30
1000	Low	T-TMLE 0.1/0.9	0.477	0.350	0.532	-0.003	-0.001	0.001	0.034	0.185	0.183	0.989	94.10	83.70
1000	Low	T-TMLE 0.05/0.95	0.477	0.350	0.532	-0.002	-0.001	0.001	0.034	0.188	0.185	0.986	93.90	82.60
1000	Low	T-TMLE 0.025/0.975	0.477	0.350	0.532	-0.001	-0.001	0.001	0.035	0.189	0.186	0.983	93.70	81.80
1000	None	Crude	0.473	0.350	0.516	-0.044	-0.003	0.008	0.028	0.160	0.166	1.033	94.60	83.00
1000	None	GC	0.473	0.350	0.516	-0.003	-0.000	0.001	0.031	0.176	0.177	1.006	94.10	84.60
1000	None	IPW	0.473	0.350	0.516	0.003	-0.000	0.000	0.036	0.192	0.190	0.989	94.20	78.60
1000	None	T-IPW 0.1/0.9	0.473	0.350	0.516	-0.021	-0.001	0.004	0.025	0.157	0.162	1.034	95.10	88.40
1000	None	T-IPW 0.05/0.95	0.473	0.350	0.516	-0.003	-0.001	0.001	0.025	0.159	0.164	1.032	95.00	89.40
1000	None	T-IPW 0.025/0.975	0.473	0.350	0.516	0.006	-0.000	-0.001	0.026	0.161	0.166	1.029	94.90	89.60
1000	None	TMLE	0.473	0.350	0.516	0.000	-0.000	0.001	0.035	0.190	0.189	0.992	94.70	79.10
1000	None	T-TMLE 0.1/0.9	0.473	0.350	0.516	-0.001	-0.000	0.001	0.034	0.185	0.183	0.988	94.10	81.60
1000	None	T-TMLE 0.05/0.95	0.473	0.350	0.516	-0.000	-0.001	0.001	0.034	0.188	0.185	0.985	93.80	80.30
1000	None	T-TMLE 0.025/0.975	0.473	0.350	0.516	0.000	-0.000	0.001	0.035	0.189	0.186	0.982	94.00	80.10
500	High	Crude	0.494	0.350	0.603	-0.021	0.003	0.008	0.059	0.238	0.237	0.993	93.50	71.20
500	High	GC	0.494	0.350	0.603	-0.012	0.001	0.005	0.061	0.257	0.244	0.950	94.00	70.30
500	High	IPW	0.494	0.350	0.603	0.018	0.001	-0.001	0.073	0.278	0.265	0.951	93.90	66.70
500	High	T-IPW 0.1/0.9	0.494	0.350	0.603	-0.013	0.000	0.004	0.055	0.231	0.231	0.999	94.30	74.10
500	High	T-IPW 0.05/0.95	0.494	0.350	0.603	0.006	0.001	0.001	0.056	0.234	0.234	0.999	94.10	76.00
500	High	T-IPW 0.025/0.975	0.494	0.350	0.603	0.016	0.001	-0.001	0.058	0.238	0.238	1.000	94.60	76.00
500	High	TMLE	0.494	0.350	0.603	0.005	0.001	0.002	0.072	0.283	0.273	0.967	95.80	63.70
500	High	T-TMLE 0.1/0.9	0.494	0.350	0.603	-0.011	0.001	0.005	0.067	0.269	0.252	0.939	93.60	67.50
500	High	T-TMLE 0.05/0.95	0.494	0.350	0.603	-0.010	0.001	0.005	0.069	0.273	0.257	0.939	94.20	66.40
500	High	T-TMLE 0.025/0.975	0.494	0.350	0.603	-0.009	0.001	0.005	0.070	0.276	0.260	0.943	94.00	65.20
500	Moderate	Crude	0.483	0.350	0.558	-0.028	0.001	0.008	0.059	0.237	0.237	0.997	93.70	62.90
500	Moderate	GC	0.483	0.350	0.558	-0.002	0.001	0.003	0.060	0.255	0.243	0.954	94.30	65.00
500	Moderate	IPW	0.483	0.350	0.558	0.019	0.001	-0.001	0.072	0.278	0.265	0.954	94.10	61.20
500	Moderate	T-IPW 0.1/0.9	0.483	0.350	0.558	-0.013	0.000	0.004	0.055	0.230	0.231	1.004	94.20	67.70
500	Moderate	T-IPW 0.05/0.95	0.483	0.350	0.558	0.006	0.001	0.001	0.056	0.233	0.234	1.004	94.30	70.20
500	Moderate	T-IPW 0.025/0.975	0.483	0.350	0.558	0.017	0.001	-0.001	0.057	0.236	0.238	1.005	94.00	70.30
500	Moderate	TMLE	0.483	0.350	0.558	0.006	0.001	0.002	0.072	0.282	0.273	0.971	95.40	58.10
500	Moderate	T-TMLE 0.1/0.9	0.483	0.350	0.558	0.000	0.001	0.003	0.066	0.268	0.252	0.943	94.30	63.10
500	Moderate	T-TMLE 0.05/0.95	0.483	0.350	0.558	0.001	0.001	0.003	0.069	0.272	0.257	0.944	93.90	61.60
500	Moderate	T-TMLE 0.025/0.975	0.483	0.350	0.558	0.002	0.001	0.003	0.070	0.274	0.260	0.948	94.00	61.00
500	Low	Crude	0.476	0.350	0.530	-0.034	-0.001	0.008	0.059	0.237	0.237	0.999	93.40	58.10
500	Low	GC	0.476	0.350	0.530	0.004	0.001	0.002	0.060	0.254	0.243	0.957	93.70	61.40
500	Low	IPW	0.476	0.350	0.530	0.019	0.001	-0.001	0.072	0.277	0.265	0.957	94.30	56.90
500	Low	T-IPW 0.1/0.9	0.476	0.350	0.530	-0.013	0.000	0.004	0.055	0.230	0.231	1.005	94.10	63.90
500	Low	T-IPW 0.05/0.95	0.476	0.350	0.530	0.006	0.001	0.001	0.055	0.233	0.234	1.005	94.60	65.90
500	Low	T-IPW 0.025/0.975	0.476	0.350	0.530	0.017	0.001	-0.001	0.057	0.236	0.238	1.007	94.30	66.00
500	Low	TMLE	0.476	0.350	0.530	0.006	0.001	0.002	0.071	0.281	0.273	0.974	95.60	52.90
500	Low	T-TMLE 0.1/0.9	0.476	0.350	0.530	0.006	0.001	0.002	0.066	0.267	0.253	0.946	94.40	58.80
500	Low	T-TMLE 0.05/0.95	0.476	0.350	0.530	0.006	0.001	0.002	0.069	0.271	0.257	0.947	93.70	58.10
500	Low	T-TMLE 0.025/0.975	0.476	0.350	0.530	0.007	0.001	0.002	0.070	0.274	0.260	0.951	93.90	57.00
500	None	Crude	0.472	0.350	0.516	-0.037	-0.001	0.008	0.059	0.236	0.237	1.001	93.30	54.70
500	None	GC	0.472	0.350	0.516	0.007	0.001	0.001	0.060	0.255	0.244	0.957	94.00	59.60
500	None	IPW	0.472	0.350	0.516	0.019	0.001	-0.001	0.072	0.277	0.265	0.958	94.30	55.50
500	None	T-IPW 0.1/0.9	0.472	0.350	0.516	-0.013	-0.000	0.004	0.054	0.229	0.231	1.007	93.90	61.00
500	None	T-IPW 0.05/0.95	0.472	0.350	0.516	0.006	0.001	0.001	0.055	0.232	0.234	1.008	94.00	62.30
500	None	T-IPW 0.025/0.975	0.472	0.350	0.516	0.016	0.001	-0.001	0.057	0.236	0.238	1.009	94.20	63.20
500	None	TMLE	0.472	0.350	0.516	0.005	0.001	0.002	0.071	0.280	0.273	0.975	95.70	50.10
500	None	T-TMLE 0.1/0.9	0.472	0.350	0.516	0.009	0.001	0.001	0.066	0.267	0.253	0.948	94.40	56.40
500	None	T-TMLE 0.05/0.95	0.472	0.350	0.516	0.009	0.001	0.001	0.068	0.271	0.257	0.949	93.80	55.60
500	None	T-TMLE 0.025/0.975	0.472	0.350	0.516	0.009	0.001	0.001	0.070	0.273	0.260	0.954	93.80	54.00

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S5. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 80% with a binary outcome for sample sizes of 200 and 100 subjects. The target parameter was log(OR), and variance was estimated by bootstrapping for each method .

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Power
200	High	Crude	0.496	0.350	0.623	-0.021	0.001	0.010	0.157	0.379	0.404	1.064	94.80	36.50
200	High	GC	0.496	0.350	0.623	-0.041	-0.000	0.013	0.151	0.392	0.393	1.003	93.70	35.90
200	High	IPW	0.496	0.350	0.623	0.051	-0.000	-0.004	0.205	0.445	0.466	1.047	93.60	35.30
200	High	T-IPW 0.1/0.9	0.496	0.350	0.623	-0.010	-0.002	0.005	0.153	0.376	0.394	1.046	94.50	40.20
200	High	T-IPW 0.05/0.95	0.496	0.350	0.623	0.015	-0.001	0.001	0.159	0.385	0.401	1.041	94.00	41.70
200	High	T-IPW 0.025/0.975	0.496	0.350	0.623	0.030	-0.001	-0.002	0.167	0.394	0.412	1.045	94.00	41.70
200	High	TMLE	0.496	0.350	0.623	0.024	-0.000	0.003	0.194	0.460	0.475	1.033	96.70	31.10
200	High	T-TMLE 0.1/0.9	0.496	0.350	0.623	-0.033	-0.000	0.012	0.170	0.419	0.414	0.989	94.10	34.80
200	High	T-TMLE 0.05/0.95	0.496	0.350	0.623	-0.026	-0.000	0.011	0.182	0.434	0.431	0.995	94.40	32.80
200	High	T-TMLE 0.025/0.975	0.496	0.350	0.623	-0.016	-0.000	0.010	0.189	0.442	0.448	1.014	95.20	31.90
200	Moderate	Crude	0.484	0.350	0.574	-0.025	0.000	0.010	0.158	0.381	0.404	1.061	94.70	31.10
200	Moderate	GC	0.484	0.350	0.574	-0.009	0.001	0.007	0.153	0.394	0.397	1.008	94.50	33.80
200	Moderate	IPW	0.484	0.350	0.574	0.056	0.001	-0.004	0.207	0.447	0.466	1.043	93.50	31.30
200	Moderate	T-IPW 0.1/0.9	0.484	0.350	0.574	-0.006	-0.001	0.005	0.153	0.377	0.394	1.043	94.40	35.40
200	Moderate	T-IPW 0.05/0.95	0.484	0.350	0.574	0.019	0.000	0.001	0.160	0.386	0.401	1.039	94.30	36.80
200	Moderate	T-IPW 0.025/0.975	0.484	0.350	0.574	0.034	0.000	-0.002	0.168	0.395	0.412	1.042	93.80	36.50
200	Moderate	TMLE	0.484	0.350	0.574	0.029	0.001	0.003	0.196	0.460	0.477	1.036	97.00	27.20
200	Moderate	T-TMLE 0.1/0.9	0.484	0.350	0.574	-0.001	0.001	0.006	0.173	0.421	0.419	0.995	94.10	32.10
200	Moderate	T-TMLE 0.05/0.95	0.484	0.350	0.574	0.004	0.001	0.006	0.185	0.436	0.436	1.000	94.70	30.70
200	Moderate	T-TMLE 0.025/0.975	0.484	0.350	0.574	0.011	0.001	0.005	0.193	0.445	0.452	1.017	95.40	28.70
200	Low	Crude	0.478	0.350	0.550	-0.035	-0.002	0.010	0.157	0.379	0.404	1.065	94.90	27.70
200	Low	GC	0.478	0.350	0.550	0.003	-0.000	0.003	0.154	0.393	0.399	1.015	94.50	32.50
200	Low	IPW	0.478	0.350	0.550	0.051	-0.000	-0.004	0.206	0.446	0.466	1.047	93.60	28.70
200	Low	T-IPW 0.1/0.9	0.478	0.350	0.550	-0.011	-0.002	0.005	0.152	0.376	0.394	1.048	94.50	32.00
200	Low	T-IPW 0.05/0.95	0.478	0.350	0.550	0.015	-0.001	0.001	0.159	0.385	0.401	1.044	94.40	34.40
200	Low	T-IPW 0.025/0.975	0.478	0.350	0.550	0.030	-0.001	-0.002	0.166	0.393	0.412	1.047	94.00	34.00
200	Low	TMLE	0.478	0.350	0.550	0.021	-0.000	0.003	0.194	0.459	0.476	1.037	96.80	24.30
200	Low	T-TMLE 0.1/0.9	0.478	0.350	0.550	0.010	-0.000	0.003	0.173	0.419	0.420	1.002	93.90	31.10
200	Low	T-TMLE 0.05/0.95	0.478	0.350	0.550	0.015	-0.000	0.003	0.186	0.435	0.438	1.006	94.90	29.40
200	Low	T-TMLE 0.025/0.975	0.478	0.350	0.550	0.018	-0.000	0.003	0.192	0.443	0.453	1.021	95.50	27.50
200	None	Crude	0.474	0.350	0.535	-0.038	-0.003	0.010	0.156	0.378	0.404	1.070	95.20	25.70
200	None	GC	0.474	0.350	0.535	0.011	-0.000	0.002	0.154	0.392	0.399	1.018	94.50	31.40
200	None	IPW	0.474	0.350	0.535	0.051	-0.000	-0.004	0.205	0.445	0.467	1.049	93.70	26.90
200	None	T-IPW 0.1/0.9	0.474	0.350	0.535	-0.012	-0.002	0.005	0.151	0.375	0.394	1.051	95.00	29.90
200	None	T-IPW 0.05/0.95	0.474	0.350	0.535	0.014	-0.001	0.001	0.158	0.384	0.401	1.046	94.30	32.90
200	None	T-IPW 0.025/0.975	0.474	0.350	0.535	0.029	-0.001	-0.002	0.165	0.392	0.412	1.050	94.10	32.50
200	None	TMLE	0.474	0.350	0.535	0.023	-0.000	0.003	0.194	0.459	0.476	1.039	96.40	23.10
200	None	T-TMLE 0.1/0.9	0.474	0.350	0.535	0.018	-0.000	0.001	0.173	0.419	0.421	1.004	94.10	30.20
200	None	T-TMLE 0.05/0.95	0.474	0.350	0.535	0.021	-0.001	0.001	0.185	0.435	0.438	1.008	94.70	28.40
200	None	T-TMLE 0.025/0.975	0.474	0.350	0.535	0.026	-0.001	0.001	0.193	0.443	0.453	1.024	95.40	26.50
100	High	Crude	0.494	0.352	0.627	0.185	0.002	0.005	1.120	0.583	1.207	2.069	94.20	22.70
100	High	GC	0.494	0.352	0.627	0.107	0.001	0.013	0.905	0.572	1.109	1.939	94.00	21.60
100	High	IPW	0.494	0.352	0.627	0.294	0.001	-0.011	1.256	0.690	1.315	1.905	95.30	18.40
100	High	T-IPW 0.1/0.9	0.494	0.352	0.627	0.194	-0.001	-0.000	1.091	0.575	1.188	2.065	94.10	24.10
100	High	T-IPW 0.05/0.95	0.494	0.352	0.627	0.224	-0.000	-0.005	1.104	0.590	1.191	2.020	94.10	25.20
100	High	T-IPW 0.025/0.975	0.494	0.352	0.627	0.246	0.000	-0.008	1.129	0.605	1.206	1.995	94.00	25.20
100	High	TMLE	0.494	0.352	0.627	0.162	0.001	0.002	0.929	0.715	1.282	1.792	96.80	11.90
100	High	T-TMLE 0.1/0.9	0.494	0.352	0.627	0.110	0.001	0.012	0.949	0.614	1.220	1.988	93.80	20.70
100	High	T-TMLE 0.05/0.95	0.494	0.352	0.627	0.089	0.001	0.011	0.857	0.650	1.212	1.864	95.10	18.10
100	High	T-TMLE 0.025/0.975	0.494	0.352	0.627	0.110	0.001	0.011	0.901	0.681	1.246	1.829	96.50	14.70
100	Moderate	Crude	0.483	0.352	0.580	0.178	0.000	0.005	1.111	0.580	1.207	2.081	94.30	20.30
100	Moderate	GC	0.483	0.352	0.580	0.163	0.001	0.005	1.008	0.575	1.148	1.994	94.30	21.30
100	Moderate	IPW	0.483	0.352	0.580	0.296	0.001	-0.011	1.250	0.687	1.315	1.913	95.60	16.10
100	Moderate	T-IPW 0.1/0.9	0.483	0.352	0.580	0.194	-0.000	-0.000	1.085	0.573	1.188	2.074	94.50	22.10
100	Moderate	T-IPW 0.05/0.95	0.483	0.352	0.580	0.225	0.000	-0.005	1.099	0.588	1.191	2.027	94.00	22.70
100	Moderate	T-IPW 0.025/0.975	0.483	0.352	0.580	0.248	0.001	-0.008	1.123	0.603	1.206	2.002	94.40	22.10
100	Moderate	TMLE	0.483	0.352	0.580	0.173	0.002	0.002	0.973	0.714	1.287	1.803	97.10	9.80
100	Moderate	T-TMLE 0.1/0.9	0.483	0.352	0.580	0.158	0.002	0.004	0.990	0.621	1.231	1.984	94.50	19.70
100	Moderate	T-TMLE 0.05/0.95	0.483	0.352	0.580	0.138	0.002	0.004	0.942	0.658	1.224	1.861	95.10	17.00
100	Moderate	T-TMLE 0.025/0.975	0.483	0.352	0.580	0.154	0.002	0.004	0.973	0.686	1.256	1.832	96.00	13.80
100	Low	Crude	0.476	0.352	0.554	0.166	-0.002	0.005	1.103	0.577	1.207	2.091	94.40	18.80
100	Low	GC	0.476	0.352	0.554	0.185	0.000	0.000	1.038	0.575	1.159	2.015	94.70	20.70
100	Low	IPW	0.476	0.352	0.554	0.290	0.000	-0.011	1.244	0.686	1.315	1.917	95.30	15.10
100	Low	T-IPW 0.1/0.9	0.476	0.352	0.554	0.188	-0.002	-0.000	1.079	0.570	1.188	2.083	94.40	20.70
100	Low	T-IPW 0.05/0.95	0.476	0.352	0.554	0.219	-0.001	-0.005	1.092	0.585	1.191	2.035	94.40	21.00
100	Low	T-IPW 0.025/0.975	0.476	0.352	0.554	0.242	-0.001	-0.008	1.117	0.601	1.206	2.008	94.60	20.00
100	Low	TMLE	0.476	0.352	0.554	0.169	0.000	0.002	0.909	0.710	1.289	1.817	97.20	9.20
100	Low	T-TMLE 0.1/0.9	0.476	0.352	0.554	0.177	0.000	-0.001	0.990	0.622	1.227	1.974	94.30	19.20
100	Low	T-TMLE 0.05/0.95	0.476	0.352	0.554	0.152	0.000	-0.001	0.911	0.658	1.220	1.853	95.40	16.40
100	Low	T-TMLE 0.025/0.975	0.476	0.352	0.554	0.168	0.000	0.000	0.920	0.685	1.258	1.835	96.10	12.90
100	None	Crude	0.472	0.352	0.539	0.163	-0.003	0.005	1.104	0.578	1.207	2.089	94.20	17.40
100	None	GC	0.472	0.352	0.539	0.196	-0.000	-0.002	1.055	0.576	1.167	2.027	94.70	20.50
100	None	IPW	0.472	0.352	0.539	0.290	-0.000	-0.011	1.246	0.687	1.315	1.916	95.40	14.90
100	None	T-IPW 0.1/0.9	0.472	0.352	0.539	0.187	-0.002	-0.000	1.080	0.571	1.188	2.080	94.00	20.40
100	None	T-IPW 0.05/0.95	0.472	0.352	0.539	0.218	-0.001	-0.005	1.094	0.586	1.191	2.032	94.40	21.00
100	None	T-IPW 0.025/0.975	0.472	0.352	0.539	0.241	-0.001	-0.008	1.118	0.602	1.207	2.005	94.80	19.40
100	None	TMLE	0.472	0.352	0.539	0.173	0.000	0.002	0.960	0.712	1.303	1.831	97.10	9.20
100	None	T-TMLE 0.1/0.9	0.472	0.352	0.539	0.196	0.000	-0.003	1.006	0.622	1.235	1.985	94.20	18.80
100	None	T-TMLE 0.05/0.95	0.472	0.352	0.539	0.159	-0.001	-0.003	0.918	0.659	1.222	1.853	95.50	16.10
100	None	T-TMLE 0.025/0.975	0.472	0.352	0.539	0.187	-0.001	-0.001	0.981	0.686	1.270	1.850	96.10	13.00

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S6. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 50% with a binary outcome for sample sizes of 1,000 and 500 subjects. The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Power
1000	High	Crude	0.496	0.350	0.605	0.022	0.009	0.003	0.017	0.130	0.130	0.997	94.30	99.80
1000	High	GC	0.496	0.350	0.605	-0.003	-0.000	0.000	0.016	0.128	0.132	1.024	95.00	99.60
1000	High	IPW	0.496	0.350	0.605	-0.003	-0.000	0.001	0.017	0.130	0.140	1.072	95.80	99.30
1000	High	T-IPW 0.1/0.9	0.496	0.350	0.605	0.009	0.001	-0.001	0.016	0.126	0.133	1.056	95.70	99.70
1000	High	T-IPW 0.05/0.95	0.496	0.350	0.605	0.005	-0.000	-0.001	0.016	0.127	0.134	1.059	95.40	99.70
1000	High	T-IPW 0.025/0.975	0.496	0.350	0.605	0.003	-0.000	-0.001	0.016	0.127	0.135	1.060	95.60	99.70
1000	High	TMLE	0.496	0.350	0.605	-0.003	-0.000	0.001	0.017	0.130	0.132	1.016	94.40	99.50
1000	High	T-TMLE 0.1/0.9	0.496	0.350	0.605	-0.003	-0.000	0.001	0.017	0.130	0.131	1.009	94.20	99.50
1000	High	T-TMLE 0.05/0.95	0.496	0.350	0.605	-0.003	-0.000	0.001	0.017	0.130	0.132	1.016	94.40	99.50
1000	High	T-TMLE 0.025/0.975	0.496	0.350	0.605	-0.003	-0.000	0.001	0.017	0.130	0.132	1.016	94.40	99.50
1000	Moderate	Crude	0.485	0.350	0.559	-0.004	0.003	0.003	0.017	0.130	0.130	1.001	94.70	99.20
1000	Moderate	GC	0.485	0.350	0.559	-0.002	0.000	0.000	0.016	0.128	0.132	1.029	95.30	99.10
1000	Moderate	IPW	0.485	0.350	0.559	-0.002	0.000	0.001	0.017	0.130	0.140	1.075	96.10	98.50
1000	Moderate	T-IPW 0.1/0.9	0.485	0.350	0.559	0.005	-0.000	-0.001	0.016	0.126	0.133	1.059	95.80	99.30
1000	Moderate	T-IPW 0.05/0.95	0.485	0.350	0.559	0.004	-0.001	-0.001	0.016	0.127	0.134	1.061	95.80	99.10
1000	Moderate	T-IPW 0.025/0.975	0.485	0.350	0.559	0.004	-0.000	-0.001	0.016	0.127	0.135	1.062	95.40	99.00
1000	Moderate	TMLE	0.485	0.350	0.559	-0.002	0.000	0.001	0.017	0.129	0.132	1.020	94.90	98.90
1000	Moderate	T-TMLE 0.1/0.9	0.485	0.350	0.559	-0.002	0.000	0.001	0.017	0.130	0.131	1.013	95.00	98.90
1000	Moderate	T-TMLE 0.05/0.95	0.485	0.350	0.559	-0.002	0.000	0.001	0.017	0.129	0.132	1.020	94.90	98.90
1000	Moderate	T-TMLE 0.025/0.975	0.485	0.350	0.559	-0.002	0.000	0.001	0.017	0.129	0.132	1.020	94.90	98.90
1000	Low	Crude	0.477	0.350	0.530	-0.021	-0.002	0.003	0.018	0.131	0.130	0.995	93.80	97.20
1000	Low	GC	0.477	0.350	0.530	-0.002	0.000	0.000	0.016	0.128	0.132	1.028	95.30	98.60
1000	Low	IPW	0.477	0.350	0.530	-0.003	0.000	0.001	0.017	0.130	0.140	1.071	96.00	97.40
1000	Low	T-IPW 0.1/0.9	0.477	0.350	0.530	0.001	-0.001	-0.001	0.016	0.127	0.133	1.052	95.80	98.10
1000	Low	T-IPW 0.05/0.95	0.477	0.350	0.530	0.002	-0.001	-0.001	0.016	0.127	0.134	1.054	95.60	98.10
1000	Low	T-IPW 0.025/0.975	0.477	0.350	0.530	0.003	-0.000	-0.001	0.016	0.128	0.135	1.056	95.50	98.10
1000	Low	TMLE	0.477	0.350	0.530	-0.002	0.000	0.001	0.017	0.130	0.132	1.017	94.70	98.10
1000	Low	T-TMLE 0.1/0.9	0.477	0.350	0.530	-0.002	0.000	0.001	0.017	0.130	0.131	1.010	94.70	98.10
1000	Low	T-TMLE 0.05/0.95	0.477	0.350	0.530	-0.002	0.000	0.001	0.017	0.130	0.132	1.017	94.70	98.10
1000	Low	T-TMLE 0.025/0.975	0.477	0.350	0.530	-0.002	0.000	0.001	0.017	0.130	0.132	1.017	94.70	98.10
1000	None	Crude	0.474	0.350	0.515	-0.029	-0.003	0.003	0.018	0.131	0.130	0.993	93.40	95.80
1000	None	GC	0.474	0.350	0.515	-0.001	0.000	0.000	0.017	0.129	0.132	1.023	95.40	97.70
1000	None	IPW	0.474	0.350	0.515	-0.002	0.000	0.001	0.017	0.131	0.140	1.068	95.80	96.70
1000	None	T-IPW 0.1/0.9	0.474	0.350	0.515	-0.000	-0.002	-0.001	0.016	0.127	0.133	1.049	95.40	97.50
1000	None	T-IPW 0.05/0.95	0.474	0.350	0.515	0.003	-0.001	-0.001	0.016	0.128	0.134	1.052	95.30	97.60
1000	None	T-IPW 0.025/0.975	0.474	0.350	0.515	0.004	-0.000	-0.001	0.016	0.128	0.135	1.054	95.40	97.40
1000	None	TMLE	0.474	0.350	0.515	-0.001	0.000	0.001	0.017	0.130	0.132	1.014	94.80	97.40
1000	None	T-TMLE 0.1/0.9	0.474	0.350	0.515	-0.001	0.000	0.001	0.017	0.131	0.132	1.008	94.60	97.40
1000	None	T-TMLE 0.05/0.95	0.474	0.350	0.515	-0.001	0.000	0.001	0.017	0.130	0.132	1.014	94.80	97.40
1000	None	T-TMLE 0.025/0.975	0.474	0.350	0.515	-0.001	0.000	0.001	0.017	0.130	0.132	1.014	94.80	97.40
500	High	Crude	0.495	0.349	0.606	0.018	0.009	0.004	0.037	0.191	0.185	0.967	93.90	91.60
500	High	GC	0.495	0.349	0.606	-0.007	-0.000	0.001	0.037	0.193	0.189	0.980	94.00	88.70
500	High	IPW	0.495	0.349	0.606	-0.007	0.000	0.002	0.039	0.197	0.199	1.013	95.40	85.10
500	High	T-IPW 0.1/0.9	0.495	0.349	0.606	0.003	0.001	-0.000	0.035	0.188	0.189	1.006	94.80	90.30
500	High	T-IPW 0.05/0.95	0.495	0.349	0.606	-0.001	-0.000	-0.000	0.035	0.188	0.191	1.011	94.50	89.00
500	High	T-IPW 0.025/0.975	0.495	0.349	0.606	-0.002	-0.000	0.000	0.036	0.189	0.192	1.015	95.20	88.90
500	High	TMLE	0.495	0.349	0.606	-0.007	0.000	0.002	0.039	0.197	0.186	0.948	93.20	88.20
500	High	T-TMLE 0.1/0.9	0.495	0.349	0.606	-0.008	0.000	0.002	0.038	0.196	0.185	0.946	93.10	88.60
500	High	T-TMLE 0.05/0.95	0.495	0.349	0.606	-0.007	0.000	0.002	0.039	0.197	0.186	0.948	93.20	88.20
500	High	T-TMLE 0.025/0.975	0.495	0.349	0.606	-0.007	0.000	0.002	0.039	0.197	0.186	0.948	93.20	88.20
500	Moderate	Crude	0.483	0.349	0.560	-0.009	0.003	0.004	0.036	0.189	0.185	0.976	94.70	83.80
500	Moderate	GC	0.483	0.349	0.560	-0.007	-0.000	0.001	0.037	0.191	0.189	0.989	94.00	83.70
500	Moderate	IPW	0.483	0.349	0.560	-0.007	0.000	0.002	0.038	0.195	0.199	1.021	95.50	78.80
500	Moderate	T-IPW 0.1/0.9	0.483	0.349	0.560	-0.002	-0.001	-0.000	0.035	0.186	0.189	1.014	95.50	83.60
500	Moderate	T-IPW 0.05/0.95	0.483	0.349	0.560	-0.002	-0.001	-0.000	0.035	0.187	0.191	1.018	95.30	83.10
500	Moderate	T-IPW 0.025/0.975	0.483	0.349	0.560	-0.002	-0.000	0.000	0.035	0.188	0.192	1.022	95.60	83.00
500	Moderate	TMLE	0.483	0.349	0.560	-0.007	0.000	0.002	0.038	0.195	0.187	0.957	93.10	82.80
500	Moderate	T-TMLE 0.1/0.9	0.483	0.349	0.560	-0.007	0.000	0.002	0.038	0.194	0.185	0.954	93.10	83.60
500	Moderate	T-TMLE 0.05/0.95	0.483	0.349	0.560	-0.007	0.000	0.002	0.038	0.195	0.187	0.956	93.10	82.80
500	Moderate	T-TMLE 0.025/0.975	0.483	0.349	0.560	-0.007	0.000	0.002	0.038	0.195	0.187	0.957	93.10	82.80
500	Low	Crude	0.476	0.349	0.533	-0.025	-0.001	0.004	0.035	0.186	0.185	0.992	94.70	80.10
500	Low	GC	0.476	0.349	0.533	-0.006	-0.000	0.001	0.036	0.190	0.189	0.995	94.10	80.20
500	Low	IPW	0.476	0.349	0.533	-0.007	0.000	0.002	0.038	0.194	0.199	1.026	95.10	76.60
500	Low	T-IPW 0.1/0.9	0.476	0.349	0.533	-0.005	-0.001	-0.000	0.034	0.185	0.189	1.023	95.50	80.50
500	Low	T-IPW 0.05/0.95	0.476	0.349	0.533	-0.003	-0.001	-0.000	0.035	0.186	0.191	1.025	95.70	80.50
500	Low	T-IPW 0.025/0.975	0.476	0.349	0.533	-0.002	-0.000	0.000	0.035	0.187	0.192	1.029	95.80	80.00
500	Low	TMLE	0.476	0.349	0.533	-0.007	0.000	0.002	0.038	0.194	0.187	0.963	93.60	79.20
500	Low	T-TMLE 0.1/0.9	0.476	0.349	0.533	-0.007	0.000	0.002	0.037	0.193	0.185	0.960	93.70	79.50
500	Low	T-TMLE 0.05/0.95	0.476	0.349	0.533	-0.007	0.000	0.002	0.038	0.194	0.187	0.963	93.60	79.30
500	Low	T-TMLE 0.025/0.975	0.476	0.349	0.533	-0.007	0.000	0.002	0.038	0.194	0.187	0.963	93.60	79.20
500	None	Crude	0.473	0.349	0.518	-0.034	-0.004	0.004	0.036	0.186	0.185	0.992	94.60	76.40
500	None	GC	0.473	0.349	0.518	-0.007	-0.000	0.001	0.036	0.190	0.189	0.995	94.50	77.20
500	None	IPW	0.473	0.349	0.518	-0.007	0.000	0.002	0.038	0.194	0.199	1.027	95.60	73.50
500	None	T-IPW 0.1/0.9	0.473	0.349	0.518	-0.007	-0.002	-0.000	0.034	0.185	0.189	1.024	95.70	77.80
500	None	T-IPW 0.05/0.95	0.473	0.349	0.518	-0.004	-0.001	-0.000	0.035	0.186	0.191	1.026	95.70	77.40
500	None	T-IPW 0.025/0.975	0.473	0.349	0.518	-0.003	-0.001	0.000	0.035	0.187	0.192	1.029	96.00	76.50
500	None	TMLE	0.473	0.349	0.518	-0.007	0.000	0.002	0.038	0.194	0.187	0.964	93.80	76.90
500	None	T-TMLE 0.1/0.9	0.473	0.349	0.518	-0.007	0.000	0.002	0.037	0.193	0.186	0.961	93.90	77.30
500	None	T-TMLE 0.05/0.95	0.473	0.349	0.518	-0.007	0.000	0.002	0.038	0.194	0.187	0.963	93.80	76.90
500	None	T-TMLE 0.025/0.975	0.473	0.349	0.518	-0.007	0.000	0.002	0.038	0.194	0.187	0.964	93.80	76.90

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S7. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 50% with a binary outcome for sample sizes of 200 and 100 subjects. The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Power
200	High	Crude	0.497	0.349	0.617	0.002	0.005	0.004	0.089	0.298	0.296	0.993	94.10	57.70
200	High	GC	0.497	0.349	0.617	-0.024	-0.005	0.001	0.099	0.314	0.305	0.972	94.00	50.80
200	High	IPW	0.497	0.349	0.617	-0.020	-0.006	-0.000	0.099	0.315	0.320	1.016	95.50	46.50
200	High	T-IPW 0.1/0.9	0.497	0.349	0.617	-0.015	-0.005	-0.001	0.090	0.299	0.301	1.008	95.40	51.80
200	High	T-IPW 0.05/0.95	0.497	0.349	0.617	-0.018	-0.006	-0.001	0.091	0.301	0.305	1.012	95.30	49.70
200	High	T-IPW 0.025/0.975	0.497	0.349	0.617	-0.017	-0.006	-0.001	0.092	0.303	0.307	1.015	95.20	48.90
200	High	TMLE	0.497	0.349	0.617	-0.021	-0.005	0.000	0.100	0.315	0.291	0.922	92.90	51.90
200	High	T-TMLE 0.1/0.9	0.497	0.349	0.617	-0.023	-0.005	0.001	0.100	0.315	0.290	0.919	92.90	52.10
200	High	T-TMLE 0.05/0.95	0.497	0.349	0.617	-0.023	-0.005	0.001	0.100	0.315	0.291	0.923	92.90	51.60
200	High	T-TMLE 0.025/0.975	0.497	0.349	0.617	-0.022	-0.005	0.000	0.100	0.315	0.291	0.923	92.80	51.60
200	Moderate	Crude	0.485	0.349	0.570	-0.021	-0.001	0.004	0.089	0.298	0.296	0.995	94.60	48.20
200	Moderate	GC	0.485	0.349	0.570	-0.021	-0.005	0.001	0.099	0.313	0.305	0.974	94.10	45.00
200	Moderate	IPW	0.485	0.349	0.570	-0.017	-0.005	-0.000	0.099	0.315	0.320	1.016	95.30	39.30
200	Moderate	T-IPW 0.1/0.9	0.485	0.349	0.570	-0.018	-0.006	-0.001	0.090	0.299	0.301	1.008	95.10	44.60
200	Moderate	T-IPW 0.05/0.95	0.485	0.349	0.570	-0.017	-0.006	-0.001	0.091	0.301	0.305	1.011	95.40	43.90
200	Moderate	T-IPW 0.025/0.975	0.485	0.349	0.570	-0.015	-0.005	-0.001	0.092	0.303	0.307	1.014	95.30	43.20
200	Moderate	TMLE	0.485	0.349	0.570	-0.019	-0.005	0.000	0.100	0.316	0.291	0.924	92.90	47.40
200	Moderate	T-TMLE 0.1/0.9	0.485	0.349	0.570	-0.019	-0.005	0.000	0.100	0.315	0.290	0.921	92.80	47.60
200	Moderate	T-TMLE 0.05/0.95	0.485	0.349	0.570	-0.019	-0.005	0.000	0.100	0.315	0.291	0.925	92.80	47.30
200	Moderate	T-TMLE 0.025/0.975	0.485	0.349	0.570	-0.019	-0.005	0.000	0.100	0.315	0.291	0.925	92.90	47.30
200	Low	Crude	0.479	0.349	0.545	-0.041	-0.006	0.004	0.090	0.297	0.296	0.998	94.30	40.60
200	Low	GC	0.479	0.349	0.545	-0.024	-0.005	0.001	0.099	0.314	0.305	0.973	94.00	41.00
200	Low	IPW	0.479	0.349	0.545	-0.020	-0.006	-0.000	0.099	0.315	0.320	1.017	95.60	36.00
200	Low	T-IPW 0.1/0.9	0.479	0.349	0.545	-0.024	-0.007	-0.001	0.090	0.299	0.302	1.008	95.20	41.20
200	Low	T-IPW 0.05/0.95	0.479	0.349	0.545	-0.021	-0.007	-0.001	0.091	0.302	0.305	1.011	95.20	41.10
200	Low	T-IPW 0.025/0.975	0.479	0.349	0.545	-0.018	-0.006	-0.001	0.092	0.303	0.307	1.013	95.10	40.60
200	Low	TMLE	0.479	0.349	0.545	-0.022	-0.005	0.000	0.100	0.316	0.292	0.924	92.90	43.50
200	Low	T-TMLE 0.1/0.9	0.479	0.349	0.545	-0.022	-0.005	0.000	0.100	0.315	0.291	0.922	92.90	43.80
200	Low	T-TMLE 0.05/0.95	0.479	0.349	0.545	-0.022	-0.005	0.000	0.100	0.315	0.292	0.926	92.90	43.40
200	Low	T-TMLE 0.025/0.975	0.479	0.349	0.545	-0.022	-0.005	0.000	0.100	0.315	0.292	0.926	92.90	43.30
200	None	Crude	0.475	0.349	0.528	-0.049	-0.008	0.004	0.090	0.297	0.297	1.000	94.00	38.10
200	None	GC	0.475	0.349	0.528	-0.022	-0.005	0.001	0.099	0.314	0.306	0.972	94.10	39.00
200	None	IPW	0.475	0.349	0.528	-0.018	-0.005	-0.000	0.099	0.315	0.320	1.016	95.60	34.40
200	None	T-IPW 0.1/0.9	0.475	0.349	0.528	-0.024	-0.007	-0.001	0.090	0.299	0.302	1.008	95.20	38.70
200	None	T-IPW 0.05/0.95	0.475	0.349	0.528	-0.020	-0.007	-0.001	0.091	0.302	0.305	1.011	95.20	39.00
200	None	T-IPW 0.025/0.975	0.475	0.349	0.528	-0.017	-0.006	-0.001	0.092	0.303	0.307	1.013	95.20	38.70
200	None	TMLE	0.475	0.349	0.528	-0.020	-0.005	0.000	0.100	0.316	0.292	0.924	92.90	41.30
200	None	T-TMLE 0.1/0.9	0.475	0.349	0.528	-0.020	-0.005	0.000	0.100	0.315	0.291	0.922	92.90	41.30
200	None	T-TMLE 0.05/0.95	0.475	0.349	0.528	-0.020	-0.005	0.000	0.100	0.315	0.292	0.926	92.90	41.10
200	None	T-TMLE 0.025/0.975	0.475	0.349	0.528	-0.020	-0.005	0.000	0.100	0.315	0.292	0.926	92.90	41.10
100	High	Crude	0.492	0.352	0.593	0.039	0.011	0.001	0.192	0.436	0.431	0.988	94.40	35.30
100	High	GC	0.492	0.352	0.593	-0.009	0.000	0.003	0.197	0.444	0.444	0.999	94.10	25.70
100	High	IPW	0.492	0.352	0.593	0.020	0.000	-0.003	0.207	0.455	0.458	1.007	95.90	25.60
100	High	T-IPW 0.1/0.9	0.492	0.352	0.593	0.014	0.000	-0.003	0.179	0.424	0.432	1.020	95.90	28.30
100	High	T-IPW 0.05/0.95	0.492	0.352	0.593	0.012	-0.001	-0.003	0.184	0.428	0.438	1.023	95.80	27.50
100	High	T-IPW 0.025/0.975	0.492	0.352	0.593	0.014	-0.001	-0.004	0.189	0.434	0.443	1.019	95.90	26.90
100	High	TMLE	0.492	0.352	0.593	0.008	0.000	-0.001	0.207	0.455	0.405	0.889	92.00	31.10
100	High	T-TMLE 0.1/0.9	0.492	0.352	0.593	-0.006	0.000	0.002	0.203	0.451	0.402	0.892	91.90	30.90
100	High	T-TMLE 0.05/0.95	0.492	0.352	0.593	-0.004	0.000	0.002	0.204	0.451	0.403	0.892	92.10	30.60
100	High	T-TMLE 0.025/0.975	0.492	0.352	0.593	-0.003	0.000	0.001	0.204	0.452	0.403	0.892	92.20	30.90
100	Moderate	Crude	0.481	0.352	0.548	0.010	0.004	0.001	0.189	0.435	0.431	0.991	94.00	28.20
100	Moderate	GC	0.481	0.352	0.548	-0.004	0.000	0.001	0.197	0.444	0.445	1.002	94.00	23.60
100	Moderate	IPW	0.481	0.352	0.548	0.019	0.000	-0.003	0.207	0.455	0.458	1.007	96.20	22.30
100	Moderate	T-IPW 0.1/0.9	0.481	0.352	0.548	0.007	-0.002	-0.003	0.180	0.424	0.432	1.020	96.20	24.10
100	Moderate	T-IPW 0.05/0.95	0.481	0.352	0.548	0.008	-0.002	-0.003	0.184	0.429	0.438	1.022	95.80	23.40
100	Moderate	T-IPW 0.025/0.975	0.481	0.352	0.548	0.012	-0.001	-0.004	0.189	0.435	0.443	1.019	96.00	23.70
100	Moderate	TMLE	0.481	0.352	0.548	0.007	0.000	-0.001	0.205	0.453	0.406	0.896	92.00	28.40
100	Moderate	T-TMLE 0.1/0.9	0.481	0.352	0.548	-0.000	0.000	0.001	0.202	0.450	0.403	0.896	92.00	28.40
100	Moderate	T-TMLE 0.05/0.95	0.481	0.352	0.548	0.001	0.000	0.000	0.203	0.451	0.404	0.897	92.10	28.40
100	Moderate	T-TMLE 0.025/0.975	0.481	0.352	0.548	0.002	0.000	0.000	0.203	0.451	0.404	0.896	92.10	28.40
100	Low	Crude	0.474	0.352	0.519	-0.011	-0.001	0.001	0.189	0.434	0.431	0.992	93.40	23.70
100	Low	GC	0.474	0.352	0.519	-0.002	-0.000	0.001	0.197	0.444	0.446	1.004	94.30	21.70
100	Low	IPW	0.474	0.352	0.519	0.016	-0.000	-0.003	0.207	0.455	0.458	1.007	96.00	20.80
100	Low	T-IPW 0.1/0.9	0.474	0.352	0.519	0.001	-0.003	-0.003	0.180	0.425	0.433	1.019	96.30	20.70
100	Low	T-IPW 0.05/0.95	0.474	0.352	0.519	0.005	-0.003	-0.003	0.184	0.429	0.439	1.022	96.00	21.30
100	Low	T-IPW 0.025/0.975	0.474	0.352	0.519	0.009	-0.002	-0.004	0.189	0.435	0.443	1.019	96.10	21.90
100	Low	TMLE	0.474	0.352	0.519	0.004	-0.000	-0.001	0.205	0.453	0.406	0.897	91.90	26.40
100	Low	T-TMLE 0.1/0.9	0.474	0.352	0.519	0.000	-0.000	-0.000	0.203	0.451	0.404	0.897	92.20	26.40
100	Low	T-TMLE 0.05/0.95	0.474	0.352	0.519	0.001	-0.000	-0.000	0.203	0.451	0.405	0.897	92.00	26.20
100	Low	T-TMLE 0.025/0.975	0.474	0.352	0.519	0.001	-0.000	-0.000	0.204	0.452	0.405	0.897	91.90	26.20
100	None	Crude	0.471	0.352	0.505	-0.022	-0.004	0.001	0.188	0.434	0.431	0.994	93.90	21.90
100	None	GC	0.471	0.352	0.505	-0.002	-0.001	0.000	0.198	0.446	0.447	1.002	94.00	21.30
100	None	IPW	0.471	0.352	0.505	0.014	-0.001	-0.003	0.207	0.455	0.458	1.007	95.60	20.10
100	None	T-IPW 0.1/0.9	0.471	0.352	0.505	-0.004	-0.004	-0.003	0.180	0.425	0.433	1.019	95.80	19.30
100	None	T-IPW 0.05/0.95	0.471	0.352	0.505	0.002	-0.003	-0.003	0.184	0.429	0.439	1.022	95.80	20.00
100	None	T-IPW 0.025/0.975	0.471	0.352	0.505	0.007	-0.002	-0.004	0.189	0.435	0.444	1.019	95.80	20.20
100	None	TMLE	0.471	0.352	0.505	0.001	-0.001	-0.000	0.204	0.452	0.406	0.899	91.90	24.70
100	None	T-TMLE 0.1/0.9	0.471	0.352	0.505	0.000	-0.001	-0.000	0.204	0.452	0.405	0.896	92.00	24.90
100	None	T-TMLE 0.05/0.95	0.471	0.352	0.505	0.001	-0.001	-0.001	0.204	0.452	0.405	0.897	92.00	24.80
100	None	T-TMLE 0.025/0.975	0.471	0.352	0.505	0.001	-0.001	-0.001	0.204	0.452	0.406	0.897	91.90	25.00

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S8. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 80% with a binary outcome, under the null hypothesis, for sample sizes of 1,000 and 500 subjects, The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Type I error
1000	High	Crude	0.350	0.350	0.004	-0.046	-0.002	0.008	0.031	0.169	0.167	0.986	93.40	6.40
1000	High	GC	0.350	0.350	0.004	-0.003	-0.000	0.001	0.035	0.188	0.179	0.953	93.40	6.50
1000	High	IPW	0.350	0.350	0.004	-0.000	-0.000	0.000	0.040	0.201	0.196	0.977	94.10	5.90
1000	High	T-IPW 0.1/0.9	0.350	0.350	0.004	-0.022	-0.001	0.004	0.028	0.167	0.168	1.007	94.30	5.60
1000	High	T-IPW 0.05/0.95	0.350	0.350	0.004	-0.006	-0.000	0.001	0.028	0.168	0.171	1.015	95.30	4.60
1000	High	T-IPW 0.025/0.975	0.350	0.350	0.004	0.004	-0.000	-0.001	0.029	0.171	0.173	1.015	95.30	4.60
1000	High	TMLE	0.350	0.350	0.004	-0.002	-0.000	0.001	0.040	0.199	0.177	0.888	90.50	9.40
1000	High	T-TMLE 0.1/0.9	0.350	0.350	0.004	-0.004	-0.000	0.001	0.038	0.195	0.159	0.816	89.00	11.50
1000	High	T-TMLE 0.05/0.95	0.350	0.350	0.004	-0.003	-0.000	0.001	0.039	0.197	0.169	0.860	89.80	9.60
1000	High	T-TMLE 0.025/0.975	0.350	0.350	0.004	-0.002	-0.000	0.001	0.039	0.198	0.175	0.886	90.60	9.50
1000	Moderate	Crude	0.350	0.350	0.004	-0.046	-0.002	0.008	0.031	0.169	0.167	0.986	93.40	6.40
1000	Moderate	GC	0.350	0.350	0.004	-0.003	-0.000	0.001	0.035	0.188	0.179	0.953	93.40	6.50
1000	Moderate	IPW	0.350	0.350	0.004	-0.000	-0.000	0.000	0.040	0.201	0.196	0.977	94.10	5.90
1000	Moderate	T-IPW 0.1/0.9	0.350	0.350	0.004	-0.022	-0.001	0.004	0.028	0.167	0.168	1.007	94.30	5.60
1000	Moderate	T-IPW 0.05/0.95	0.350	0.350	0.004	-0.006	-0.000	0.001	0.028	0.168	0.171	1.015	95.30	4.60
1000	Moderate	T-IPW 0.025/0.975	0.350	0.350	0.004	0.004	-0.000	-0.001	0.029	0.171	0.173	1.015	95.30	4.60
1000	Moderate	TMLE	0.350	0.350	0.004	-0.002	-0.000	0.001	0.040	0.199	0.177	0.888	90.50	9.40
1000	Moderate	T-TMLE 0.1/0.9	0.350	0.350	0.004	-0.004	-0.000	0.001	0.038	0.195	0.159	0.816	89.00	11.50
1000	Moderate	T-TMLE 0.05/0.95	0.350	0.350	0.004	-0.003	-0.000	0.001	0.039	0.197	0.169	0.860	89.80	9.90
1000	Moderate	T-TMLE 0.025/0.975	0.350	0.350	0.004	-0.002	-0.000	0.001	0.039	0.198	0.175	0.886	90.60	9.50
1000	Low	Crude	0.350	0.350	0.004	-0.046	-0.002	0.008	0.031	0.169	0.167	0.986	93.40	6.40
1000	Low	GC	0.350	0.350	0.004	-0.003	-0.000	0.001	0.035	0.188	0.179	0.953	93.40	6.50
1000	Low	IPW	0.350	0.350	0.004	-0.000	-0.000	0.000	0.040	0.201	0.196	0.977	94.10	5.90
1000	Low	T-IPW 0.1/0.9	0.350	0.350	0.004	-0.022	-0.001	0.004	0.028	0.167	0.168	1.007	94.30	5.60
1000	Low	T-IPW 0.05/0.95	0.350	0.350	0.004	-0.006	-0.000	0.001	0.028	0.168	0.171	1.015	95.30	4.60
1000	Low	T-IPW 0.025/0.975	0.350	0.350	0.004	0.004	-0.000	-0.001	0.029	0.171	0.173	1.015	95.30	4.60
1000	Low	TMLE	0.350	0.350	0.004	-0.002	-0.000	0.001	0.040	0.199	0.177	0.888	90.50	9.40
1000	Low	T-TMLE 0.1/0.9	0.350	0.350	0.004	-0.004	-0.000	0.001	0.038	0.195	0.159	0.816	89.00	11.50
1000	Low	T-TMLE 0.05/0.95	0.350	0.350	0.004	-0.003	-0.000	0.001	0.039	0.197	0.169	0.860	89.80	9.90
1000	Low	T-TMLE 0.025/0.975	0.350	0.350	0.004	-0.002	-0.000	0.001	0.039	0.198	0.175	0.886	90.60	9.50
1000	None	Crude	0.350	0.350	0.004	-0.046	-0.002	0.008	0.031	0.169	0.167	0.986	93.40	6.40
1000	None	GC	0.350	0.350	0.004	-0.003	-0.000	0.001	0.035	0.188	0.179	0.953	93.40	6.50
1000	None	IPW	0.350	0.350	0.004	-0.000	-0.000	0.000	0.040	0.201	0.196	0.977	94.10	5.90
1000	None	T-IPW 0.1/0.9	0.350	0.350	0.004	-0.022	-0.001	0.004	0.028	0.167	0.168	1.007	94.30	5.60
1000	None	T-IPW 0.05/0.95	0.350	0.350	0.004	-0.006	-0.000	0.001	0.028	0.168	0.171	1.015	95.30	4.60
1000	None	T-IPW 0.025/0.975	0.350	0.350	0.004	0.004	-0.000	-0.001	0.029	0.171	0.173	1.015	95.30	4.60
1000	None	TMLE	0.350	0.350	0.004	-0.002	-0.000	0.001	0.040	0.199	0.177	0.888	90.50	9.40
1000	None	T-TMLE 0.1/0.9	0.350	0.350	0.004	-0.004	-0.000	0.001	0.038	0.195	0.159	0.816	89.00	11.50
1000	None	T-TMLE 0.05/0.95	0.350	0.350	0.004	-0.003	-0.000	0.001	0.039	0.197	0.169	0.860	89.80	9.90
1000	None	T-TMLE 0.025/0.975	0.350	0.350	0.004	-0.002	-0.000	0.001	0.039	0.198	0.175	0.886	90.60	9.50
500	High	Crude	0.350	0.350	0.009	-0.049	-0.003	0.008	0.061	0.242	0.239	0.988	93.00	6.90
500	High	GC	0.350	0.350	0.009	-0.005	-0.000	0.001	0.069	0.262	0.246	0.939	93.70	6.20
500	High	IPW	0.350	0.350	0.009	0.000	-0.001	0.000	0.080	0.283	0.277	0.978	95.60	4.40
500	High	T-IPW 0.1/0.9	0.350	0.350	0.009	-0.026	-0.001	0.004	0.056	0.235	0.238	1.015	94.90	5.30
500	High	T-IPW 0.05/0.95	0.350	0.350	0.009	-0.008	-0.001	0.001	0.056	0.237	0.243	1.024	95.00	4.60
500	High	T-IPW 0.025/0.975	0.350	0.350	0.009	0.002	-0.001	-0.001	0.058	0.241	0.247	1.025	95.50	4.50
500	High	TMLE	0.350	0.350	0.009	-0.008	-0.001	0.002	0.082	0.286	0.237	0.831	89.70	10.20
500	High	T-TMLE 0.1/0.9	0.350	0.350	0.009	-0.004	-0.001	0.001	0.074	0.272	0.223	0.820	89.60	10.30
500	High	T-TMLE 0.05/0.95	0.350	0.350	0.009	-0.004	-0.001	0.001	0.077	0.277	0.234	0.844	90.40	9.50
500	High	T-TMLE 0.025/0.975	0.350	0.350	0.009	-0.004	-0.001	0.001	0.078	0.279	0.237	0.849	90.60	9.40
500	Moderate	Crude	0.350	0.350	0.009	-0.049	-0.003	0.008	0.061	0.242	0.239	0.988	93.00	6.90
500	Moderate	GC	0.350	0.350	0.009	-0.005	-0.000	0.001	0.069	0.262	0.246	0.939	93.70	6.20
500	Moderate	IPW	0.350	0.350	0.009	0.000	-0.001	0.000	0.080	0.283	0.277	0.978	95.60	4.40
500	Moderate	T-IPW 0.1/0.9	0.350	0.350	0.009	-0.026	-0.001	0.004	0.056	0.235	0.238	1.015	94.90	5.30
500	Moderate	T-IPW 0.05/0.95	0.350	0.350	0.009	-0.008	-0.001	0.001	0.056	0.237	0.243	1.024	95.00	4.60
500	Moderate	T-IPW 0.025/0.975	0.350	0.350	0.009	0.002	-0.001	-0.001	0.058	0.241	0.247	1.025	95.50	4.50
500	Moderate	TMLE	0.350	0.350	0.009	-0.008	-0.001	0.002	0.082	0.286	0.237	0.831	89.70	10.20
500	Moderate	T-TMLE 0.1/0.9	0.350	0.350	0.009	-0.004	-0.001	0.001	0.074	0.272	0.223	0.820	89.60	10.30
500	Moderate	T-TMLE 0.05/0.95	0.350	0.350	0.009	-0.004	-0.001	0.001	0.077	0.277	0.234	0.844	90.40	9.70
500	Moderate	T-TMLE 0.025/0.975	0.350	0.350	0.009	-0.004	-0.001	0.001	0.078	0.279	0.237	0.849	90.60	9.40
500	Low	Crude	0.350	0.350	0.009	-0.049	-0.003	0.008	0.061	0.242	0.239	0.988	93.00	6.90
500	Low	GC	0.350	0.350	0.009	-0.005	-0.000	0.001	0.069	0.262	0.246	0.939	93.70	6.20
500	Low	IPW	0.350	0.350	0.009	0.000	-0.001	0.000	0.080	0.283	0.277	0.978	95.60	4.40
500	Low	T-IPW 0.1/0.9	0.350	0.350	0.009	-0.026	-0.001	0.004	0.056	0.235	0.238	1.015	94.90	5.30
500	Low	T-IPW 0.05/0.95	0.350	0.350	0.009	-0.008	-0.001	0.001	0.056	0.237	0.243	1.024	95.00	4.60
500	Low	T-IPW 0.025/0.975	0.350	0.350	0.009	0.002	-0.001	-0.001	0.058	0.241	0.247	1.025	95.50	4.50
500	Low	TMLE	0.350	0.350	0.009	-0.008	-0.001	0.002	0.082	0.286	0.237	0.831	89.70	10.20
500	Low	T-TMLE 0.1/0.9	0.350	0.350	0.009	-0.004	-0.001	0.001	0.074	0.272	0.223	0.820	89.60	10.30
500	Low	T-TMLE 0.05/0.95	0.350	0.350	0.009	-0.004	-0.001	0.001	0.077	0.277	0.234	0.844	90.40	9.70
500	Low	T-TMLE 0.025/0.975	0.350	0.350	0.009	-0.004	-0.001	0.001	0.078	0.279	0.237	0.849	90.60	9.40
500	None	Crude	0.350	0.350	0.009	-0.049	-0.003	0.008	0.061	0.242	0.239	0.988	93.00	6.90
500	None	GC	0.350	0.350	0.009	-0.005	-0.000	0.001	0.069	0.262	0.246	0.939	93.70	6.20
500	None	IPW	0.350	0.350	0.009	0.000	-0.001	0.000	0.080	0.283	0.277	0.978	95.60	4.40
500	None	T-IPW 0.1/0.9	0.350	0.350	0.009	-0.026	-0.001	0.004	0.056	0.235	0.238	1.015	94.90	5.30
500	None	T-IPW 0.05/0.95	0.350	0.350	0.009	-0.008	-0.001	0.001	0.056	0.237	0.243	1.024	95.00	4.60
500	None	T-IPW 0.025/0.975	0.350	0.350	0.009	0.002	-0.001	-0.001	0.058	0.241	0.247	1.025	95.50	4.50
500	None	TMLE	0.350	0.350	0.009	-0.008	-0.001	0.002	0.082	0.286	0.237	0.831	89.70	10.20
500	None	T-TMLE 0.1/0.9	0.350	0.350	0.009	-0.004	-0.001	0.001	0.074	0.272	0.223	0.820	89.60	10.30
500	None	T-TMLE 0.05/0.95	0.350	0.350	0.009	-0.004	-0.001	0.001	0.077	0.277	0.234	0.844	90.40	9.70
500	None	T-TMLE 0.025/0.975	0.350	0.350	0.009	-0.004	-0.001	0.001	0.078	0.279	0.237	0.849	90.60	9.40

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S9. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 80% with a binary outcome, under the null hypothesis, for sample sizes of 200 and 100 subjects, The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Type I error
200	High	Crude	0.351	0.350	0.018	-0.055	-0.002	0.010	0.151	0.385	0.408	1.058	94.60	5.50
200	High	GC	0.351	0.350	0.018	-0.006	0.000	0.002	0.161	0.402	0.404	1.005	94.20	5.80
200	High	IPW	0.351	0.350	0.018	0.022	-0.000	-0.003	0.204	0.451	0.432	0.957	95.10	4.90
200	High	T-IPW 0.1/0.9	0.351	0.350	0.018	-0.030	-0.001	0.006	0.147	0.383	0.384	1.003	96.20	3.90
200	High	T-IPW 0.05/0.95	0.351	0.350	0.018	-0.008	-0.001	0.001	0.154	0.392	0.394	1.005	95.90	4.00
200	High	T-IPW 0.025/0.975	0.351	0.350	0.018	0.006	-0.000	-0.001	0.161	0.401	0.402	1.002	95.60	4.30
200	High	TMLE	0.351	0.350	0.018	0.009	-0.000	-0.000	0.212	0.461	0.369	0.801	88.90	10.90
200	High	T-TMLE 0.1/0.9	0.351	0.350	0.018	0.002	-0.000	0.000	0.183	0.428	0.348	0.815	89.30	10.20
200	High	T-TMLE 0.05/0.95	0.351	0.350	0.018	0.005	-0.000	0.000	0.194	0.441	0.364	0.827	90.00	9.10
200	High	T-TMLE 0.025/0.975	0.351	0.350	0.018	0.007	-0.000	0.000	0.201	0.448	0.367	0.818	89.80	10.00
200	Moderate	Crude	0.351	0.350	0.018	-0.055	-0.002	0.010	0.151	0.385	0.408	1.058	94.60	5.50
200	Moderate	GC	0.351	0.350	0.018	-0.006	0.000	0.002	0.161	0.402	0.404	1.005	94.20	5.80
200	Moderate	IPW	0.351	0.350	0.018	0.022	-0.000	-0.003	0.204	0.451	0.432	0.957	95.10	4.90
200	Moderate	T-IPW 0.1/0.9	0.351	0.350	0.018	-0.030	-0.001	0.006	0.147	0.383	0.384	1.003	96.20	3.90
200	Moderate	T-IPW 0.05/0.95	0.351	0.350	0.018	-0.008	-0.001	0.001	0.154	0.392	0.394	1.005	95.90	4.00
200	Moderate	T-IPW 0.025/0.975	0.351	0.350	0.018	0.006	-0.000	-0.001	0.161	0.401	0.402	1.002	95.60	4.30
200	Moderate	TMLE	0.351	0.350	0.018	0.009	-0.000	-0.000	0.212	0.461	0.369	0.801	88.90	10.90
200	Moderate	T-TMLE 0.1/0.9	0.351	0.350	0.018	0.002	-0.000	0.000	0.183	0.428	0.348	0.815	89.30	10.20
200	Moderate	T-TMLE 0.05/0.95	0.351	0.350	0.018	0.005	-0.000	0.000	0.194	0.441	0.364	0.827	90.00	9.70
200	Moderate	T-TMLE 0.025/0.975	0.351	0.350	0.018	0.007	-0.000	0.000	0.201	0.448	0.367	0.818	89.80	10.00
200	Low	Crude	0.351	0.350	0.018	-0.055	-0.002	0.010	0.151	0.385	0.408	1.058	94.60	5.50
200	Low	GC	0.351	0.350	0.018	-0.006	0.000	0.002	0.161	0.402	0.404	1.005	94.20	5.80
200	Low	IPW	0.351	0.350	0.018	0.022	-0.000	-0.003	0.204	0.451	0.432	0.957	95.10	4.90
200	Low	T-IPW 0.1/0.9	0.351	0.350	0.018	-0.030	-0.001	0.006	0.147	0.383	0.384	1.003	96.20	3.90
200	Low	T-IPW 0.05/0.95	0.351	0.350	0.018	-0.008	-0.001	0.001	0.154	0.392	0.394	1.005	95.90	4.00
200	Low	T-IPW 0.025/0.975	0.351	0.350	0.018	0.006	-0.000	-0.001	0.161	0.401	0.402	1.002	95.60	4.30
200	Low	TMLE	0.351	0.350	0.018	0.009	-0.000	-0.000	0.212	0.461	0.369	0.801	88.90	10.90
200	Low	T-TMLE 0.1/0.9	0.351	0.350	0.018	0.002	-0.000	0.000	0.183	0.428	0.348	0.815	89.30	10.20
200	Low	T-TMLE 0.05/0.95	0.351	0.350	0.018	0.005	-0.000	0.000	0.194	0.441	0.364	0.827	90.00	9.70
200	Low	T-TMLE 0.025/0.975	0.351	0.350	0.018	0.007	-0.000	0.000	0.201	0.448	0.367	0.818	89.80	10.00
200	None	Crude	0.351	0.350	0.018	-0.055	-0.002	0.010	0.151	0.385	0.408	1.058	94.60	5.50
200	None	GC	0.351	0.350	0.018	-0.006	0.000	0.002	0.161	0.402	0.404	1.005	94.20	5.80
200	None	IPW	0.351	0.350	0.018	0.022	-0.000	-0.003	0.204	0.451	0.432	0.957	95.10	4.90
200	None	T-IPW 0.1/0.9	0.351	0.350	0.018	-0.030	-0.001	0.006	0.147	0.383	0.384	1.003	96.20	3.90
200	None	T-IPW 0.05/0.95	0.351	0.350	0.018	-0.008	-0.001	0.001	0.154	0.392	0.394	1.005	95.90	4.00
200	None	T-IPW 0.025/0.975	0.351	0.350	0.018	0.006	-0.000	-0.001	0.161	0.401	0.402	1.002	95.60	4.30
200	None	TMLE	0.351	0.350	0.018	0.009	-0.000	-0.000	0.212	0.461	0.369	0.801	88.90	10.90
200	None	T-TMLE 0.1/0.9	0.351	0.350	0.018	0.002	-0.000	0.000	0.183	0.428	0.348	0.815	89.30	10.20
200	None	T-TMLE 0.05/0.95	0.351	0.350	0.018	0.005	-0.000	0.000	0.194	0.441	0.364	0.827	90.00	9.70
200	None	T-TMLE 0.025/0.975	0.351	0.350	0.018	0.007	-0.000	0.000	0.201	0.448	0.367	0.818	89.80	10.00
100	High	Crude	0.349	0.352	0.020	-0.021	-0.001	0.004	0.332	0.576	1.214	2.106	94.90	5.00
100	High	GC	0.349	0.352	0.020	0.019	0.001	-0.002	0.339	0.582	1.197	2.055	94.70	5.10
100	High	IPW	0.349	0.352	0.020	0.070	0.001	-0.009	0.495	0.700	0.628	0.897	94.30	5.60
100	High	T-IPW 0.1/0.9	0.349	0.352	0.020	0.001	-0.000	0.000	0.328	0.573	0.563	0.983	96.10	3.80
100	High	T-IPW 0.05/0.95	0.349	0.352	0.020	0.029	0.000	-0.005	0.349	0.591	0.580	0.982	96.20	4.00
100	High	T-IPW 0.025/0.975	0.349	0.352	0.020	0.046	0.001	-0.007	0.371	0.608	0.592	0.974	95.50	4.50
100	High	TMLE	0.349	0.352	0.020	0.026	0.001	0.001	0.526	0.725	0.514	0.709	86.00	14.00
100	High	T-TMLE 0.1/0.9	0.349	0.352	0.020	0.030	0.001	-0.003	0.400	0.632	0.478	0.756	87.90	12.10
100	High	T-TMLE 0.05/0.95	0.349	0.352	0.020	0.038	0.001	-0.003	0.453	0.672	0.502	0.747	87.90	11.80
100	High	T-TMLE 0.025/0.975	0.349	0.352	0.020	0.038	0.001	-0.002	0.491	0.700	0.509	0.728	86.60	13.40
100	Moderate	Crude	0.349	0.352	0.020	-0.021	-0.001	0.004	0.332	0.576	1.214	2.106	94.90	5.00
100	Moderate	GC	0.349	0.352	0.020	0.019	0.001	-0.002	0.339	0.582	1.197	2.055	94.70	5.10
100	Moderate	IPW	0.349	0.352	0.020	0.070	0.001	-0.009	0.495	0.700	0.628	0.897	94.30	5.60
100	Moderate	T-IPW 0.1/0.9	0.349	0.352	0.020	0.001	-0.000	0.000	0.328	0.573	0.563	0.983	96.10	3.80
100	Moderate	T-IPW 0.05/0.95	0.349	0.352	0.020	0.029	0.000	-0.005	0.349	0.591	0.580	0.982	96.20	4.00
100	Moderate	T-IPW 0.025/0.975	0.349	0.352	0.020	0.046	0.001	-0.007	0.371	0.608	0.592	0.974	95.50	4.50
100	Moderate	TMLE	0.349	0.352	0.020	0.026	0.001	0.001	0.526	0.725	0.514	0.709	86.00	14.00
100	Moderate	T-TMLE 0.1/0.9	0.349	0.352	0.020	0.030	0.001	-0.003	0.400	0.632	0.478	0.756	87.90	12.10
100	Moderate	T-TMLE 0.05/0.95	0.349	0.352	0.020	0.038	0.001	-0.003	0.453	0.672	0.502	0.747	87.90	11.90
100	Moderate	T-TMLE 0.025/0.975	0.349	0.352	0.020	0.038	0.001	-0.002	0.491	0.700	0.509	0.728	86.60	13.40
100	Low	Crude	0.349	0.352	0.020	-0.021	-0.001	0.004	0.332	0.576	1.214	2.106	94.90	5.00
100	Low	GC	0.349	0.352	0.020	0.019	0.001	-0.002	0.339	0.582	1.197	2.055	94.70	5.10
100	Low	IPW	0.349	0.352	0.020	0.070	0.001	-0.009	0.495	0.700	0.628	0.897	94.30	5.60
100	Low	T-IPW 0.1/0.9	0.349	0.352	0.020	0.001	-0.000	0.000	0.328	0.573	0.563	0.983	96.10	3.80
100	Low	T-IPW 0.05/0.95	0.349	0.352	0.020	0.029	0.000	-0.005	0.349	0.591	0.580	0.982	96.20	4.00
100	Low	T-IPW 0.025/0.975	0.349	0.352	0.020	0.046	0.001	-0.007	0.371	0.608	0.592	0.974	95.50	4.50
100	Low	TMLE	0.349	0.352	0.020	0.026	0.001	0.001	0.526	0.725	0.514	0.709	86.00	14.00
100	Low	T-TMLE 0.1/0.9	0.349	0.352	0.020	0.030	0.001	-0.003	0.400	0.632	0.478	0.756	87.90	12.10
100	Low	T-TMLE 0.05/0.95	0.349	0.352	0.020	0.038	0.001	-0.003	0.453	0.672	0.502	0.747	87.90	11.90
100	Low	T-TMLE 0.025/0.975	0.349	0.352	0.020	0.038	0.001	-0.002	0.491	0.700	0.509	0.728	86.60	13.40
100	None	Crude	0.349	0.352	0.020	-0.021	-0.001	0.004	0.332	0.576	1.214	2.106	94.90	5.00
100	None	GC	0.349	0.352	0.020	0.019	0.001	-0.002	0.339	0.582	1.197	2.055	94.70	5.10
100	None	IPW	0.349	0.352	0.020	0.070	0.001	-0.009	0.495	0.700	0.628	0.897	94.30	5.60
100	None	T-IPW 0.1/0.9	0.349	0.352	0.020	0.001	-0.000	0.000	0.328	0.573	0.563	0.983	96.10	3.80
100	None	T-IPW 0.05/0.95	0.349	0.352	0.020	0.029	0.000	-0.005	0.349	0.591	0.580	0.982	96.20	4.00
100	None	T-IPW 0.025/0.975	0.349	0.352	0.020	0.046	0.001	-0.007	0.371	0.608	0.592	0.974	95.50	4.50
100	None	TMLE	0.349	0.352	0.020	0.026	0.001	0.001	0.526	0.725	0.514	0.709	86.00	14.00
100	None	T-TMLE 0.1/0.9	0.349	0.352	0.020	0.030	0.001	-0.003	0.400	0.632	0.478	0.756	87.90	12.10
100	None	T-TMLE 0.05/0.95	0.349	0.352	0.020	0.038	0.001	-0.003	0.453	0.672	0.502	0.747	87.90	11.90
100	None	T-TMLE 0.025/0.975	0.349	0.352	0.020	0.038	0.001	-0.002	0.491	0.700	0.509	0.728	86.60	13.40

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S10. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 50% with a binary outcome, under the null hypothesis, for sample sizes of 1,000 and 500 subjects, The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Type I error
1000	High	Crude	0.350	0.350	0.003	-0.031	-0.004	0.003	0.020	0.139	0.133	0.956	92.80	7.10
1000	High	GC	0.350	0.350	0.003	-0.004	-0.000	0.000	0.019	0.138	0.135	0.980	94.40	5.80
1000	High	IPW	0.350	0.350	0.003	-0.006	-0.001	0.001	0.020	0.141	0.143	1.017	95.70	4.20
1000	High	T-IPW 0.1/0.9	0.350	0.350	0.003	-0.004	-0.002	-0.001	0.018	0.136	0.136	1.002	95.10	4.80
1000	High	T-IPW 0.05/0.95	0.350	0.350	0.003	-0.001	-0.002	-0.001	0.019	0.137	0.138	1.007	95.20	4.90
1000	High	T-IPW 0.025/0.975	0.350	0.350	0.003	-0.000	-0.001	-0.001	0.019	0.137	0.139	1.010	95.30	4.80
1000	High	TMLE	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	High	T-TMLE 0.1/0.9	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.140	0.135	0.967	94.30	5.50
1000	High	T-TMLE 0.05/0.95	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	High	T-TMLE 0.025/0.975	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	Moderate	Crude	0.350	0.350	0.003	-0.031	-0.004	0.003	0.020	0.139	0.133	0.956	92.80	7.10
1000	Moderate	GC	0.350	0.350	0.003	-0.004	-0.000	0.000	0.019	0.138	0.135	0.980	94.40	5.80
1000	Moderate	IPW	0.350	0.350	0.003	-0.006	-0.001	0.001	0.020	0.141	0.143	1.017	95.70	4.20
1000	Moderate	T-IPW 0.1/0.9	0.350	0.350	0.003	-0.004	-0.002	-0.001	0.018	0.136	0.136	1.002	95.10	4.80
1000	Moderate	T-IPW 0.05/0.95	0.350	0.350	0.003	-0.001	-0.002	-0.001	0.019	0.137	0.138	1.007	95.20	4.90
1000	Moderate	T-IPW 0.025/0.975	0.350	0.350	0.003	-0.000	-0.001	-0.001	0.019	0.137	0.139	1.010	95.30	4.80
1000	Moderate	TMLE	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	Moderate	T-TMLE 0.1/0.9	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.140	0.135	0.967	94.30	5.50
1000	Moderate	T-TMLE 0.05/0.95	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	Moderate	T-TMLE 0.025/0.975	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	Low	Crude	0.350	0.350	0.003	-0.031	-0.004	0.003	0.020	0.139	0.133	0.956	92.80	7.10
1000	Low	GC	0.350	0.350	0.003	-0.004	-0.000	0.000	0.019	0.138	0.135	0.980	94.40	5.80
1000	Low	IPW	0.350	0.350	0.003	-0.006	-0.001	0.001	0.020	0.141	0.143	1.017	95.70	4.20
1000	Low	T-IPW 0.1/0.9	0.350	0.350	0.003	-0.004	-0.002	-0.001	0.018	0.136	0.136	1.002	95.10	4.80
1000	Low	T-IPW 0.05/0.95	0.350	0.350	0.003	-0.001	-0.002	-0.001	0.019	0.137	0.138	1.007	95.20	4.90
1000	Low	T-IPW 0.025/0.975	0.350	0.350	0.003	-0.000	-0.001	-0.001	0.019	0.137	0.139	1.010	95.30	4.80
1000	Low	TMLE	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	Low	T-TMLE 0.1/0.9	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.140	0.135	0.967	94.30	5.50
1000	Low	T-TMLE 0.05/0.95	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	Low	T-TMLE 0.025/0.975	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	None	Crude	0.350	0.350	0.003	-0.031	-0.004	0.003	0.020	0.139	0.133	0.956	92.80	7.10
1000	None	GC	0.350	0.350	0.003	-0.004	-0.000	0.000	0.019	0.138	0.135	0.980	94.40	5.80
1000	None	IPW	0.350	0.350	0.003	-0.006	-0.001	0.001	0.020	0.141	0.143	1.017	95.70	4.20
1000	None	T-IPW 0.1/0.9	0.350	0.350	0.003	-0.004	-0.002	-0.001	0.018	0.136	0.136	1.002	95.10	4.80
1000	None	T-IPW 0.05/0.95	0.350	0.350	0.003	-0.001	-0.002	-0.001	0.019	0.137	0.138	1.007	95.20	4.90
1000	None	T-IPW 0.025/0.975	0.350	0.350	0.003	-0.000	-0.001	-0.001	0.019	0.137	0.139	1.010	95.30	4.80
1000	None	TMLE	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	None	T-TMLE 0.1/0.9	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.140	0.135	0.967	94.30	5.50
1000	None	T-TMLE 0.05/0.95	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
1000	None	T-TMLE 0.025/0.975	0.350	0.350	0.003	-0.005	-0.001	0.001	0.019	0.139	0.136	0.972	94.50	5.40
500	High	Crude	0.352	0.349	0.014	-0.041	-0.005	0.004	0.037	0.188	0.189	1.008	93.80	5.60
500	High	GC	0.352	0.349	0.014	-0.014	-0.002	0.002	0.036	0.190	0.194	1.022	94.70	5.20
500	High	IPW	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.194	0.204	1.052	96.00	4.00
500	High	T-IPW 0.1/0.9	0.352	0.349	0.014	-0.015	-0.004	-0.000	0.034	0.185	0.194	1.047	95.30	4.50
500	High	T-IPW 0.05/0.95	0.352	0.349	0.014	-0.013	-0.003	-0.000	0.035	0.186	0.195	1.049	95.20	5.20
500	High	T-IPW 0.025/0.975	0.352	0.349	0.014	-0.011	-0.002	0.000	0.035	0.187	0.197	1.051	94.90	4.90
500	High	TMLE	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	High	T-TMLE 0.1/0.9	0.352	0.349	0.014	-0.016	-0.002	0.002	0.037	0.193	0.190	0.988	94.80	5.20
500	High	T-TMLE 0.05/0.95	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.30
500	High	T-TMLE 0.025/0.975	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	Moderate	Crude	0.352	0.349	0.014	-0.041	-0.005	0.004	0.037	0.188	0.189	1.008	93.80	5.60
500	Moderate	GC	0.352	0.349	0.014	-0.014	-0.002	0.002	0.036	0.190	0.194	1.022	94.70	5.20
500	Moderate	IPW	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.194	0.204	1.052	96.00	4.00
500	Moderate	T-IPW 0.1/0.9	0.352	0.349	0.014	-0.015	-0.004	-0.000	0.034	0.185	0.194	1.047	95.30	4.50
500	Moderate	T-IPW 0.05/0.95	0.352	0.349	0.014	-0.013	-0.003	-0.000	0.035	0.186	0.195	1.049	95.20	5.20
500	Moderate	T-IPW 0.025/0.975	0.352	0.349	0.014	-0.011	-0.002	0.000	0.035	0.187	0.197	1.051	94.90	4.90
500	Moderate	TMLE	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	Moderate	T-TMLE 0.1/0.9	0.352	0.349	0.014	-0.016	-0.002	0.002	0.037	0.193	0.190	0.988	94.80	5.20
500	Moderate	T-TMLE 0.05/0.95	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	Moderate	T-TMLE 0.025/0.975	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	Low	Crude	0.352	0.349	0.014	-0.041	-0.005	0.004	0.037	0.188	0.189	1.008	93.80	5.60
500	Low	GC	0.352	0.349	0.014	-0.014	-0.002	0.002	0.036	0.190	0.194	1.022	94.70	5.20
500	Low	IPW	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.194	0.204	1.052	96.00	4.00
500	Low	T-IPW 0.1/0.9	0.352	0.349	0.014	-0.015	-0.004	-0.000	0.034	0.185	0.194	1.047	95.30	4.50
500	Low	T-IPW 0.05/0.95	0.352	0.349	0.014	-0.013	-0.003	-0.000	0.035	0.186	0.195	1.049	95.20	5.20
500	Low	T-IPW 0.025/0.975	0.352	0.349	0.014	-0.011	-0.002	0.000	0.035	0.187	0.197	1.051	94.90	4.90
500	Low	TMLE	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	Low	T-TMLE 0.1/0.9	0.352	0.349	0.014	-0.016	-0.002	0.002	0.037	0.193	0.190	0.988	94.80	5.20
500	Low	T-TMLE 0.05/0.95	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	Low	T-TMLE 0.025/0.975	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	None	Crude	0.352	0.349	0.014	-0.041	-0.005	0.004	0.037	0.188	0.189	1.008	93.80	5.60
500	None	GC	0.352	0.349	0.014	-0.014	-0.002	0.002	0.036	0.190	0.194	1.022	94.70	5.20
500	None	IPW	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.194	0.204	1.052	96.00	4.00
500	None	T-IPW 0.1/0.9	0.352	0.349	0.014	-0.015	-0.004	-0.000	0.034	0.185	0.194	1.047	95.30	4.50
500	None	T-IPW 0.05/0.95	0.352	0.349	0.014	-0.013	-0.003	-0.000	0.035	0.186	0.195	1.049	95.20	5.20
500	None	T-IPW 0.025/0.975	0.352	0.349	0.014	-0.011	-0.002	0.000	0.035	0.187	0.197	1.051	94.90	4.90
500	None	TMLE	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	None	T-TMLE 0.1/0.9	0.352	0.349	0.014	-0.016	-0.002	0.002	0.037	0.193	0.190	0.988	94.80	5.20
500	None	T-TMLE 0.05/0.95	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50
500	None	T-TMLE 0.025/0.975	0.352	0.349	0.014	-0.016	-0.002	0.002	0.038	0.193	0.192	0.991	94.70	5.50

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S11. Comparative simulation results for GC, IPW, truncated IPW, TMLE and truncated TMLE models for an exposure prevalence of 50% with a binary outcome, under the null hypothesis, for sample sizes of 200 and 100 subjects, The target parameter was log(OR).

Sample size	Extrapolation	Method	π_1 theo	π_0 theo	log(OR) theo	log(OR) bias	π_1 bias	π_0 bias	MSE	Estimated Sd	Empirical Sd	Sd ratio	Coverage	Type I error
200	High	Crude	0.352	0.349	0.010	-0.034	-0.004	0.004	0.091	0.299	0.304	1.017	95.20	4.90
200	High	GC	0.352	0.349	0.010	-0.005	-0.001	0.001	0.097	0.311	0.314	1.008	94.60	5.50
200	High	IPW	0.352	0.349	0.010	-0.002	-0.001	-0.000	0.098	0.313	0.328	1.049	96.70	3.40
200	High	T-IPW 0.1/0.9	0.352	0.349	0.010	-0.009	-0.003	-0.001	0.088	0.297	0.309	1.040	96.20	3.50
200	High	T-IPW 0.05/0.95	0.352	0.349	0.010	-0.005	-0.002	-0.001	0.089	0.298	0.312	1.047	96.20	3.90
200	High	T-IPW 0.025/0.975	0.352	0.349	0.010	-0.001	-0.002	-0.001	0.090	0.300	0.315	1.051	96.10	3.80
200	High	TMLE	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.959	94.60	5.30
200	High	T-TMLE 0.1/0.9	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.299	0.957	94.40	5.80
200	High	T-TMLE 0.05/0.95	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.60
200	High	T-TMLE 0.025/0.975	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.30
200	Moderate	Crude	0.352	0.349	0.010	-0.034	-0.004	0.004	0.091	0.299	0.304	1.017	95.20	4.90
200	Moderate	GC	0.352	0.349	0.010	-0.005	-0.001	0.001	0.097	0.311	0.314	1.008	94.60	5.50
200	Moderate	IPW	0.352	0.349	0.010	-0.002	-0.001	-0.000	0.098	0.313	0.328	1.049	96.70	3.40
200	Moderate	T-IPW 0.1/0.9	0.352	0.349	0.010	-0.009	-0.003	-0.001	0.088	0.297	0.309	1.040	96.20	3.50
200	Moderate	T-IPW 0.05/0.95	0.352	0.349	0.010	-0.005	-0.002	-0.001	0.089	0.298	0.312	1.047	96.20	3.90
200	Moderate	T-IPW 0.025/0.975	0.352	0.349	0.010	-0.001	-0.002	-0.001	0.090	0.300	0.315	1.051	96.10	3.80
200	Moderate	TMLE	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.959	94.60	5.30
200	Moderate	T-TMLE 0.1/0.9	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.299	0.957	94.40	5.80
200	Moderate	T-TMLE 0.05/0.95	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.30
200	Moderate	T-TMLE 0.025/0.975	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.30
200	Low	Crude	0.352	0.349	0.010	-0.034	-0.004	0.004	0.091	0.299	0.304	1.017	95.20	4.90
200	Low	GC	0.352	0.349	0.010	-0.005	-0.001	0.001	0.097	0.311	0.314	1.008	94.60	5.50
200	Low	IPW	0.352	0.349	0.010	-0.002	-0.001	-0.000	0.098	0.313	0.328	1.049	96.70	3.40
200	Low	T-IPW 0.1/0.9	0.352	0.349	0.010	-0.009	-0.003	-0.001	0.088	0.297	0.309	1.040	96.20	3.50
200	Low	T-IPW 0.05/0.95	0.352	0.349	0.010	-0.005	-0.002	-0.001	0.089	0.298	0.312	1.047	96.20	3.90
200	Low	T-IPW 0.025/0.975	0.352	0.349	0.010	-0.001	-0.002	-0.001	0.090	0.300	0.315	1.051	96.10	3.80
200	Low	TMLE	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.959	94.60	5.30
200	Low	T-TMLE 0.1/0.9	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.299	0.957	94.40	5.80
200	Low	T-TMLE 0.05/0.95	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.30
200	Low	T-TMLE 0.025/0.975	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.30
200	None	Crude	0.352	0.349	0.010	-0.034	-0.004	0.004	0.091	0.299	0.304	1.017	95.20	4.90
200	None	GC	0.352	0.349	0.010	-0.005	-0.001	0.001	0.097	0.311	0.314	1.008	94.60	5.50
200	None	IPW	0.352	0.349	0.010	-0.002	-0.001	-0.000	0.098	0.313	0.328	1.049	96.70	3.40
200	None	T-IPW 0.1/0.9	0.352	0.349	0.010	-0.009	-0.003	-0.001	0.088	0.297	0.309	1.040	96.20	3.50
200	None	T-IPW 0.05/0.95	0.352	0.349	0.010	-0.005	-0.002	-0.001	0.089	0.298	0.312	1.047	96.20	3.90
200	None	T-IPW 0.025/0.975	0.352	0.349	0.010	-0.001	-0.002	-0.001	0.090	0.300	0.315	1.051	96.10	3.80
200	None	TMLE	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.959	94.60	5.30
200	None	T-TMLE 0.1/0.9	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.299	0.957	94.40	5.80
200	None	T-TMLE 0.05/0.95	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.30
200	None	T-TMLE 0.025/0.975	0.352	0.349	0.010	-0.004	-0.001	0.000	0.097	0.312	0.300	0.960	94.60	5.30
100	High	Crude	0.348	0.352	-0.016	-0.012	-0.001	0.001	0.199	0.446	0.446	0.999	93.00	6.90
100	High	GC	0.348	0.352	-0.016	0.002	0.001	0.001	0.217	0.467	0.464	0.994	93.10	6.70
100	High	IPW	0.348	0.352	-0.016	0.018	0.001	-0.003	0.229	0.478	0.470	0.983	94.00	6.20
100	High	T-IPW 0.1/0.9	0.348	0.352	-0.016	0.003	-0.002	-0.003	0.198	0.445	0.444	0.997	94.30	5.60
100	High	T-IPW 0.05/0.95	0.348	0.352	-0.016	0.007	-0.002	-0.003	0.204	0.452	0.450	0.996	94.50	5.50
100	High	T-IPW 0.025/0.975	0.348	0.352	-0.016	0.012	-0.001	-0.004	0.210	0.458	0.455	0.994	94.50	5.60
100	High	TMLE	0.348	0.352	-0.016	0.004	0.000	-0.000	0.227	0.477	0.419	0.878	90.80	9.00
100	High	T-TMLE 0.1/0.9	0.348	0.352	-0.016	0.003	0.001	-0.000	0.226	0.475	0.417	0.877	90.90	9.10
100	High	T-TMLE 0.05/0.95	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.476	0.418	0.878	91.00	9.00
100	High	T-TMLE 0.025/0.975	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.475	0.418	0.879	91.10	8.80
100	Moderate	Crude	0.348	0.352	-0.016	-0.012	-0.001	0.001	0.199	0.446	0.446	0.999	93.00	6.90
100	Moderate	GC	0.348	0.352	-0.016	0.002	0.001	0.001	0.217	0.467	0.464	0.994	93.10	6.70
100	Moderate	IPW	0.348	0.352	-0.016	0.018	0.001	-0.003	0.229	0.478	0.470	0.983	94.00	6.20
100	Moderate	T-IPW 0.1/0.9	0.348	0.352	-0.016	0.003	-0.002	-0.003	0.198	0.445	0.444	0.997	94.30	5.60
100	Moderate	T-IPW 0.05/0.95	0.348	0.352	-0.016	0.007	-0.002	-0.003	0.204	0.452	0.450	0.996	94.50	5.50
100	Moderate	T-IPW 0.025/0.975	0.348	0.352	-0.016	0.012	-0.001	-0.004	0.210	0.458	0.455	0.994	94.50	5.60
100	Moderate	TMLE	0.348	0.352	-0.016	0.004	0.000	-0.000	0.227	0.477	0.419	0.878	90.80	9.00
100	Moderate	T-TMLE 0.1/0.9	0.348	0.352	-0.016	0.003	0.001	-0.000	0.226	0.475	0.417	0.877	90.90	9.10
100	Moderate	T-TMLE 0.05/0.95	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.476	0.418	0.878	91.00	8.90
100	Moderate	T-TMLE 0.025/0.975	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.475	0.418	0.879	91.10	8.80
100	Low	Crude	0.348	0.352	-0.016	-0.012	-0.001	0.001	0.199	0.446	0.446	0.999	93.00	6.90
100	Low	GC	0.348	0.352	-0.016	0.002	0.001	0.001	0.217	0.467	0.464	0.994	93.10	6.70
100	Low	IPW	0.348	0.352	-0.016	0.018	0.001	-0.003	0.229	0.478	0.470	0.983	94.00	6.20
100	Low	T-IPW 0.1/0.9	0.348	0.352	-0.016	0.003	-0.002	-0.003	0.198	0.445	0.444	0.997	94.30	5.60
100	Low	T-IPW 0.05/0.95	0.348	0.352	-0.016	0.007	-0.002	-0.003	0.204	0.452	0.450	0.996	94.50	5.50
100	Low	T-IPW 0.025/0.975	0.348	0.352	-0.016	0.012	-0.001	-0.004	0.210	0.458	0.455	0.994	94.50	5.60
100	Low	TMLE	0.348	0.352	-0.016	0.004	0.000	-0.000	0.227	0.477	0.419	0.878	90.80	9.00
100	Low	T-TMLE 0.1/0.9	0.348	0.352	-0.016	0.003	0.001	-0.000	0.226	0.475	0.417	0.877	90.90	9.10
100	Low	T-TMLE 0.05/0.95	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.476	0.418	0.878	91.00	8.90
100	Low	T-TMLE 0.025/0.975	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.475	0.418	0.879	91.10	8.80
100	None	Crude	0.348	0.352	-0.016	-0.012	-0.001	0.001	0.199	0.446	0.446	0.999	93.00	6.90
100	None	GC	0.348	0.352	-0.016	0.002	0.001	0.001	0.217	0.467	0.464	0.994	93.10	6.70
100	None	IPW	0.348	0.352	-0.016	0.018	0.001	-0.003	0.229	0.478	0.470	0.983	94.00	6.20
100	None	T-IPW 0.1/0.9	0.348	0.352	-0.016	0.003	-0.002	-0.003	0.198	0.445	0.444	0.997	94.30	5.60
100	None	T-IPW 0.05/0.95	0.348	0.352	-0.016	0.007	-0.002	-0.003	0.204	0.452	0.450	0.996	94.50	5.50
100	None	T-IPW 0.025/0.975	0.348	0.352	-0.016	0.012	-0.001	-0.004	0.210	0.458	0.455	0.994	94.50	5.60
100	None	TMLE	0.348	0.352	-0.016	0.004	0.000	-0.000	0.227	0.477	0.419	0.878	90.80	9.00
100	None	T-TMLE 0.1/0.9	0.348	0.352	-0.016	0.003	0.001	-0.000	0.226	0.475	0.417	0.877	90.90	9.10
100	None	T-TMLE 0.05/0.95	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.476	0.418	0.878	91.00	8.90
100	None	T-TMLE 0.025/0.975	0.348	0.352	-0.016	0.003	0.000	-0.000	0.226	0.475	0.418	0.879	91.10	8.80

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE; MSE, Mean Squared Error; Sd, Standard deviation; Sd ratio, the ratio between the estimated standard deviation and the empirical standard deviation.
For the truncated methods, the methods differed according to the indicated threshold (i.e. 0.1/0.9, 0.05/0.95 or 0.025/0.975).
π_1 theo, π_0 theo and log(OR) theo for the theoretical values of π_1, π_0 and log(OR).

TABLE S12. Baseline characteristics between the included and the excluded patients for the analysis of the AtlanREA cohort.

	Excluded (*n* =386)	**Included (*n* =1119)**	**P-value**	**SMD**
Center [*n* (%)]			<0.001	0.749
1	36 (9.3)	296 (26.5)		
2	19 (4.9)	87 (7.8)		
3	67 (17.4)	355 (31.7)		
4	139 (36.0)	210 (18.8)		
5	125 (32.4)	171 (15.3)		
Age in years [mean ± sd]	52.4 ± 15.9	50.2 ± 17.2	0.025	0.135
BMI in kg/m^2 [mean ± sd]	24.9 ± 5.1	25.3 ± 4.9	0.228	0.076
Male [*n* (%)]	234 (60.6)	718 (64.2)	<0.001	0.172
Cognitive dysfunction [*n* (%)]	3 (0.8)	8 (0.7)	1.000	0.007
History of brain trauma [*n* (%)]	16 (4.1)	58 (5.2)	0.498	0.049
Cardiac dysfunction [*n* (%)]	5 (1.3)	21 (1.9)	0.580	0.062
Respiratory dysfunction [*n* (%)]	14 (3.6)	29 (2.6)	0.372	0.080
History of malignancy [*n* (%)]	17 (4.4)	44 (3.9)	0.829	0.035
Diabetes [*n* (%)]	20 (5.2)	69 (6.2)	0.407	0.076
Chronic Alcoholism [*n* (%)]	69 (17.9)	201 (18.0)	0.173	0.106
Active smoking [*n* (%)]	112 (29.0)	288 (25.7)	0.171	0.111
Admission Glasgow Coma Scale [*n* (%)]			<0.001	0.333
3-7	246 (66.5)	825 (73.8)		
8-11	71 (19.2)	215 (19.2)		
12-15	53 (14.3)	79 (7.1)		
Brain trauma [*n* (%)]	139 (36.0)	587 (52.5)	<0.001	0.339
Subarachnoid hemorrhage [*n* (%)]	130 (33.7)	286 (25.6)	<0.001	0.313
Arteriovenous malformation [*n* (%)]	18 (4.7)	34 (3.0)	<0.001	0.313
Cerebral hematoma [*n* (%)]	67 (17.4)	162 (14.5)	<0.001	0.327
Ischemic stroke [*n* (%)]	21 (5.4)	41 (3.7)	<0.001	0.327
Other cerebral injuries [*n* (%)]	38 (9.8)	86 (7.7)	<0.001	0.333
Systolic Arterial Pressure < 90 mmHg [*n* (%)]	85 (22.0)	264 (23.6)	0.574	0.063
Cardiac arrest [*n* (%)]	11 (2.8)	40 (3.6)	0.667	0.059
Transfusion [*n* (%)]	31 (8.0)	152 (13.6)	0.015	0.181
Orotracheal intubation at admission [*n* (%)]	348 (90.2)	1088 (97.2)	<0.001	0.294
Hemoglobin in g/dL [mean ± sd]	12.1 ± 2.1	11.7 ± 2.2	0.004	0.177
Platelets in count/mm^3 [mean ± sd]	220.4 ± 88.3	200.8 ± 79.0	<0.001	0.234
Leukocytes in count/mm^3 [mean ± sd]	15.2 ± 5.3	16.3 ± 6.3	0.002	0.201
Serum Creatinine in mmol/L [mean ± sd]	68.5 ± 31.4	74.1 ± 44.1	0.022	0.148
Arterial pH [mean ± sd]	7.3 ± 0.1	7.3 ± 0.1	<0.001	0.272
Serum urea [mean ± sd]	5.1 ± 2.4	5.3 ± 3.1	0.348	0.066
Fibrinogen in g/L [mean ± sd]	3.1 ± 1.3	2.7 ± 1.2	0.002	0.245
Natremia in mmol/L [mean ± sd]	139.7 ± 4.9	140.3 ± 4.8	0.049	0.122
Prothrombin in % [mean ± sd]	78.8 ± 17.2	77.2 ± 18.4	0.142	0.094
Serum glucose in mmol/L [mean ± sd]	8.4 ± 2.9	8.6 ± 2.8	0.308	0.064
Arterial Lactate in mmol/L [mean ± sd]	2.8 ± 2.7	2.3 ± 1.8	0.004	0.247
PaO_2/fiO_2 [mean ± sd]	315.2 ± 147.8	312.8 ± 184.7	0.820	0.015
SAPS II Score [mean ± sd]	44.3 ± 13.5	47.5 ± 11.8	<0.001	0.257
Intracranial pressure in mmHg [mean ± sd]	15.6 ± 14.3	16.0 ± 14.9	0.707	0.024
Mortality in ICU [*n* (%)]	89 (24.1)	322 (28.8)	0.095	0.106

Abbreviations: BMI, Body Mass Index; SMD, Standard mean difference; sd, standard deviation.

TABLE S13. Baseline characteristics of the studied AtlanREA patients.

	Control (*n* =992)	**Barbiturate (*n* =127)**	**P-value**	**SMD**
Center [*n* (%)]			<0.001	0.612
1	262 (26.4)	34 (26.8)		
2	85 (8.6)	2 (1.6)		
3	313 (31.6)	42 (33.1)		
4	200 (20.2)	10 (7.9)		
5	132 (13.3)	39 (30.7)		
Age in years [mean ± sd]	51.0 ± 17.2	44.4 ± 16.1	<0.001	0.394
BMI in kg/m^2 [mean ± sd]	25.4 ± 5.0	24.9 ± 5.0	0.281	0.109
Male [*n* (%)]	631 (63.6)	87 (68.5)	0.325	0.104
Cognitive dysfunction [*n* (%)]	8 (0.8)	0 (0.0)	0.648	0.128
History of brain trauma [*n* (%)]	50 (5.0)	8 (6.3)	0.697	0.054
Cardiac dysfunction [*n* (%)]	18 (1.8)	3 (2.4)	0.220	0.216
Respiratory dysfunction [*n* (%)]	28 (2.8)	1 (0.8)	0.083	0.269
History of malignancy [*n* (%)]	41 (4.1)	3 (2.4)	0.245	0.198
Diabetes [*n* (%)]	64 (6.5)	5 (3.9)	0.209	0.206
Chronic Alcoholism [*n* (%)]	185 (18.6)	16 (12.6)	0.125	0.203
Active smoking [*n* (%)]	248 (25.0)	40 (31.5)	0.242	0.157
Admission Glasgow Coma Scale [*n* (%)]			0.055	0.464
3-7	720 (72.6)	105 (82.7)		
8-11	199 (20.0)	20 (12.6)		
12-15	73 (7.4)	6 (4.7)		
Brain trauma [*n* (%)]	512 (51.6)	75 (59.1)	0.273	0.156
Subarachnoid hemorrhage [*n* (%)]	251 (25.3)	35 (27.6)	0.044	0.254
Arteriovenous malformation [*n* (%)]	28 (2.8)	6 (4.7)	0.062	0.222
Cerebral hematoma [*n* (%)]	148 (14.9)	14 (11.0)	0.240	0.162
Ischemic stroke [*n* (%)]	36 (3.6)	5 (3.9)	0.234	0.163
Other cerebral injuries [*n* (%)]	82 (8.3)	4 (3.1)	0.063	0.247
Systolic Arterial Pressure < 90 mmHg [*n* (%)]	228 (23.0)	36 (28.3)	0.351	0.136
Cardiac arrest [*n* (%)]	35 (3.5)	5 (3.9)	0.913	0.050
Transfusion [*n* (%)]	128 (12.9)	24 (18.9)	0.142	0.186
Orotracheal intubation at admission [*n* (%)]	962 (97.0)	126 (99.2)	0.246	0.164
Hemoglobin in g/dL [mean ± sd]	11.7 ± 2.2	11.5 ± 2.6	0.289	0.094
Platelets in count/mm^3 [mean ± sd]	202.5 ± 78.7	188.3 ± 81.7	0.058	0.176
Leukocytes in count/mm^3 [mean ± sd]	16.2 ± 6.3	18.1 ± 6.5	0.002	0.293
Serum Creatinine in mmol/L [mean ± sd]	73.9 ± 45.7	76.3 ± 28.9	0.564	0.063
Arterial pH [mean ± sd]	7.3 ± 0.1	7.3 ± 0.1	0.003	0.268
Serum urea [mean ± sd]	5.4 ± 3.3	5.0 ± 2.4	0.210	0.132
Fibrinogen in g/L [mean ± sd]	2.8 ± 1.2	2.5 ± 1.4	0.003	0.293
Natremia in mmol/L [mean ± sd]	140.2 ± 4.6	141.0 ± 6.2	0.106	0.138
Prothrombin in % [mean ± sd]	77.6 ± 18.3	73.4 ± 19.0	0.016	0.227
Serum glucose in mmol/L [mean ± sd]	8.5 ± 2.7	9.1 ± 3.5	0.048	0.176
Arterial Lactate in mmol/L [mean ± sd]	2.2 ± 1.8	2.6 ± 1.6	0.016	0.235
PaO_2/fiO_2 [mean ± sd]	311.0 ± 185.7	327.2 ± 177.1	0.360	0.089
SAPS II Score [mean ± sd]	47.4 ± 11.8	49.0 ± 12.1	0.145	0.136
Intracranial pressure in mmHg [mean ± sd]	14.0 ± 12.4	31.0 ± 22.1	<0.001	0.947
Mortality in ICU [*n* (%)]	256 (25.8)	66 (52.0)	<0.001	0.557

Abbreviations: BMI, Body Mass Index; SMD, Standard mean difference; sd, standard deviation.

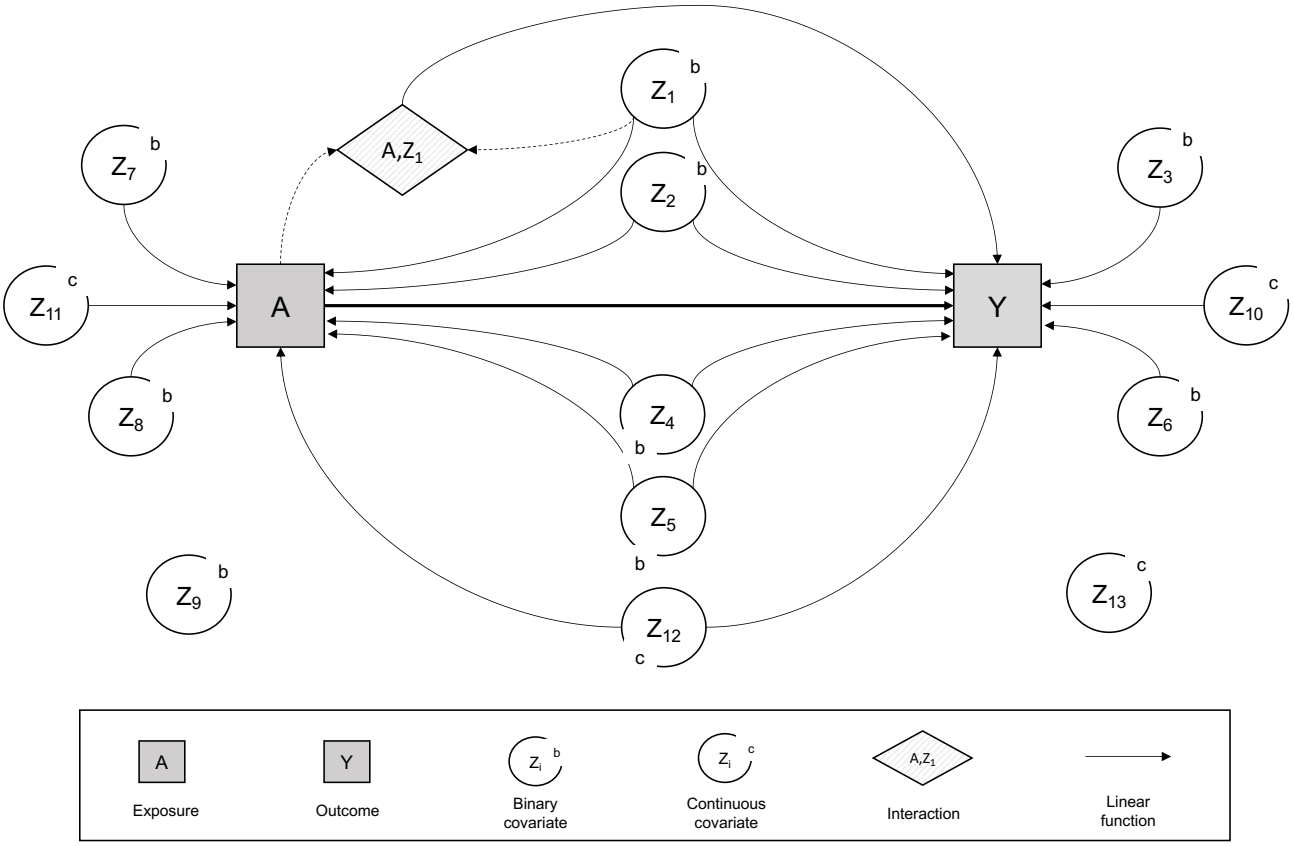


FIGURE S1. Directed acyclic graph related to the simulations.

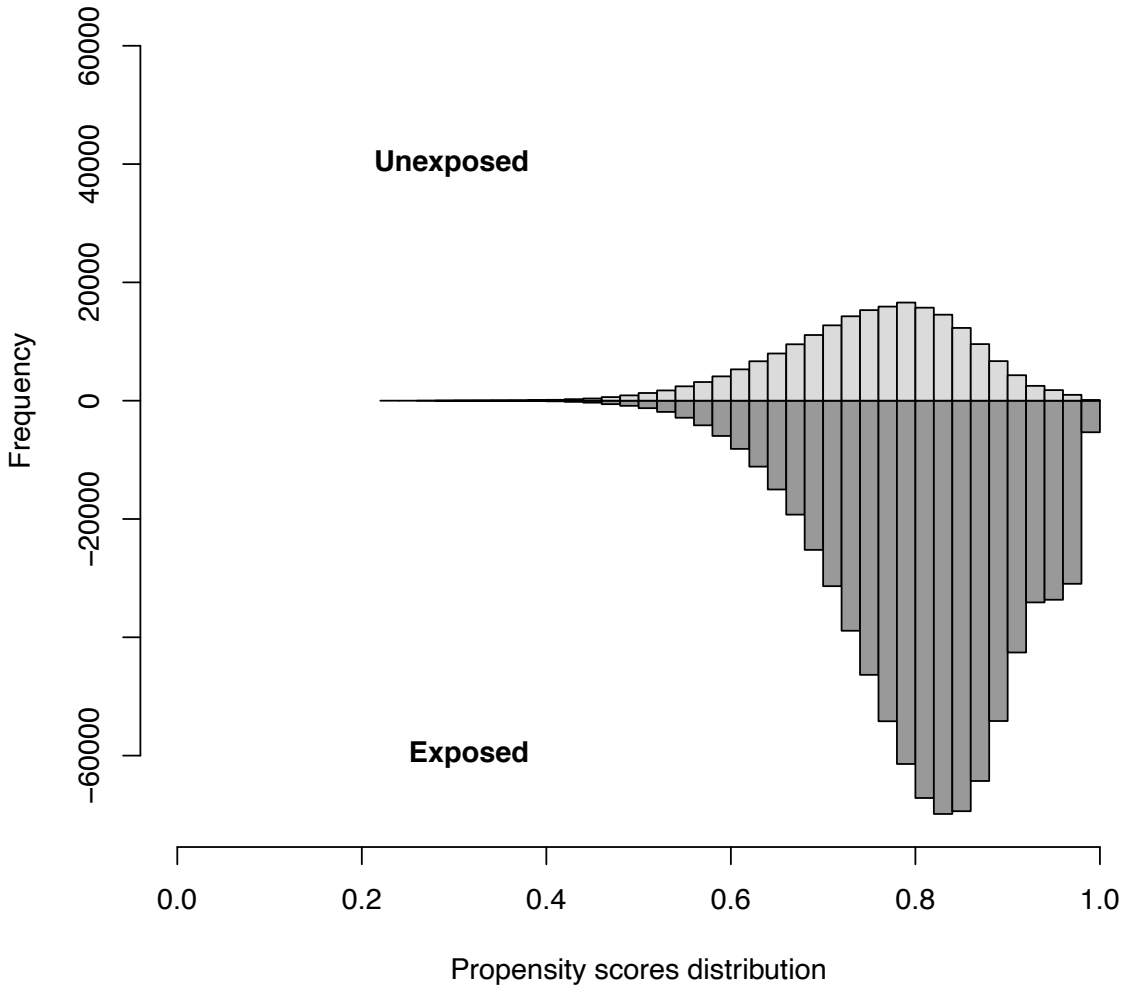


FIGURE S2. Propensity scores distribution between exposed and unexposed subjects in the scenario of positivity near-violation with an exposure prevalence at 80% (from a population of 1,000,000 subjects)

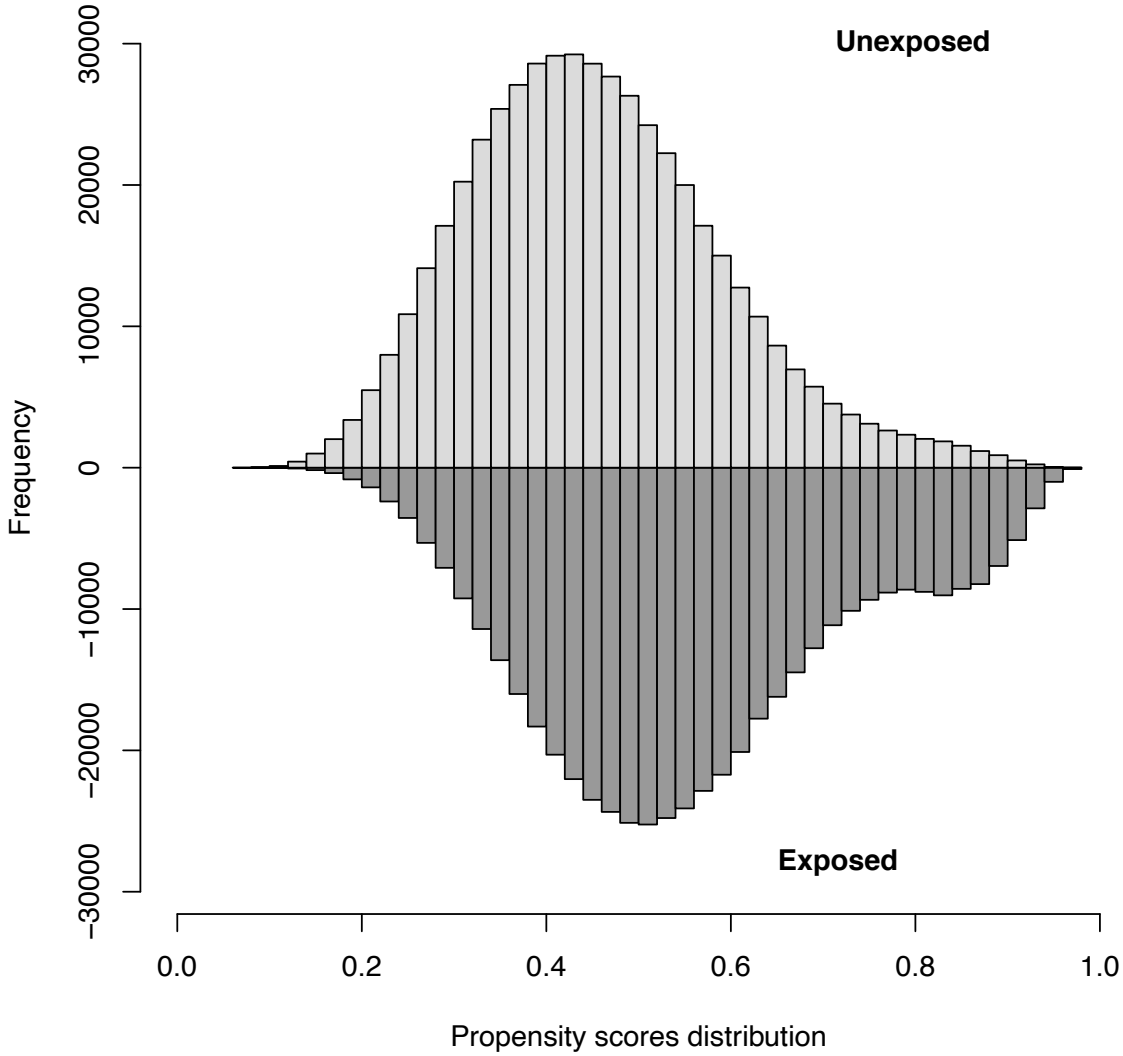


FIGURE S3. Propensity scores distribution between exposed and unexposed subjects in the scenario of positivity near-violation with an exposure prevalence at 50% (from a population of 1,000,000 subjects)

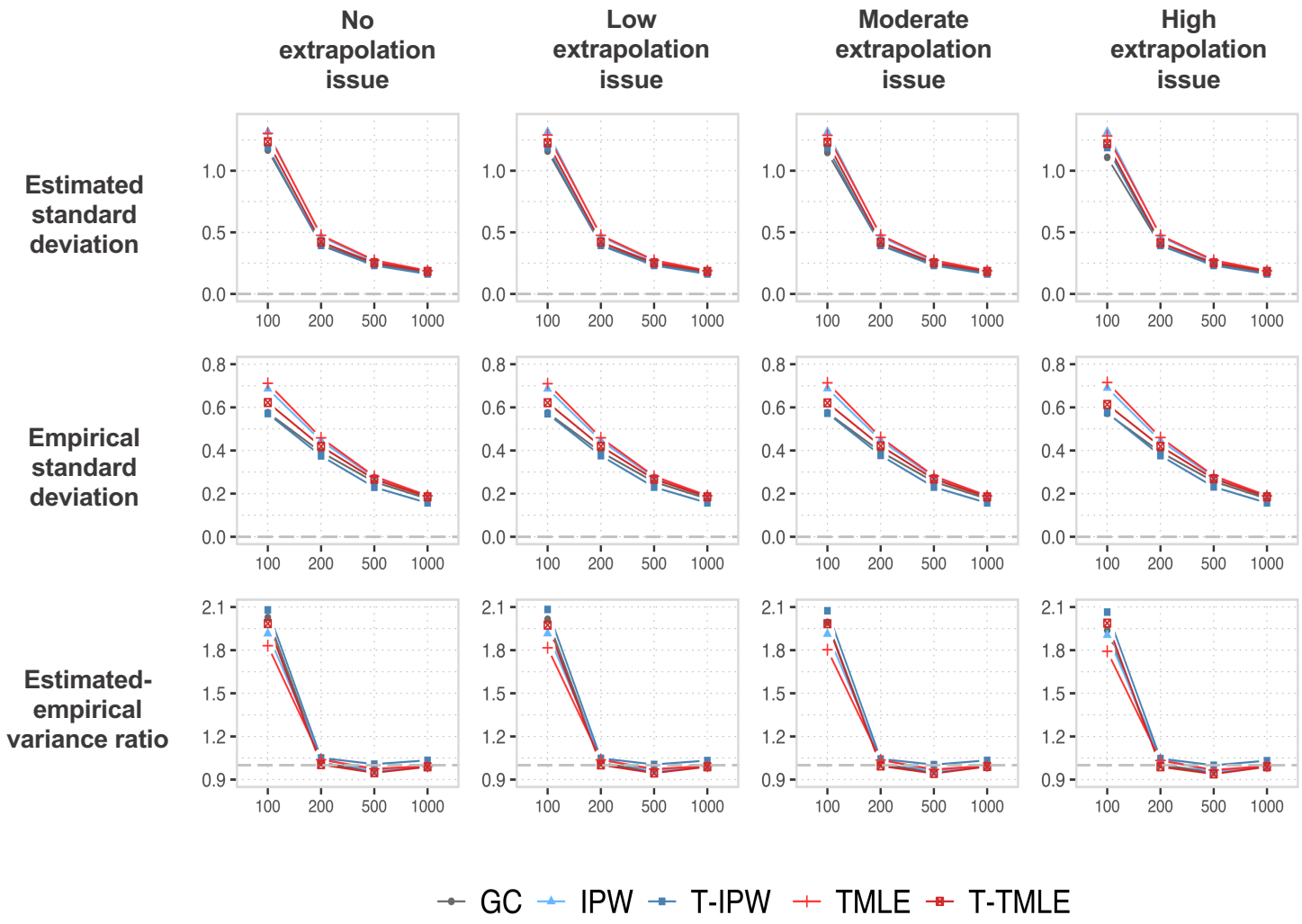


FIGURE S4. Graphical representation of the evolution of accuracy (empirical standard deviation, estimated standard deviation and, the variance estimation ratio) according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue. The target parameter was log(OR). Variance estimation was done using bootstrap (1000 datasets). Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9).

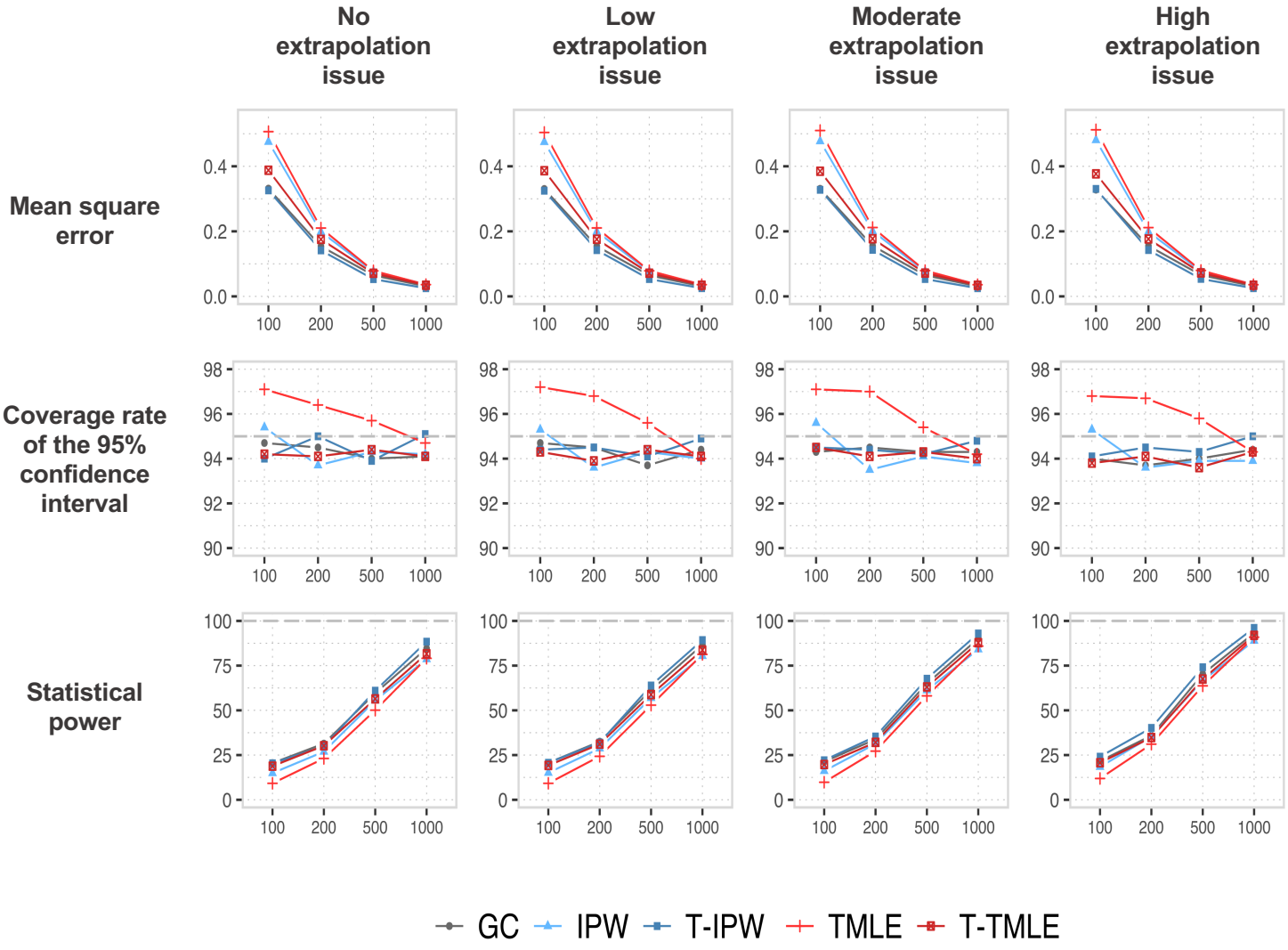


FIGURE S5. The mean square error, the coverage of the 95% confidence interval and the statistical power according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue. The target parameter was $log(\widehat{\text{OR}})$. Variance estimation was done using bootstrap (1000 datasets). Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9)

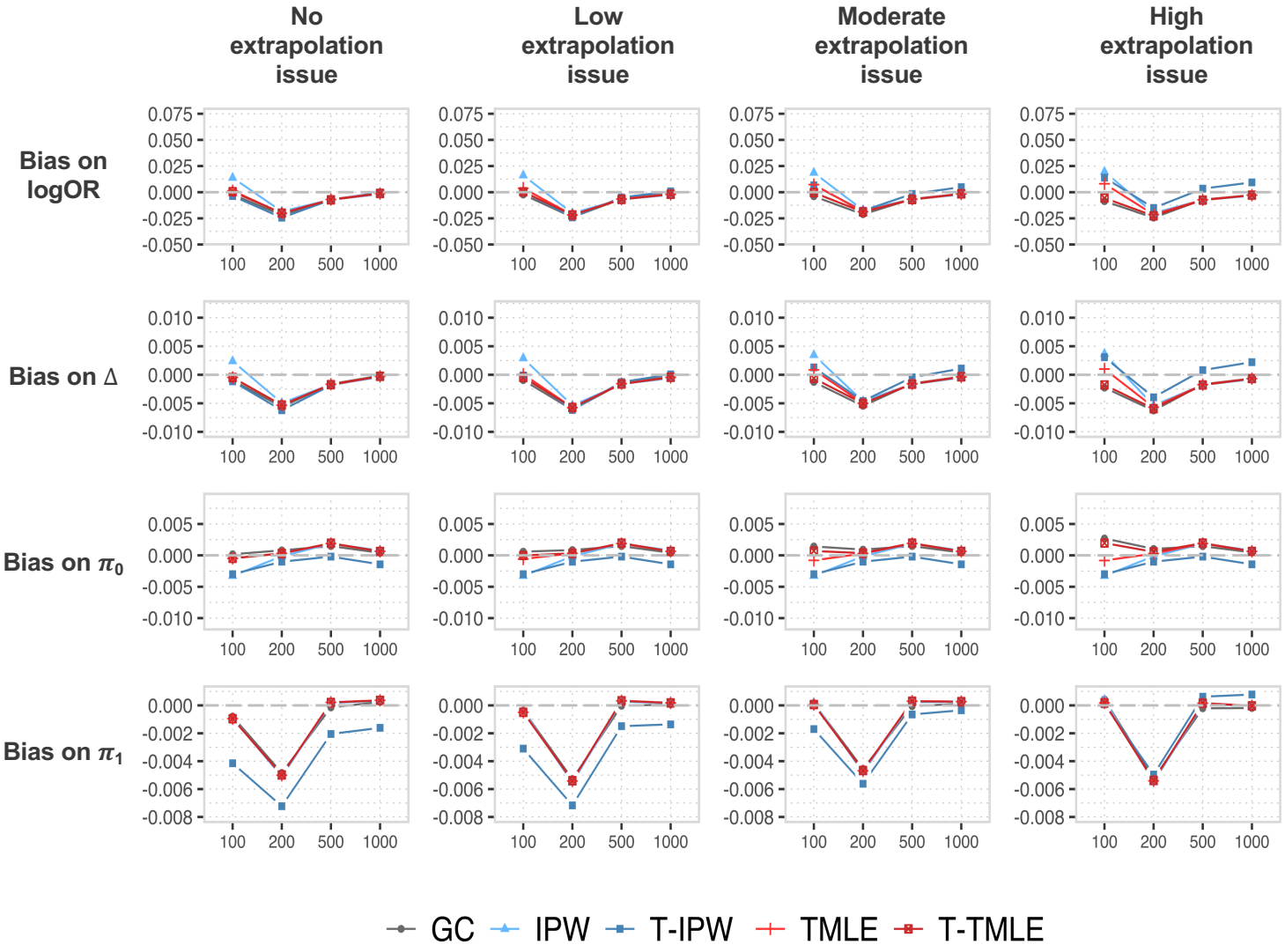


FIGURE S6. The mean absolute bias (y-axis) according to different sample size (from 100 to 1,000, x-axis) and extrapolation issue for an exposure prevalence of 50%. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9); π_1, the expected proportions of event if the entire population is exposed; π_0, the expected proportions of event if the entire population is unexposed; Δ, the corresponding difference (π_1 - π_0); OR, the corresponding odds-ratio.

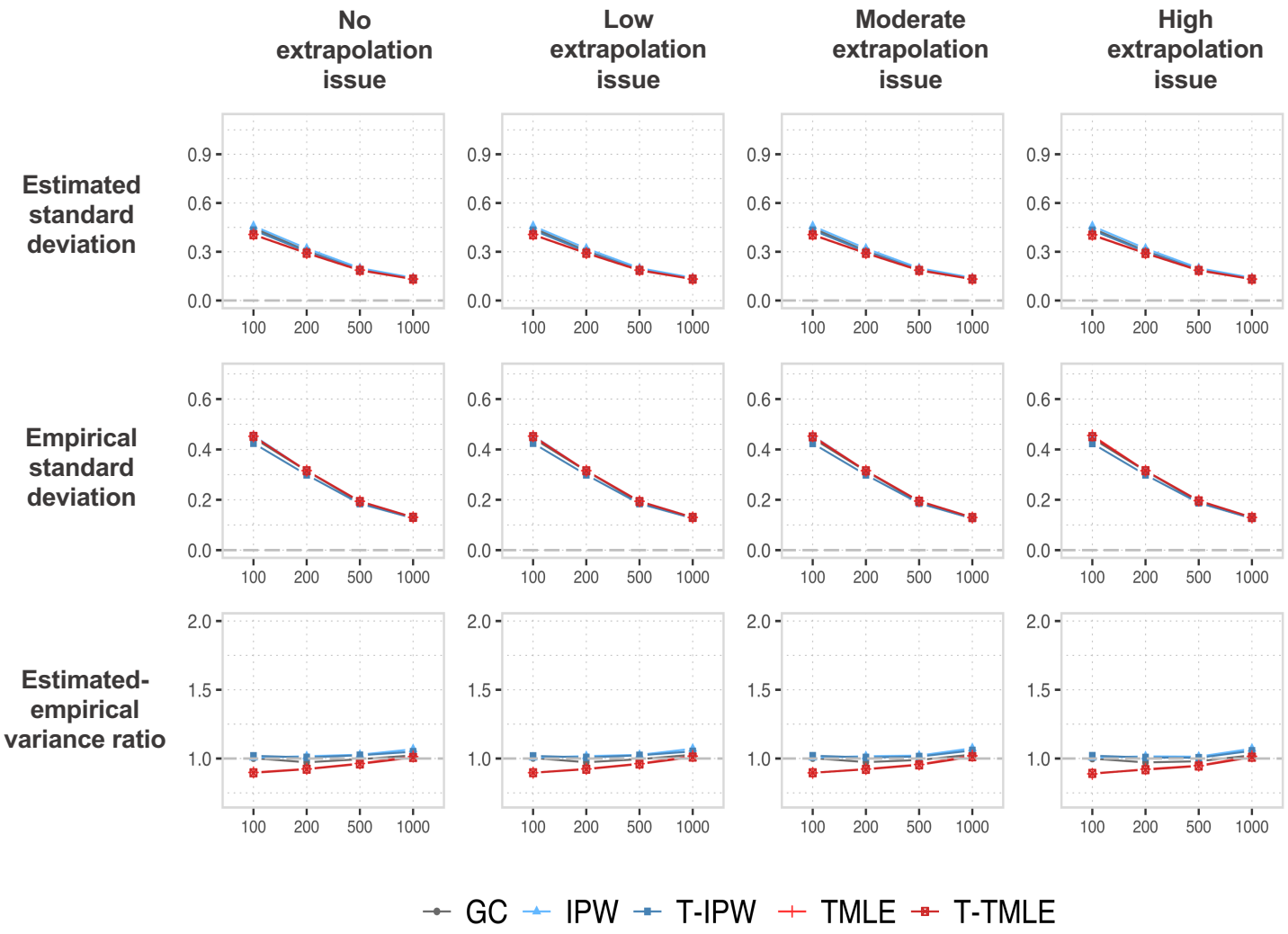


FIGURE S7. Graphical representation of the evolution of accuracy (empirical standard deviation, estimated standard deviation and, the variance estimation ratio) according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue. The target parameter was log(OR). The exposure prevalence was 50%. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9).

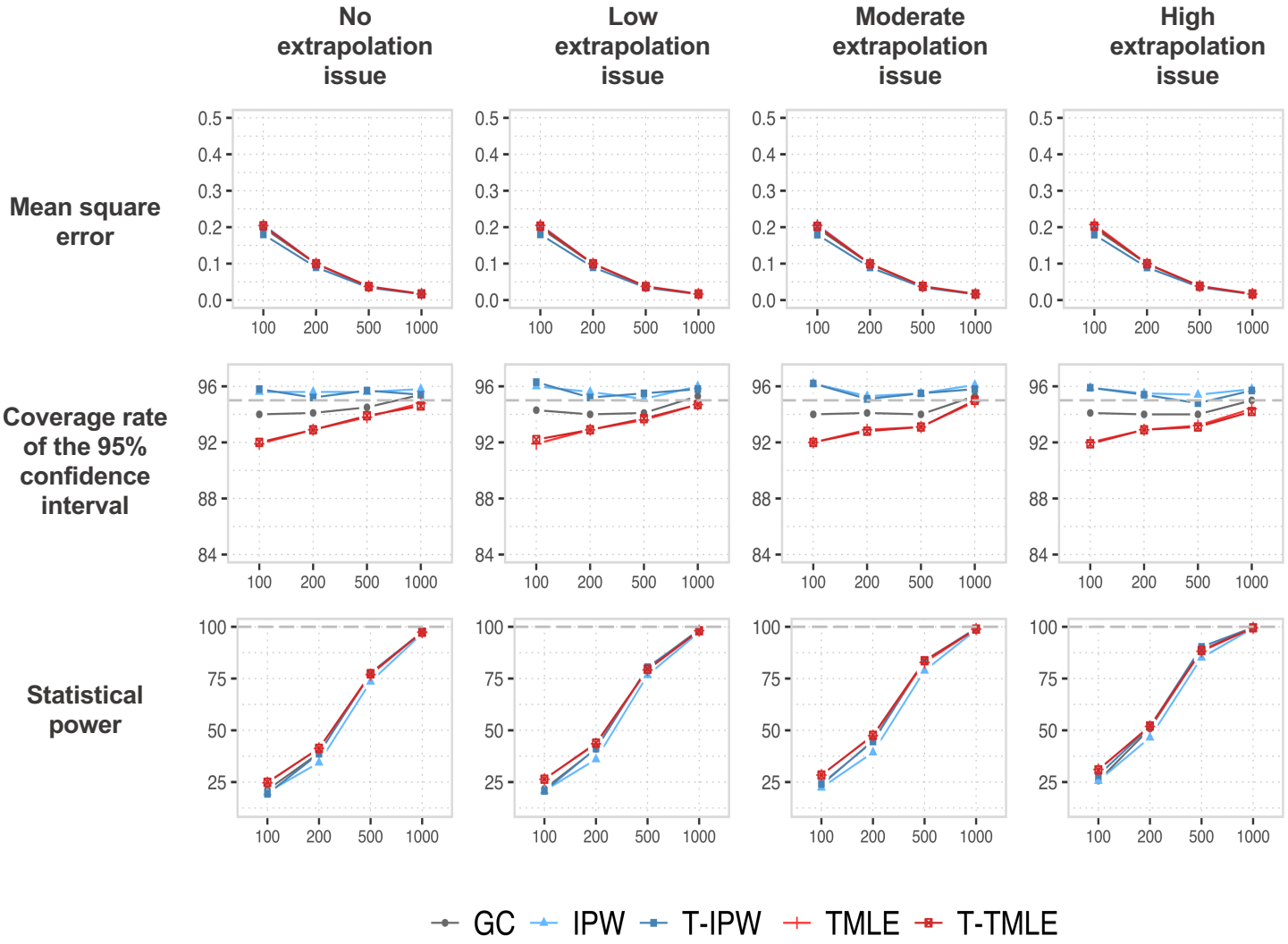


FIGURE S8. The mean square error, the coverage of the 95% confidence interval and the statistical power according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue. The target parameter was log(OR). The exposure prevalence was 50%. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9).

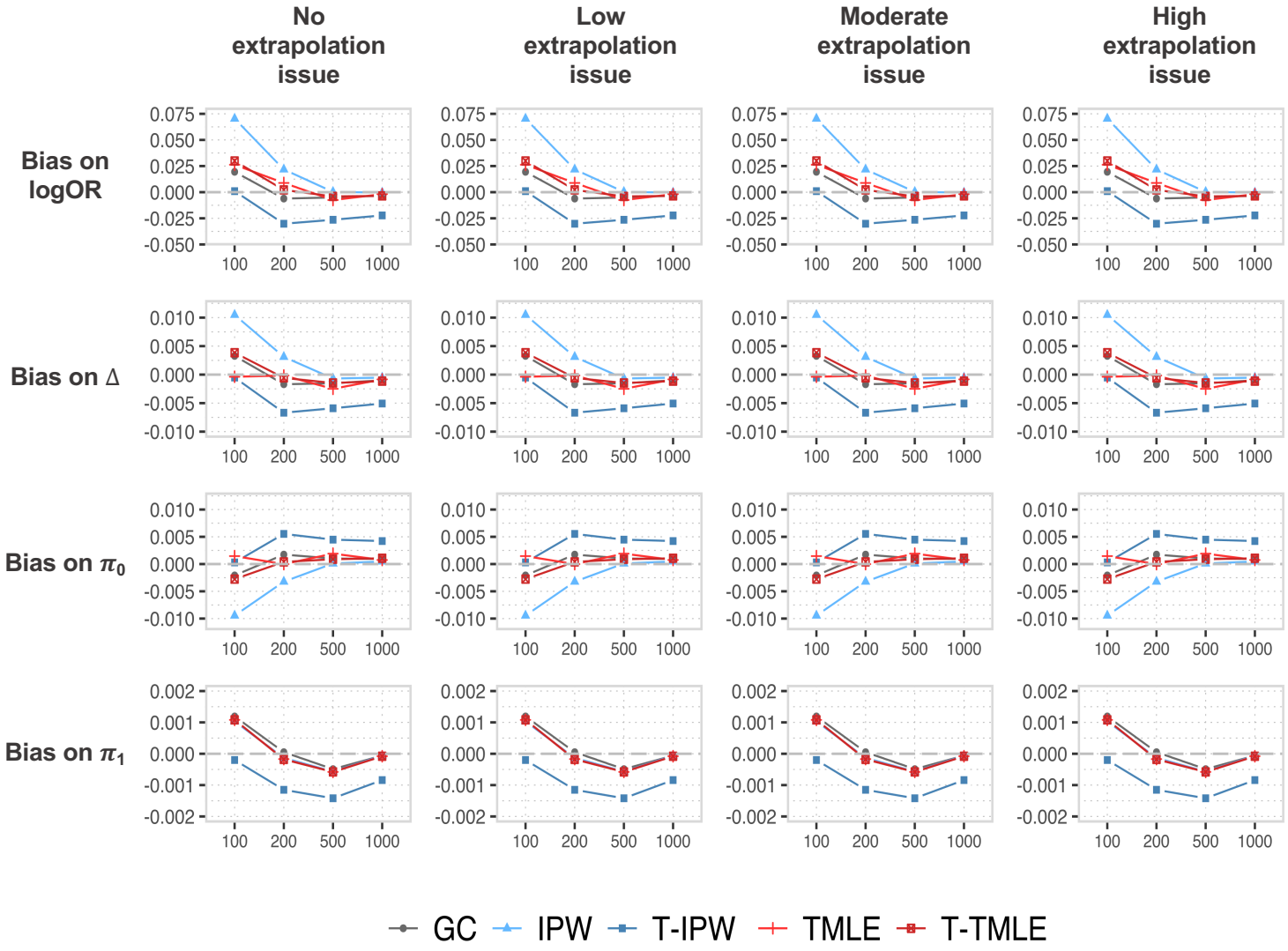


FIGURE S9. The mean absolute bias (y-axis) according to different sample size (from 100 to 1,000, x-axis) and extrapolation issue for an exposure prevalence of 80%, under the null hypothesis. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9); π_1, the expected proportions of event if the entire population is exposed; π_0, the expected proportions of event if the entire population is unexposed; Δ, the corresponding difference (π_1 - π_0); OR, the corresponding odds-ratio.

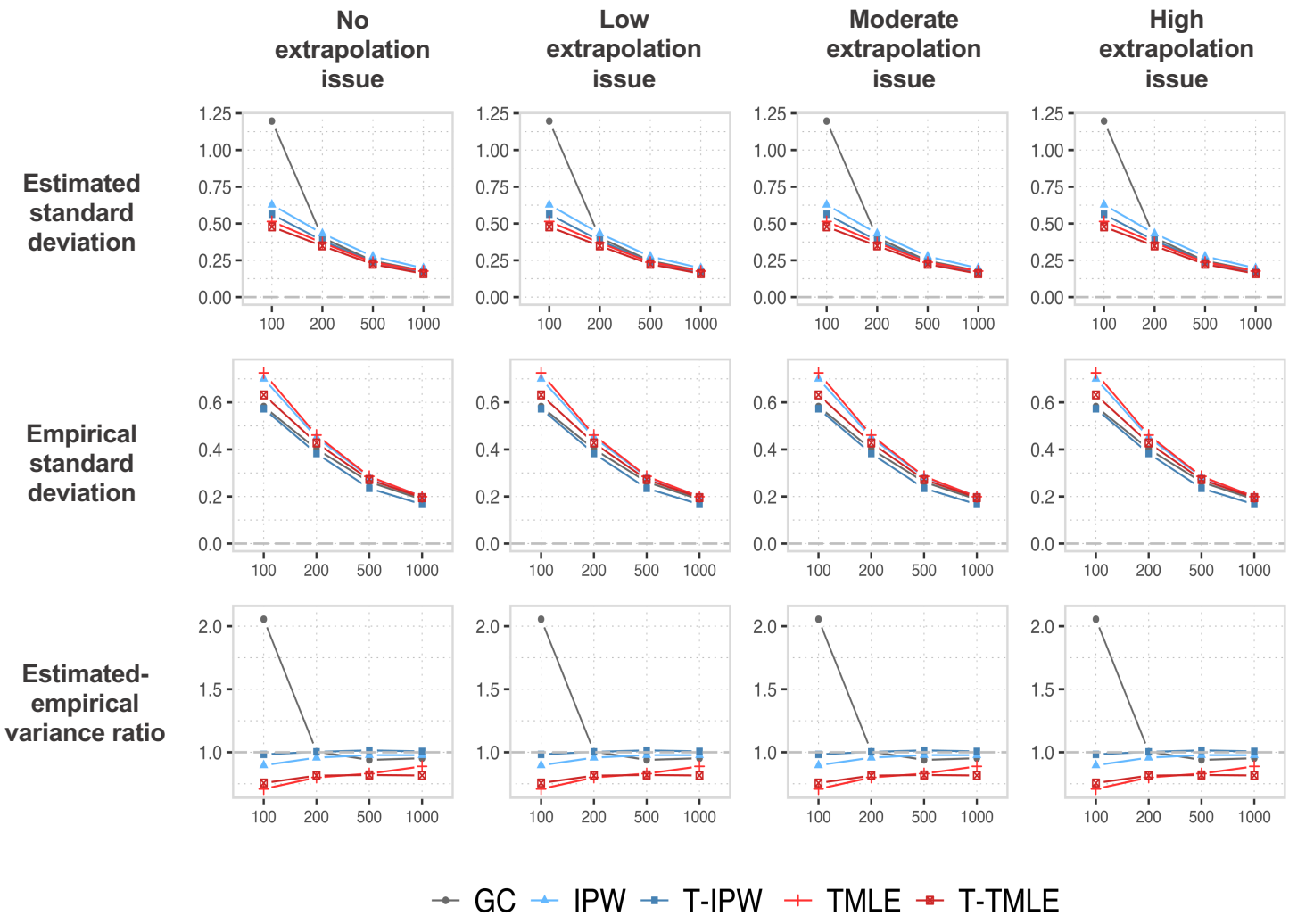


FIGURE S10. Graphical representation of the evolution of accuracy (empirical standard deviation, estimated standard deviation and, the variance estimation ratio) according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue, under the null hypothesis. The target parameter was log(OR). The exposure prevalence was 80%. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9).

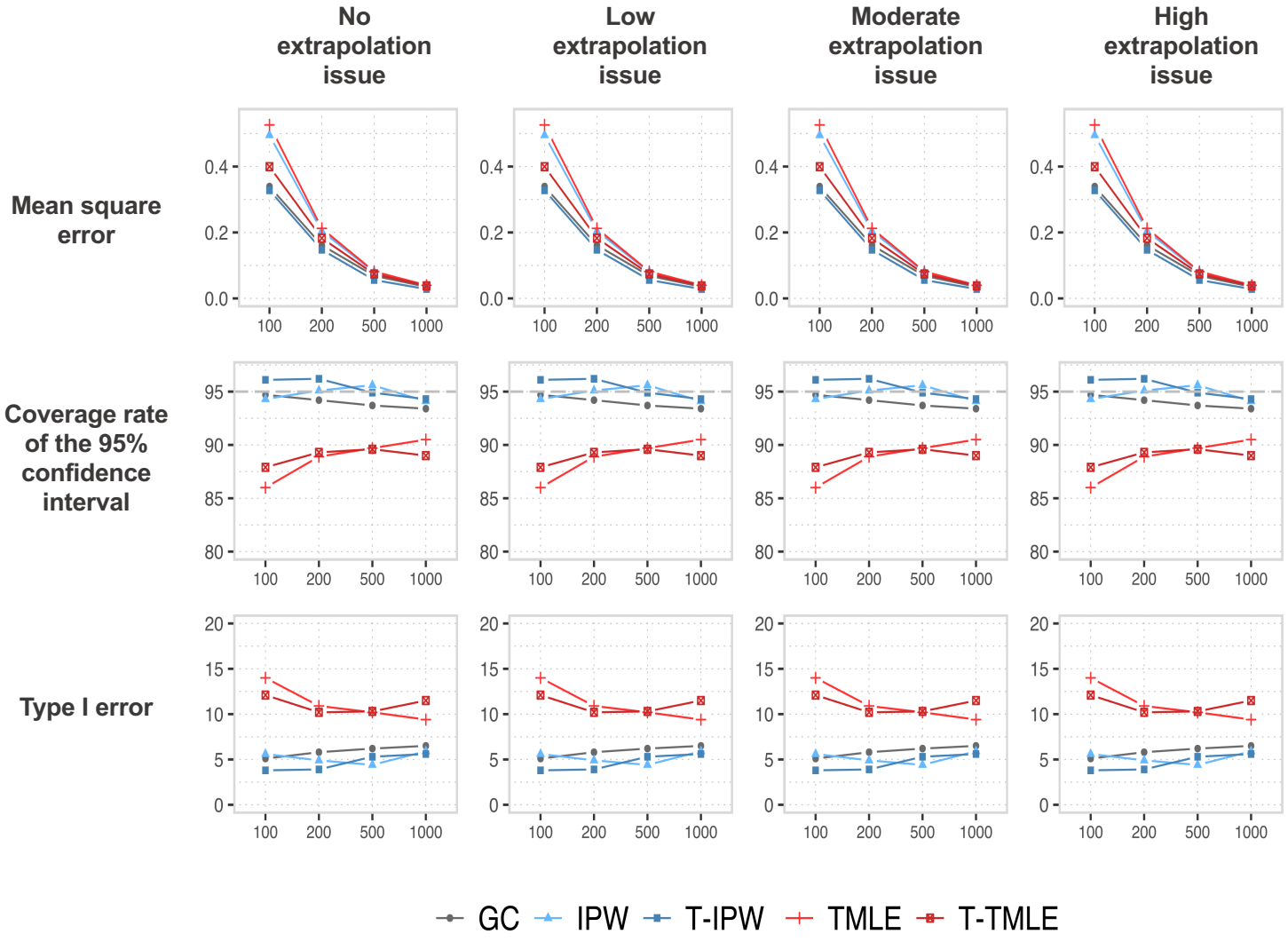


FIGURE S11. The mean square error, the coverage of the 95% confidence interval and the statistical power according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue, under the null hypothesis. The target parameter was log(OR). The exposure prevalence was 80%. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9).

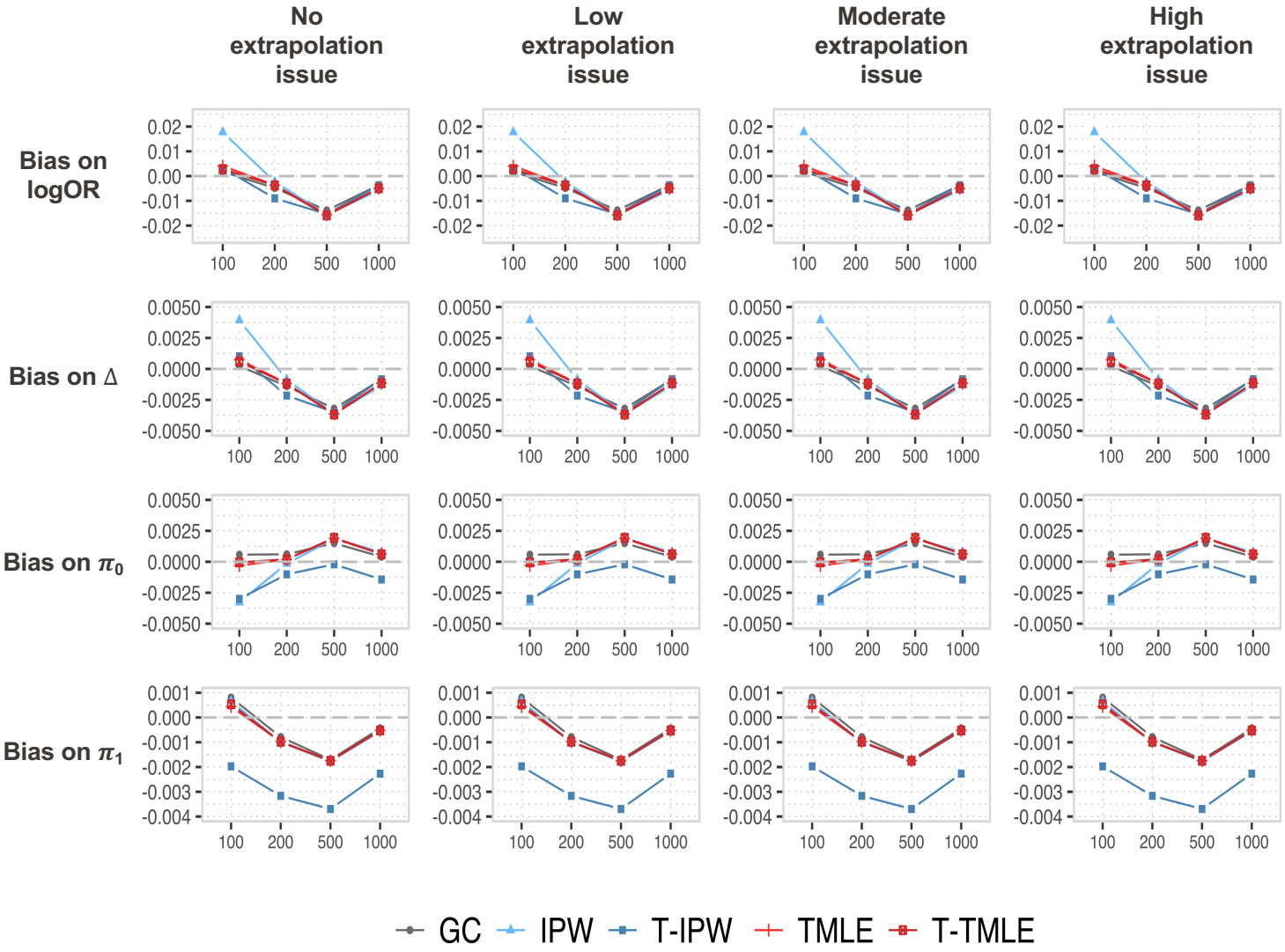


FIGURE S12. The mean absolute bias (y-axis) according to different sample size (from 100 to 1,000, x-axis) and extrapolation issue for an exposure prevalence of 50%, under the null hypothesis. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9); π_1, the expected proportions of event if the entire population is exposed; π_0, the expected proportions of event if the entire population is unexposed; Δ, the corresponding difference (π_1 - π_0); OR, the corresponding odds-ratio.

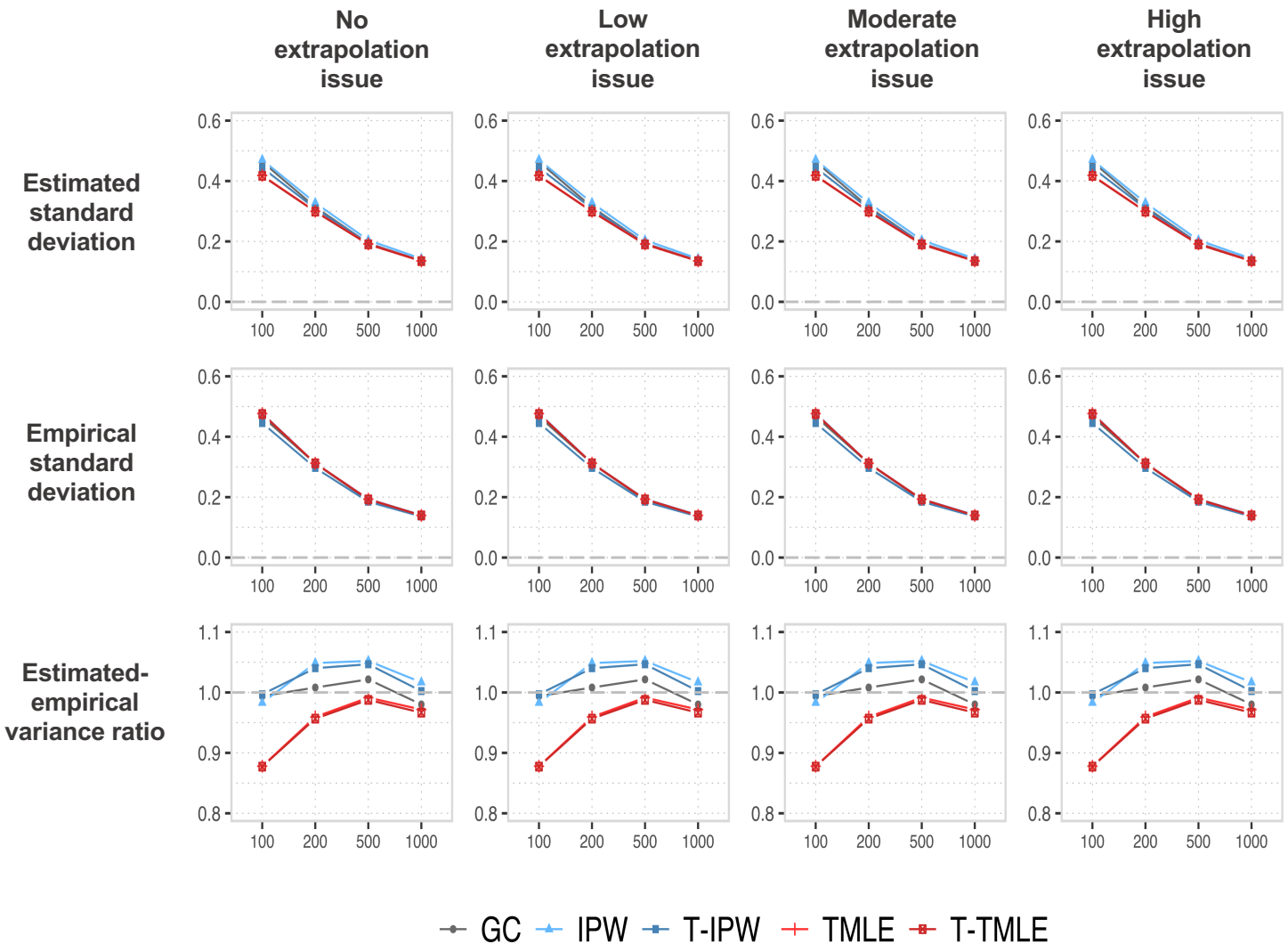


FIGURE S13. Graphical representation of the evolution of accuracy (empirical standard deviation, estimated standard deviation and, the variance estimation ratio) according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue, under the null hypothesis. The target parameter was log(OR). The exposure prevalence was 50%. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9).

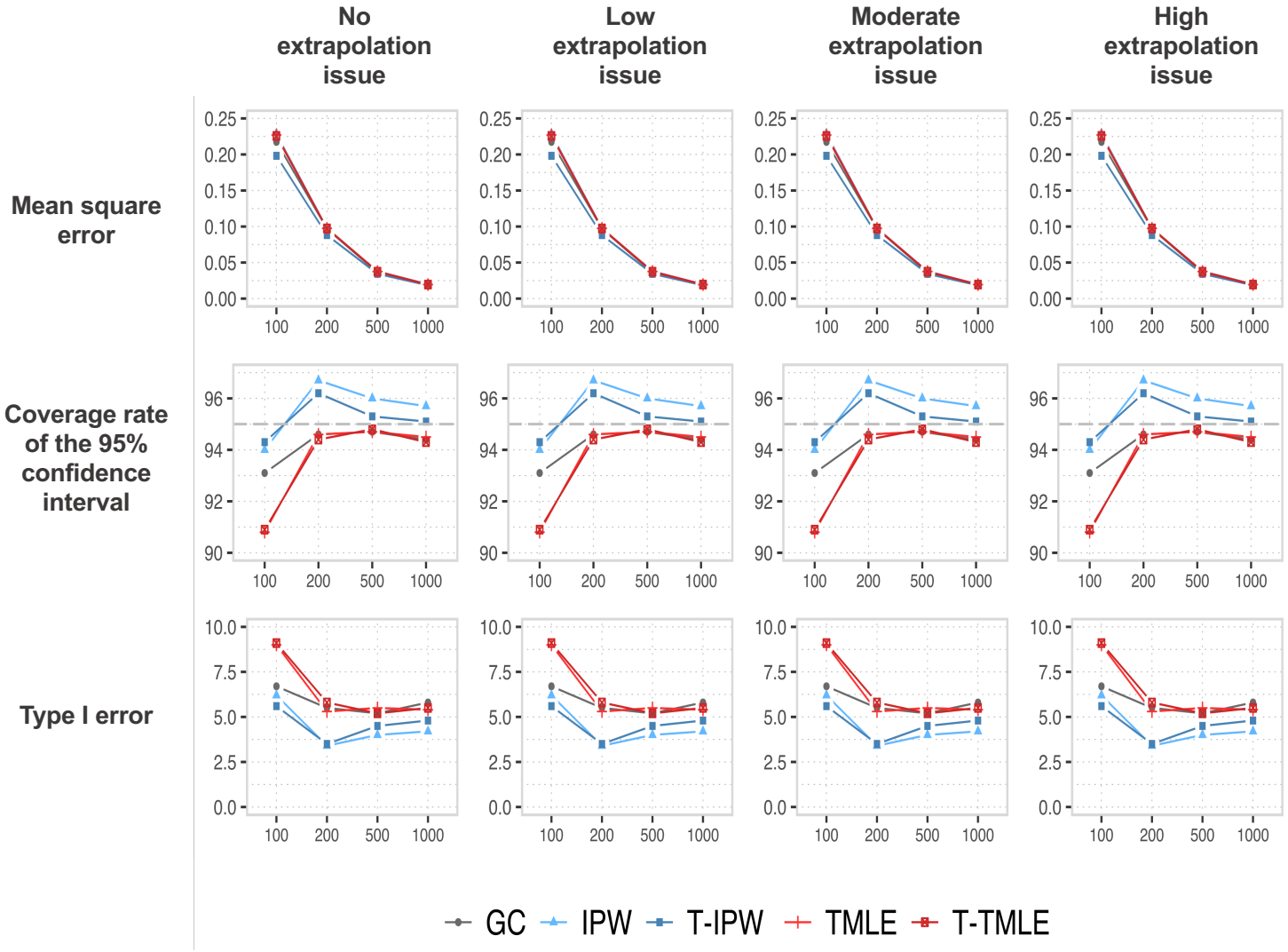


FIGURE S14. The mean square error, the coverage of the 95% confidence interval and the statistical power according to different sample size (from 100 to 1,000, x-axis), in different situations of extrapolation issue, under the null hypothesis. The target parameter was log(OR). The exposure prevalence was 50%. Abbreviations: GC, g-computation; IPW, inverse probability weighting; T-IPW, truncated inverse probability weighting (thresholds: 10^{th} and 90^{th} percentiles) ; TMLE, targeting maximum likelihood estimator; T-TMLE, truncated targeting maximum likelihood estimator (thresholds: bounds at 0.1 and 0.9).

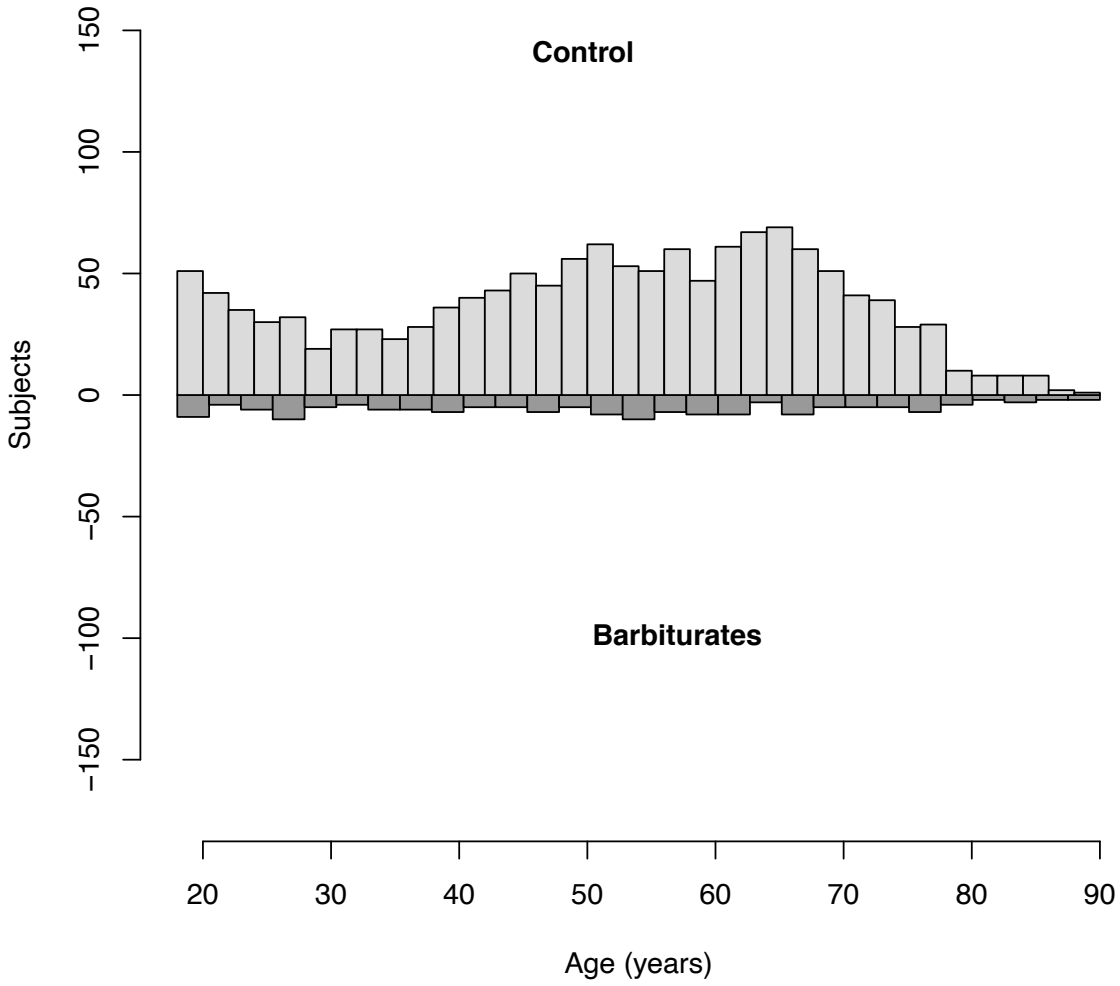


FIGURE S15. Distribution of the patient age in the AtlanREA cohort according to the treatment group: barbiturates and control.

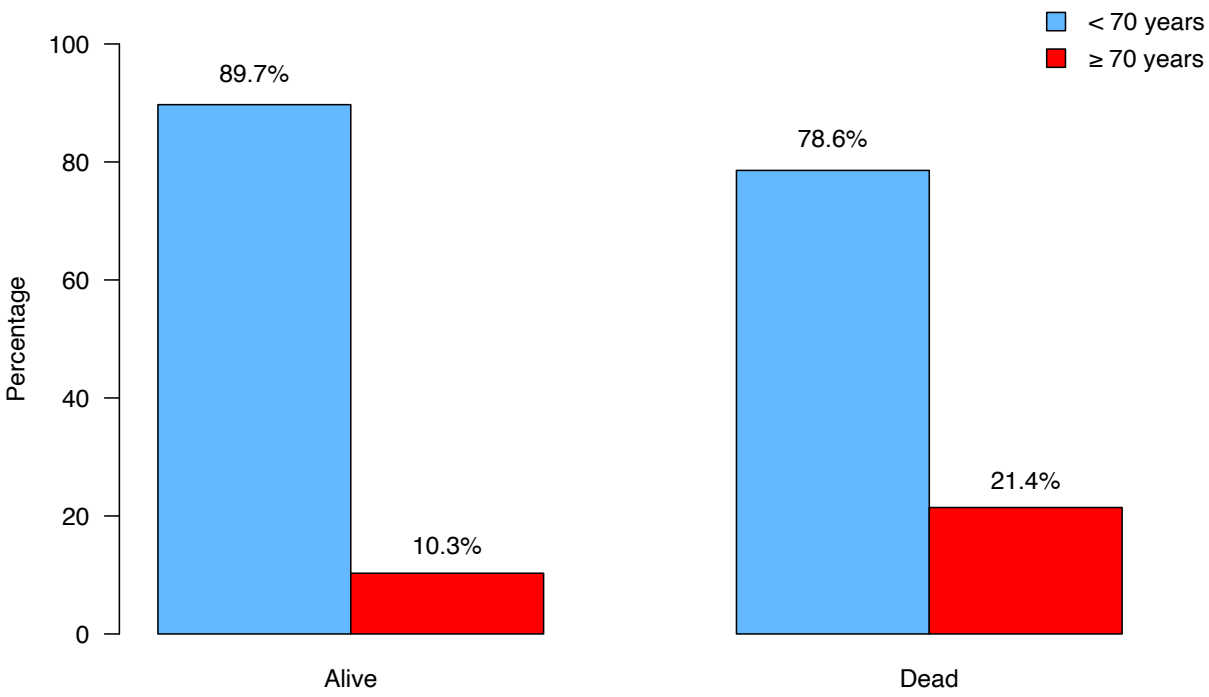


FIGURE S16. Proportion of patients who died in the intensive care unit according to age (binarized at the 70-year threshold).

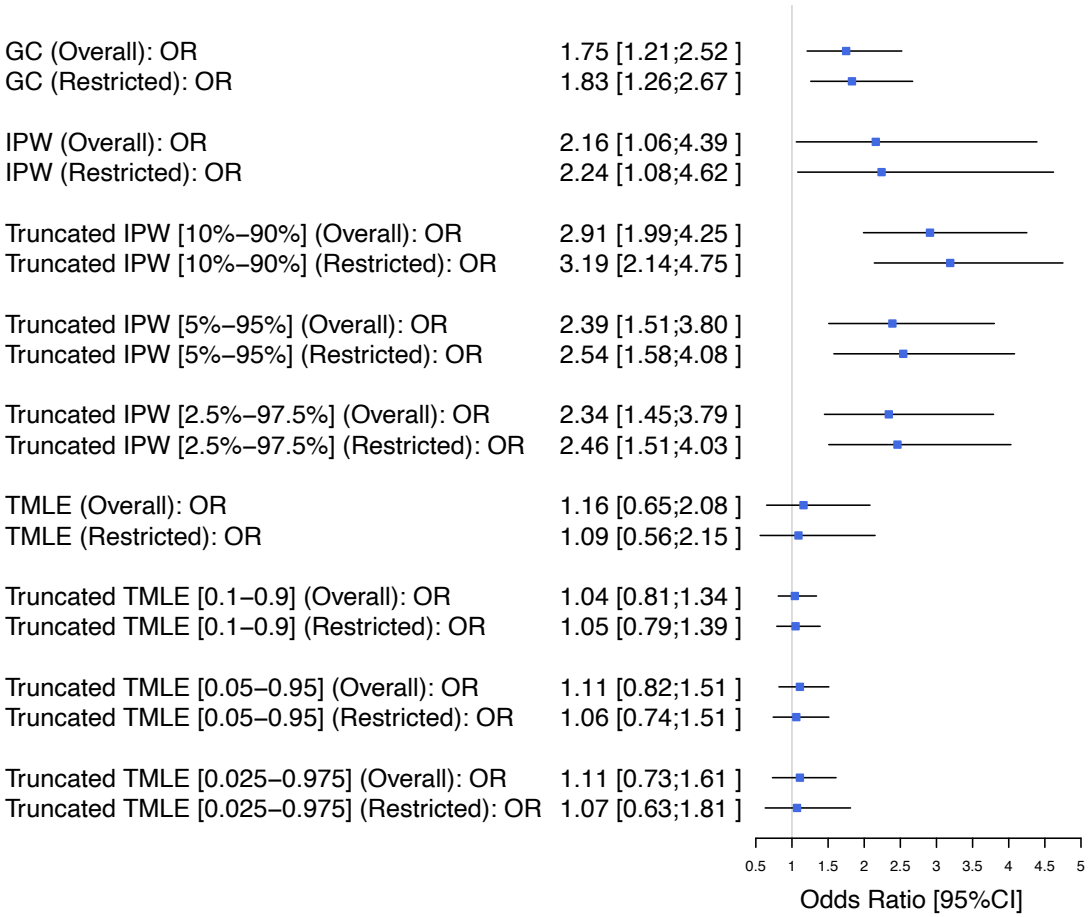


FIGURE S17. Summary table of the estimated odds ratio of observed in-hospital mortality for the group treated with barbiturates according to the different statistical methods between the general population and the age-restricted population in the AtlanREA cohort.

Abbreviations: GC, G-computation; IPW, Inverse Probability Weighting; T-IPW, Truncated Inverse Probability Weighting; TMLE, Targeted Maximum Likelihood Estimation; T-TMLE, Truncated TMLE.

Chapitre 5

Discussion et perspectives

Sommaire

5.1 Résumé des travaux réalisés 108
5.1.1 Analyse de l'impact des barbituriques 108
5.1.2 Problématique de l'hypothèse de positivité 108
5.1.3 Intérêt de la G-Computation 109
5.1.4 Intérêt des méthodes d'apprentissage automatisé 109
5.2 Perspectives pour les bases de données en soins critiques 110
5.2.1 Limites de la constitution de cohortes 110
5.2.2 Challenge informatique 111
5.2.3 Développement d'aides décisionnelles au lit du patient 112

5.1 Résumé des travaux réalisés

5.1.1 Analyse de l'impact des barbituriques

Le cheminement de ces travaux a débuté par l'application d'une méthode d'inférence causale sur les données observationnelles de la cohorte AtlanRéa, et s'intéressant à l'impact de l'utilisation précoce des barbituriques. Au-delà des résultats de cette analyse dans laquelle nous avons identifié une association entre la prescription précoce de barbituriques chez les patients traumatisés craniens et la survenue d'une mortalité hospitalière (hazard ratio de 1.85, avec un intervalle de confiance à 95% [1.03 - 3.33]), différents obstacles ont été rencontrés.

Tout d'abord, différentes limites intrinsèques à la constitution de la cohorte nous ont obligé à affiner notre question causale. En effet, actuellement, la cohorte n'est pas issue d'un recueil régulier et quotidien de différentes covariables, qui nous auraient permis d'étudier l'utilisation des barbituriques à n'importe quel moment du séjour du patient en réanimation. Cette analyse aurait pu être possible en appliquant des méthodes statistiques adpatées aux données longitudinales, comme par exemple le recours à des scores de propension temps-dépendants. [77,78] Nous avons donc dû restreindre notre question à l'utilisation précoce des barbituriques (dans les 24 premières heures suivant l'admission en réanimation).

Par ailleurs, nous nous sommes confrontés à une proportion non négligeable de données manquantes sur certaines variables, nous obligeant à rediscuter leur utilisation dans les modélisations statistiques, ou même la façon d'interpréter nos critères de jugement (le score de l'échelle *Glasgow Outcome Scale* évaluée à 3 mois présentait 26% de données manquantes). Comme la grande majorité des analyses sur données observationnelles qui ont été publiées, [79] nous avons choisi de faire une analyse sur les cas complets (*i.e.* exclusion des cas avec des données manquantes). Ce choix implique une perte de puissance statistique et un risque de biais. Il est important de savoir que différents mécanismes de constitution des données manquantes sont décrits [1], [80], et que différentes méthodes existent pour les gérer. Il existe notamment les différentes types de modélisations permettant de faire de l'imputation multiple pour reconstituer plusieurs jeux de données complets. Des tutoriels sont proposés pour favoriser l'utilisation de ces méthodes dans le cadre de l'analyse de données observationnelles. [81] Qu'importe la méthode utilisée, l'utilisation d'analyses de sensibilité est primordiale. [82]

5.1.2 Problématique de l'hypothèse de positivité

A la lumière des résultats de l'étude de simulation, un point important à souligner est que l'ensemble des méthodes statistiques sont impactées par les situations d'infraction de la positivité, et tout doit être fait par l'investigateur pour limiter ces situations. Les infractions théoriques sont un problème conceptuel de la modélisation de la question causale, alors que les infractions aléatoires pourront être plus ou moins gérées par le choix de la méthode statistique. Dans ces situations, le recours aux méthodes de troncature peut paraitre séduisant pour réduire la variance, mais souvent au prix d'un risque de biais augmenté.

1. 3 mécanismes décrits : manquante de façon aléatoire (*missing at random - MAR*), manquante complètement de façon aléatoire (*missing completely at random - MCAR*), et manquante de façon non-aléatoire (*missing not at random - MNAR*)

L'identification d'infraction de l'hypothèse de positivité peut se faire par exemple, en étudiant graphiquement la distribution des score de propension.La difficulté réside dans l'identification de la strate de population à l'origine de la violation. Pour palier ce problème, en extension des travaux menés par l'équipe, nous avons proposé un algorithme de classification basé sur les arbres de décision pour aider à identifier ces strates (Annexe C). Cet outil permet d'identifier les infractions structurelles comme celles qui sont aléatoires, et la discussion avec le clinicien expert permettra de statuer sur la nécessité de restreindre la population.

5.1.3 Intérêt de la G-Computation

L'intérêt de l'utilisation des méthodes basées sur le SP est la vérification de l'hypothèse de positivité, ce que ne permet pas la GC [18,83]. Mais, la survenue d'une violation devient problématique lors du recours à l'IPTW avec l'apparition de poids extrêmes, augmentant la variance et introduisant du biais. [84]. Au contraire, la GC à la capacité d'extrapoler la distribution du devenir (distribution de l'*outcome model*) dans les zones d'information impactées par la violation aléatoire de la positivité [18]. Les EDR présente la même limite des poids extrêmes dans des situations d'infraction de l'hypothèse de positivité, et une extrapolation dans les strates problématiques comme la GC. De plus, l'utilisation des fonctions d'influence pour la TMLE peut induire une estimation sur-optimiste de la variance dans ces situations d'infraction de la positivité. [12,85]. Nous avons montré que le recours au bootstrap permet de corriger ce problème.

Ainsi, notre travail a mis en évidence la performance des méthodes utilisant un modèle prédisant la survenue de l'évènement (*outcome model*) dans les situations d'infraction aléatoire de l'hypothèse de positivité. Ces résultats sont en accord avec d'autres études montrant une variance plus faible pour la GC dans des situations de positivité respectée [43, 45, 47, 86, 87]. La GC semble être la méthode statistique présentant la meilleure balance entre le biais induit et la variance dans l'estimation, avec une puissance statistique satisfaisante.

5.1.4 Intérêt des méthodes d'apprentissage automatisé

Une difficulté de la modélisation statistique réside dans le choix de l'ensemble de variables d'ajustement. Ces ensembles peuvent contenir des facteurs de risque, des facteurs de confusion et même des variables instrumentales. L'intégration simple des facteurs de risque semble être la meilleure option pour diminuer la variance, et que dans ce cadre le recours à l'*outcome model* est un choix cohérent (Annexe E).

A partir de ces différents résultats et en supposant que les algorithmes de ML pourrait optimiser la spécification des modèles en réduisant l'impact de la fluctuation d'échantillonage sur l'échantillon étudié, nous avons proposé une méthode associant des algorithmes de ML dans le cadre de la GC. Un autre intérêt de la GC réside dans sa modélisation de la survenue de l'évènement, et évite l'intégration de variables instrumentales. L'utilisation des méthodes de ML a pour objectif d'éviter les biais de mauvaise spécification du modèle Q(A, L). [88] Dans ce cadre, les algorithmes de ML nous évitent de devoir formaliser des hypothèses *a priori* sur la structure des données, et ils permettent de modéliser des associations complexes tout en minimisant une fonction de perte pour correspondre aux données. [89] Dans la problématique de l'infraction aléatoire de l'hypothèse

de positivité, l'utilisation des méthodes de ML permettent de rendre robuste l'extrapolation de la relation causale aux strates d'individus avec peu d'information.

Austin a initié cette réflexion de l'intégration d'algorithmes de ML dans le cadre de l'*outcome model* de la GC. [90] Nous avons proposé l'utilisation d'un *Super Learner* (SL) qui pondère les prédictions obtenues par les différents algorithmes (préselectionnés dans une bibliothéque) pour obtenir une estimation finale moyennée qui sera moins biaisée. [91] Dans cette application, l'association du SL avec la GC performait mieux sur le biais et la variance que les autres méthodes, avec une estimation correcte même pour des échantillons de petite taille.

Les limites du ML réside tout d'abord dans les difficultés des modèles à converger (nécessité d'une plus grande quantité de données), [92] mais aussi dans les biais induits par le sur-apprentissage et la régularisation. [93] La régularisation limite le sur-apprentissage en réduisant la variance, mais avec une augmentation du biais.

Le risque de sous-estimation de la variance est d'aboutir à de l'inférence inexacte. A partir de ce constat, plusieurs auteurs mettent en garde sur l'utilisation des méthodes de ML. [14, 92, 94, 95] La TMLE peut surmonter le sur-apprentissage via l'utilisation de SL incluant des approches plus simples [94]. Le SL intègre différents algorithmes qui seront ensuite pondérés, ce qui permet de diminuer le sur-apprentissage. Dans cet objectif, il sera nécessaire d'éviter d'inclure des algorithmes trop flexible, tel que la classification sur des forêts aléatoires (méthodes de *boosting* - BCART) [92,95] L'autre ressource du SL pour éviter le sur-apprentissage est d'utiliser la validation croisée.

5.2 Perspectives pour les bases de données en soins critiques

5.2.1 Limites de la constitution de cohortes

Le recours à l'utilisation de données observationnelles semble être prometteur. Cependant, la création de cohortes s'accompagnent de difficultés dans son élaboration et sa méthodologie de recueil.

Au-delà des questions méthodologiques dans le choix de la population d'étude qui sortent du cadre de cette thèse (*e.g.* les biais de sélection ou la difficulté du calcul d'une taille de population suffisante) , d'autres limites existent dans l'utilisation d'une cohorte pour de l'inférence causale. Le suivi longitudinal d'une cohorte peut entraîner un biais d'attrition avec des patients perdus de vue au cours du temps, et devant faire discuter les choix dans la gestion des données manquantes.

D'autre part, une limite que nous avons rencontrée dans l'utilisation de la cohorte AtlanRéa est la difficulté de gérer la granularité temporelle. Cette problématique est notamment vraie pour les cohortes de soins critiques dont l'évolution clinique des patients suit un référentiel temporel dont l'unité est la journée (à la différence, par exemple, des cohortes de cancérologie suivant les patients mois après mois). Lorsque l'on s'intéresse à des traitements dépendants du temps (*e.g.* une thérapie médicamenteuse donnée sur plusieurs jours), il sera indispensable de gérer les facteurs de confusion dans la même temporalité, et de recueillir l'information concernant ces facteurs jour après jour. En effet, le clinicien décide chaque jour d'arrêter ou de poursuivre son traitement selon différents arguments cliniques. Pour notre premier travail, cette limite nous a contraint à ne nous intéresser qu'à l'utilisation précoce des barbituriques (dans les 24 premières heures). En

effet, la cohorte AtlanRéa n'est pas conçue pour recueillir toutes ces informations au cours du temps. Dans les situations d'inférence causale, un point crucial dans la collection des données longitudinales sera de notifier les différents temps et échéances. La difficulté est de déterminer a posteriori le moment où l'intervention n'a pas été initiée pour le groupe contrôle (si on est dans le cadre d'une comparaison de modalités de traitement, où il existe une modalité « absence de traitement »). [96]

5.2.2 Challenge informatique

Une solution proposée pour pallier à cette difficulté est d'assurer un recueil automatique des données directement dans le dossier électronique, à une fréquence de temps définie pour chaque variable. Un des challenges pour l'informatique va résider dans le stockage des données, car pour donner un ordre d'idée, un monitorage de scope cardiaque à 5 branches génère environ 37 MB de données en 2 heures. [97]

Par ailleurs, un point important va être d'améliorer la qualité de collection des données. Il existe un champ complet de recherche sur l'analyse des variables continues de monitorage de paramètres physiologiques et la gestion du bruit dans la production des données. Au-delà, le processus d'extraction et de collection des données à partir d'un dossier électronique de soins critiques est souvent un processus complexe. Des alternatives récentes visent à automatiser ce processus pour obtenir des variables plus aisément utilisables, comme par exemple avec la méthode du *Critical carE Database for Advanced Research*. [98]

Une autre limite réside dans la faible interopérabilité entre les différentes plateformes et logiciels, avec l'existence de barrières légales et d'accès aux données. Actuellement, seulement une infime part des données de santé générées sont accessible dans un objectif de recherche. La solution proposée pour standardiser les données de santé est de suivre un format commun de codage des jeux de données en santé, le format OMOP-CDM (*Observational Medical Outcomes Partnership – Common Data Model*), qui permet d'homogénéiser les données issues de bases administratives, de dossiers médicaux, ou même de registre de pharmacovigilance. [99] L'initiative OMOP est issue d'un partenariat public-privé, présidé par la *Food and Drug Administration* (autorité de santé aux Etats-Unis) et financé par un consortium de sociétés pharmaceutiques. La mise en place de la standardisation des données de santé au format OMOP est aussi promue par l'union européenne via l'EHDEN (*European Health Data Evidence Network*) pouvant délivrer des financements pour favoriser ces transformations informatiques et la standardisation des données. Pour aller plus loin, le partage de données de manière internationale reste un challenge. En effet, au-delà de l'harmonisation du format des données, le champs légal dépendant de chaque pays reste un frein majeur. [100] Cependant, un exemple récent existe avec la base de données *AmsterdamUMCdb* qui vise à répondre à la fois aux réglementations européennes et à celles américaines. [101]

Une vraie promesse existe dans les soins critiques lorsque les données issues des dispositifs médicaux, moniteurs, et dossiers médicaux seront accessibles et agrégeables. Dans ce contexte, les algorithmes de ML couplés aux principes de l'inférence causale permettront d'approcher des estimations fiables pouvant guider le clinicien quotidiennement. Une aide est apportée par les serveurs de données de santé, de type *cloud*, qui permettent le déploiement de modèle d'apprentissage en temps réel au lit du patient, appelé *online learners*. [102] Cette perspective d'aide décisionnelle alimentée en temps réel est représentée dans la Figure 7.

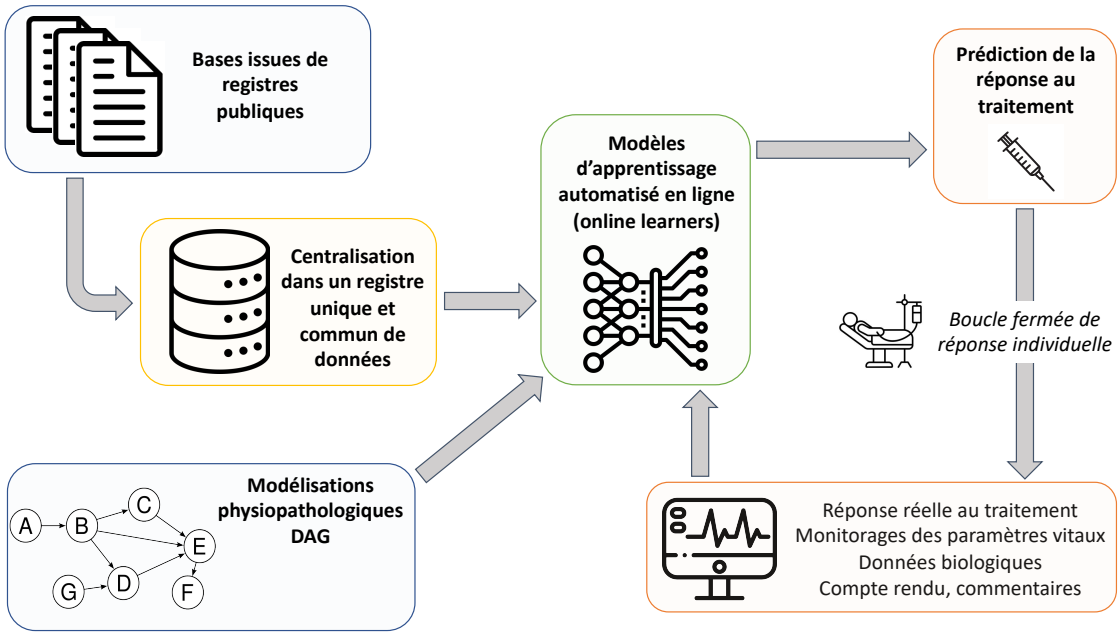


FIGURE 7 – Représentation du processus de mise en place de l'aide décisionnelle en ligne (*online learners*)

5.2.3 Développement d'aides décisionnelles au lit du patient

L'approche de la médecine personnalisée est en marche, et la gestion de ces grands jeux de données ne pourra pas se passer des méthodes algorithmiques basées sur l'apprentissage automatisé. Ainsi, des méthodes de ML peuvent même être imaginées utilisables directement au lit du patient, et en temps réel. L'objectif sera ainsi d'extraire de ces jeux de données l'information nécessaire pour aider le clinicien dans sa décision.

A l'heure actuelle, la plupart des applications cliniques existantes sont issues de modèles prédictifs entrainés sur des cohortes statiques et fermées. Mais le développement de ces technologies nous laisse imaginer à court terme l'émergence de ces plateformes avec des *online learners*, permettant la gestion, l'analyse de données longitudinales stockées en temps réel et permettant d'optimiser la personnalisation des données. Les données enregistrées du patient pourront permettre de prédire la réponse individuelle et spécifique du patient, à partir des données issues de la cohorte ouverte et alimentée en continue, mais aussi par une boucle rétrograde de la réponse du patient aux tests thérapeutiques.

Chapitre 6

Conclusion

La réalisation d'inférence causale sur des données observationnelles apparaît être une alternative intéressante aux essais randomisés, permettant d'effectuer des analyses "en vie réelle". Dans ce cadre, plusieurs hypothèses doivent être vérifiées, dont notamment l'hypothèse de positivité. Ce constat est particulièrement vrai dans le domaine des soins critiques qui a vu la création de nombreuses bases de données ces dernières années.

Le premier travail de cette thèse a analysé l'impact de l'utilisation des barbituriques chez les patients traumatisés crâniens par une analyse sur des données observationnelles issues de la cohorte AtlanRéa. La réalisation de ces analyses a permis de se confronter à la problématique de l'infraction de l'hypothèse de positivité, nous ammenant à comparer différentes méthodes statistiques dans des situations d'infraction à l'hypothèse de positivité. Ce deuxième travail a mis en lumière la robustesse des méthodes statistiques basées sur un *outcome model*, avec notamment un intérêt particulier pour la GC. Cette méthode statistique présente la meilleur balance entre le biais induit et le niveau de variance, tout en préservant la puissance statistique.

L'accumulation de données observationnelles en soins critiques, avec l'apparition de cohorte constituée automatiquement par recueil des données dans le dossier médical électronique ouvre des perspectives aux méthodes statistiques d'inférence causale. Le recours aux algorithmes de ML semble être un atout majeur dans la gestion de l'échantillonage et les problématiques d'extrapolation. Le futur proche de l'utilisation des données observationnelles en soins critiques s'imagine directement auprès du patient en recourant à des aides électroniques à la décision, permettant d'effectuer l'analyse statistique d'inférence causale en temps réel, à partir de n'importe quelle interface connectée, que ce soit un ordinateur ou même un smartphone.

Bibliographie

[1] Westreich D. Epidemiology by design : a causal approach to the health sciences. New York, NY, US : Oxford University Press ; 2020. 4

[2] Sanson-Fisher RW, Bonevski B, Green LW, D'Este C. Limitations of the randomized controlled trial in evaluating population-based health interventions. American Journal of Preventive Medicine. 2007 Aug ;33(2) :155-61. 4

[3] Hernán MA, Robins JM. Causal Inference : What If. Chapman & hall/crc. ed. Chapman & Hall/CRC monographs on statistics & applied probability ; 2020. 4, 6

[4] Hill AB. The Environment and Disease : Association or Causation ? Proceedings of the Royal Society of Medicine. 1965 May ;58(5) :295-300. Available from : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/. 4

[5] Ioannidis JPA. Exposure-wide epidemiology : revisiting Bradford Hill. Statistics in Medicine. 2016 May ;35(11) :1749-62. 4

[6] Holland PW. Statistics and Causal Inference. Journal of the American Statistical Association. 1986 ;81(396) :945-60. Publisher : [American Statistical Association, Taylor & Francis, Ltd.]. Available from : https://www.jstor.org/stable/2289064. 4

[7] Shimonovich M, Pearce A, Thomson H, Keyes K, Katikireddi SV. Assessing causality in epidemiology : revisiting Bradford Hill to incorporate developments in causal thinking. European Journal of Epidemiology. 2021 Sep ;36(9) :873-87. 4

[8] Rubin DB. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. Journal of Educational Psychology. 1974 Apr ;70(1) :41-55. Available from : https://dash.harvard.edu/handle/1/3408692. 7

[9] Hernán MA. A definition of causal effect for epidemiological research. Journal of Epidemiology and Community Health. 2004 Apr ;58(4) :265-71. 7

[10] Lundberg I, Johnson R, Stewart BM. What Is Your Estimand ? Defining the Target Quantity Connects Statistical Evidence to Theory. American Sociological Review. 2021 Jun ;86(3) :532-65. Publisher : SAGE Publications Inc. Available from : https://doi.org/10.1177/00031224211004187. 7

[11] Imbens G. Nonparametric Estimation of Average Treatment Effects under Exogeneity : A Review. Review of Economics and Statistics. 2004 Feb ;86(1) :4-29. 8

[12] Petersen ML, van der Laan MJ. Causal models and learning from data : integrating causal modeling and statistical estimation. Epidemiology (Cambridge, Mass). 2014 May ;25(3) :418-26. 8, 109

[13] Pearl J. Causality : Models, reasoning, and inference. Causality : Models, reasoning, and inference. New York, NY, US : Cambridge University Press ; 2000. Pages : xvi, 384. 8

[14] Díaz I. Machine learning in the estimation of causal effects : targeted minimum loss-based estimation and double/debiased machine learning. Biostatistics (Oxford, England). 2020 Apr;21(2) :353-8. 8, 110

[15] Hernán MA. Comment : Spherical Cows in a Vacuum : Data Analysis Competitions for Causal Inference. Statistical Science. 2019 Feb;34(1) :69-71. Publisher : Institute of Mathematical Statistics. Available from : https://projecteuclid.org/journals/statistical-science/volume-34/issue-1/Comment--Spherical-Cows-in-a-Vacuum--Data-Analysis/10.1214/18-STS684.full. 8

[16] Dawid AP. Conditional Independence in Statistical Theory. Journal of the Royal Statistical Society : Series B (Methodological). 1979;41(1) :1-15. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1979.tb01052.x. Available from : https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1979.tb01052.x. 9

[17] Westreich D, Cole SR. Invited commentary : positivity in practice. American Journal of Epidemiology. 2010 Mar;171(6) :674-7 ; discussion 678-81. 9, 11, 12

[18] Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. Statistical Methods in Medical Research. 2012 Feb;21(1) :31-54. 9, 10, 12, 109

[19] Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983. 10, 11, 13

[20] Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling. 1986 Jan;7(9) :1393-512. Available from : http://www.sciencedirect.com/science/article/pii/0270025586900886. 10, 14

[21] Vansteelandt S, Keiding N. Invited Commentary : G-Computation–Lost in Translation ? American Journal of Epidemiology. 2011 Apr;173(7) :739-42. Available from : https://academic.oup.com/aje/article/173/7/739/104202. 10

[22] Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set : demonstration of a causal inference technique. American Journal of Epidemiology. 2011 Apr;173(7) :731-8. 10, 15

[23] Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. American Journal of Epidemiology. 2011 Apr;173(7) :761-7. 10, 15

[24] Pirracchio R, Carone M. The Balance Super Learner : A robust adaptation of the Super Learner to improve estimation of the average treatment effect in the treated based on propensity score matching. Statistical Methods in Medical Research. 2018 Aug;27(8) :2504-18. 11

[25] Collier ZK, Leite WL, Zhang H. Estimating propensity scores using neural networks and traditional methods : a comparative simulation study. Communications in Statistics - Simulation and Computation. 2021 Aug;0(0) :1-16. Publisher : Taylor & Francis _eprint : https ://doi.org/10.1080/03610918.2021.1963455. Available from : https://doi.org/10.1080/03610918.2021.1963455. 11

[26] Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology (Cambridge, Mass). 2000 Sep;11(5) :550-60. 11, 14

[27] Drake C. Effects of Misspecification of the Propensity Score on Estimators of Treatment Effect. Biometrics. 1993 ;49(4) :1231-6. Publisher : [Wiley, International Biometric Society]. Available from : https://www.jstor.org/stable/2532266. 11

[28] Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. American Journal of Epidemiology. 2008 Sep ;168(6) :656-64. 12, 16

[29] Bahamyirou A, Blais L, Forget A, Schnitzer ME. Understanding and diagnosing the potential for bias when using machine learning methods with doubly robust causal estimators. Statistical Methods in Medical Research. 2019 Jun ;28(6) :1637-50. 12

[30] Austin PC. The relative ability of different propensity score methods to balance measured covariates between treated and untreated subjects in observational studies. Medical Decision Making : An International Journal of the Society for Medical Decision Making. 2009 Dec ;29(6) :661-77. 13

[31] Garrido MM. Covariate Adjustment and Propensity Scores. JAMA. 2016 Apr ;315(14) :1521-2. Available from : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6380669/. 13

[32] Austin PC. A comparison of 12 algorithms for matching on the propensity score. Statistics in Medicine. 2014 Mar ;33(6) :1057-69. 13

[33] Lunt M. Selecting an Appropriate Caliper Can Be Essential for Achieving Good Balance With Propensity Score Matching. American Journal of Epidemiology. 2014 Jan ;179(2) :226-35. Available from : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3873103/. 13

[34] Rassen JA, Shelat AA, Myers J, Glynn RJ, Rothman KJ, Schneeweiss S. One-to-many propensity score matching in cohort studies. Pharmacoepidemiology and Drug Safety. 2012 May ;21 Suppl 2 :69-80. 13

[35] Stuart EA, Lee BK, Leacy FP. Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. Journal of Clinical Epidemiology. 2013 Aug ;66(8 Suppl) :S84-90.e1. 13, 17

[36] Glynn RJ, Lunt M, Rothman KJ, Poole C, Schneeweiss S, Stürmer T. Comparison of alternative approaches to trim subjects in the tails of the propensity score distribution. Pharmacoepidemiology and Drug Safety. 2019 Oct ;28(10) :1290-8. 14, 16

[37] Yoshida K, Solomon DH, Haneuse S, Kim SC, Patorno E, Tedeschi SK, et al. A tool for empirical equipoise assessment in multigroup comparative effectiveness research. Pharmacoepidemiology and Drug Safety. 2019 Jul ;28(7) :934-41. 14

[38] Austin PC, Stuart EA. Estimating the effect of treatment on binary outcomes using full matching on the propensity score. Statistical Methods in Medical Research. 2017 Dec ;26(6) :2505-25. 14

[39] Hirano K, Imbens GW. Estimation of Causal Effects using Propensity Score Weighting : An Application to Data on Right Heart Catheterization. Health Services and Outcomes Research Methodology. 2001 Dec ;2(3) :259-78. Available from : https://doi.org/10.1023/A:1020371312283. 14

[40] Li F, Thomas LE, Li F. Addressing Extreme Propensity Scores via the Overlap Weights ;188(1) :250-7. 14

[41] Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value in Health : The Journal of the International Society for Pharmacoeconomics and Outcomes Research. 2010 Apr ;13(2) :273-7. 14

[42] Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Statistics in Medicine. 2016 Dec;35(30) :5642-55. 14

[43] Lendle SD, Fireman B, van der Laan MJ. Targeted maximum likelihood estimation in safety analysis. Journal of Clinical Epidemiology. 2013 Aug;66(8 Suppl) :S91-8. 15, 16, 109

[44] van der Laan MJ, Rubin D. Targeted Maximum Likelihood Learning;2(1). Available from : https://www.degruyter.com/view/j/ijb.2006.2.1/ijb.2006.2.1.1043/ijb.2006.2.1.1043.xml. 15

[45] Porter KE, Gruber S, van der Laan MJ, Sekhon JS. The relative performance of targeted maximum likelihood estimators. The International Journal of Biostatistics. 2011;7(1) :Article 31. 15, 109

[46] Hampel FR. The Influence Curve and Its Role in Robust Estimation. Journal of the American Statistical Association. 1974;69(346) :383-93. Publisher : [American Statistical Association, Taylor & Francis, Ltd.]. Available from : https://www.jstor.org/stable/2285666. 16

[47] Kang JDY, Schafer JL. Demystifying Double Robustness : A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data;22(4) :523-39. Available from : https://projecteuclid.org/euclid.ss/1207580167. 16, 109

[48] Lee BK, Lessler J, Stuart EA. Weight Trimming and Propensity Score Weighting. PLoS ONE. 2011 Mar;6(3) :e18174. Available from : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3069059/. 16

[49] Ju C, Schwab J, van der Laan MJ. On adaptive propensity score truncation in causal inference;28(6) :1741-60. 16

[50] Bembom O, Laan Mvd. Data-adaptive selection of the truncation level for Inverse-Probability-of-Treatment-Weighted estimators. Available from : https://biostats.bepress.com/ucbbiostat/paper230. 16

[51] Ali MS, Groenwold RHH, Belitser SV, Pestman WR, Hoes AW, Roes KCB, et al. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal : a systematic review;68(2) :112-21. 17

[52] Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, et al. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly : a matched analysis using propensity scores;54(4) :387-98. 17

[53] Imai K, King G, Stuart EA. Misunderstandings between experimentalists and observationalists about causal inference :22. 17

[54] Rubin DB. The design versus the analysis of observational studies for causal effects : parallels with the design of randomized trials. Statistics in Medicine. 2007 Jan;26(1) :20-36. 17

[55] Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation : a simulation study. Pharmacoepidemiology and Drug Safety. 2008 Jun;17(6) :546-55. 17

[56] Westreich D, Lessler J, Funk MJ. Propensity score estimation : neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. Journal of Clinical Epidemiology. 2010 Aug;63(8) :826-33. 17

[57] Low YS, Gallego B, Shah NH. Comparing high-dimensional confounder control methods for rapid cohort studies from electronic health records. Journal of Comparative Effectiveness Research. 2016 Mar ;5(2) :179-92. 17

[58] Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Statistics in Medicine. 2010 Feb ;29(3) :337-46. 17

[59] Wyss R, Lunt M, Brookhart MA, Glynn RJ, Stürmer T. Reducing Bias Amplification in the Presence of Unmeasured Confounding Through Out-of-Sample Estimation Strategies for the Disease Risk Score. Journal of Causal Inference. 2014 Sep ;2(2) :131-46. 17

[60] Colson KE, Rudolph KE, Zimmerman SC, Goin DE, Stuart EA, Laan Mvd, et al. Optimizing matching and analysis combinations for estimating causal effects. Scientific Reports. 2016 Mar ;6(1) :23222. Number : 1 Publisher : Nature Publishing Group. Available from : https://www.nature.com/articles/srep23222. 17

[61] Caruana E, Chevret S, Resche-Rigon M, Pirracchio R. A new weighted balance measure helped to select the variables to be included in a propensity score model ;68(12) :1415-22.e2. 17

[62] Vincent JL. We should abandon randomized controlled trials in the intensive care unit ;38(10) :S534-8. 20

[63] Granholm A, Alhazzani W, Derde LPG, Angus DC, Zampieri FG, Hammond NE, et al. Randomised clinical trials in critical care : past, present and future ;48(2) :164-78. Available from : https://doi.org/10.1007/s00134-021-06587-9. 20

[64] Harhay MO, Wagner J, Ratcliffe SJ, Bronheim RS, Gopal A, Green S, et al. Outcomes and statistical power in adult critical care randomized trials ;189(12) :1469-78. 20

[65] Bailly S, Pirracchio R, Timsit JF. What's new in the quantification of causal effects from longitudinal cohort studies : a brief introduction to marginal structural models for intensivists. Intensive Care Medicine. 2016 Apr ;42(4) :576-9. 21

[66] Etminan M, Collins GS, Mansournia MA. Using Causal Diagrams to Improve the Design and Interpretation of Medical Research. Chest. 2020 Jul ;158(1S) :S21-8. 21

[67] Johnson AEW, Pollard TJ, Shen L, Lehman LWH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Scientific Data. 2016 May ;3 :160035. 21

[68] Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research ;5(1) :180178. Number : 1 Publisher : Nature Publishing Group. Available from : https://www.nature.com/articles/sdata2018178. 22

[69] Seymour CW, Liu VX, Iwashyna TJ, Brunkhorst FM, Rea TD, Scherag A, et al. Assessment of Clinical Criteria for Sepsis : For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3) ;315(8) :762-74. Available from : https://doi.org/10.1001/jama.2016.0288. 22

[70] The ARDS Definition Task Force*. Acute Respiratory Distress Syndrome : The Berlin Definition ;307(23) :2526-33. Available from : https://doi.org/10.1001/jama.2012.5669. 22

[71] Stow PJ, Hart GK, Higlett T, George C, Herkes R, McWilliam D, et al. Development and implementation of a high-quality clinical database : the Australian and New Zealand Intensive Care Society Adult Patient Database. Journal of Critical Care. 2006 Jun ;21(2) :133-41. 22

[72] van de Klundert N, Holman R, Dongelmans DA, de Keizer NF. Data Resource Profile : the Dutch National Intensive Care Evaluation (NICE) Registry of Admissions to Adult Intensive Care Units. International Journal of Epidemiology. 2015 Dec;44(6) :1850-1850h. 22

[73] Harris S, Shi S, Brealey D, MacCallum NS, Denaxas S, Perez-Suarez D, et al. Critical Care Health Informatics Collaborative (CCHIC) : Data, tools and methods for reproducible research : A multi-centre UK intensive care database;112 :82-9. 22

[74] COVID-ICU Group on behalf of the REVA Network and the COVID-ICU Investigators. Clinical characteristics and day-90 outcomes of 4244 critically ill adults with COVID-19 : a prospective cohort study;47(1) :60-73. 23

[75] Gayat E, Cariou A, Deye N, Vieillard-Baron A, Jaber S, Damoisel C, et al. Determinants of long-term outcome in ICU survivors : results from the FROG-ICU study;22(1) :8. 23

[76] Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods;38(11) :2074-102. Available from : https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8086. 59

[77] Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men;11(5) :561-70. 108

[78] Lu B. Propensity score matching with time-dependent covariates;61(3) :721-8. 108

[79] Eekhout I, de Boer RM, Twisk JWR, de Vet HCW, Heymans MW. Missing data : a systematic review of how they are reported and handled;23(5) :729-32. 108

[80] Rubin DB. Inference and Missing Data;63(3) :581-92. Publisher : [Oxford University Press, Biometrika Trust]. Available from : https://www.jstor.org/stable/2335739. 108

[81] Lee KJ, Tilling KM, Cornish RP, Little RJA, Bell ML, Goetghebeur E, et al. Framework for the treatment and reporting of missing data in observational studies : The Treatment And Reporting of Missing data in Observational Studies framework;134 :79-88. 108

[82] National Research Council (US) Panel on Handling Missing Data in Clinical Trials. The Prevention and Treatment of Missing Data in Clinical Trials. National Academies Press (US);. Available from : http://www.ncbi.nlm.nih.gov/books/NBK209904/. 108

[83] Zhu Y, Mitra N, Roy J. Addressing Positivity Violations in Causal Effect Estimation using Gaussian Process Priors. Available from : http://arxiv.org/abs/2110.10266. 109

[84] Webster-Clark M, Stürmer T, Wang T, Man K, Marinac-Dabic D, Rothman KJ, et al. Using propensity scores to estimate effects of treatment initiation decisions : State of the science;40(7) :1718-35. _eprint : https ://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8866. Available from : https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8866. 109

[85] Schnitzer ME, Lok JJ, Gruber S. Variable Selection for Confounder Control, Flexible Modeling and Collaborative Targeted Minimum Loss-Based Estimation in Causal Inference;12(1) :97-115. 109

[86] Neugebauer R, van der Laan M. Why prefer double robust estimators in causal inference?;129(1) :405-26. Available from : http://www.sciencedirect.com/science/article/pii/S0378375804002022. 109

[87] Moore KL, Neugebauer R, van der Laan MJ, Tager IB. Causal inference in epidemiological studies with strong confounding ;31(13) :1380-404. 109

[88] Blakely T, Lynch J, Simons K, Bentley R, Rose S. Reflection on modern methods : when worlds collide-prediction, machine learning and causal inference ;49(6) :2058-64. 109

[89] Bi Q, Goodman KE, Kaminsky J, Lessler J. What is Machine Learning ? A Primer for the Epidemiologist ;188(12) :2222-39. 109

[90] Austin PC. Using Ensemble-Based Methods for Directly Estimating Causal Effects : An Investigation of Tree-Based G-Computation ;47(1) :115-35. Available from : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3293511/. 110

[91] Naimi AI, Balzer LB. Stacked generalization : an introduction to super learning ;33(5) :459-64. 110

[92] Naimi AI, Mishler AE, Kennedy EH. Challenges in Obtaining Valid Causal Effect Estimates with Machine Learning Algorithms :kwab201. 110

[93] Mullainathan S, Spiess J. Machine Learning : An Applied Econometric Approach ;31(2) :87-106. Available from : https://www.aeaweb.org/articles?id=10.1257/jep.31.2.87. 110

[94] Balzer LB, Westling T. Demystifying Statistical Inference When Using Machine Learning in Causal Research :kwab200. 110

[95] Balzer LB, Petersen ML. Invited Commentary : Machine Learning in Causal Inference-How Do I Love Thee ? Let Me Count the Ways ;190(8) :1483-7. 110

[96] Taylor SP, Kowalkowski MA, Admon AJ. Timing Is Everything. The Importance of Alignment of Time Anchors for Observational Causal Inference Research. Annals of the American Thoracic Society. 2021 May ;18(5) :769-72. 111

[97] Levin MA, Wanderer JP, Ehrenfeld JM. Data, Big Data, and Metadata in Anesthesiology. Anesthesia and Analgesia. 2015 Dec ;121(6) :1661-7. 111

[98] Schenck EJ, Hoffman KL, Cusick M, Kabariti J, Sholle ET, Campion TR. Critical carE Database for Advanced Research (CEDAR) : An automated method to support intensive care units with electronic health record data ;118 :103789. Available from : https://www.sciencedirect.com/science/article/pii/S1532046421001180. 111

[99] Biedermann P, Ong R, Davydov A, Orlova A, Solovyev P, Sun H, et al. Standardizing registry data to the OMOP Common Data Model : experience from three pulmonary hypertension databases ;21(1) :238. Available from : https://doi.org/10.1186/s12874-021-01434-3. 111

[100] McLennan S, Shaw D, Celi LA. The challenge of local consent requirements for global critical care databases ;45(2) :246-8. 111

[101] Thoral PJ, Peppink JM, Driessen RH, Sijbrands EJG, Kompanje EJO, Kaplan L, et al. Sharing ICU Patient Data Responsibly Under the Society of Critical Care Medicine/European Society of Intensive Care Medicine Joint Data Science Collaboration : The Amsterdam University Medical Centers Database (AmsterdamUMCdb) Example ;49(6) :e563-77. 111

[102] Benkeser D, Ju C, Lendle S, van der Laan M. Online Cross-Validation-Based Ensemble Learning. Statistics in medicine. 2018 Jan ;37(2) :249-60. Available from : https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5671383/. 111

Annexe A

Mesures de performance pour les études de simulation

TABLEAU A.1 – Résumé des principales mesures de performance des études de simulation, adapté de Morris, White et Crowther (2019)

Mesure de performances	Définition	Estimation	Erreur de Monte-Carlo
Biais absolu	$\mathrm{E}[\hat{\theta}] - \theta$	$\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \hat{\theta}_i - \theta$	$\sqrt{\frac{1}{n_{\text{sim}}(n_{\text{sim}}-1)} \sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \bar{\theta})^2}$
Erreur standard empirique (eSE)	$\sqrt{\mathrm{Var}(\hat{\theta})}$	$\sqrt{\frac{1}{n_{\text{sim}}-1} \sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \bar{\theta})^2}$	$\frac{\widehat{\text{eSE}}}{\sqrt{2(n_{\text{sim}}-1)}}$
Erreur standard asymptotique (aSE)[a]	$\sqrt{\mathrm{E}[\widehat{\mathrm{Var}}(\hat{\theta})]}$	$\sqrt{\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \widehat{\mathrm{Var}}(\hat{\theta}_i)]}$	$\sqrt{\frac{\widehat{\mathrm{Var}}[\widehat{\mathrm{Var}}(\hat{\theta})]}{4 n_{\text{sim}} \times \widehat{a\text{SE}}^2}}$
Biais d'estimation de la variance (%)[a]	$100\left(\frac{a\text{SE}}{e\text{SE}} - 1\right)$	$100\left(\frac{\widehat{a\text{SE}}}{\widehat{e\text{SE}}} - 1\right)$	$100\left(\frac{\widehat{a\text{SE}}}{\widehat{e\text{SE}}}\right) \sqrt{\frac{\widehat{\mathrm{Var}}[\widehat{\mathrm{Var}}(\hat{\theta})]}{4 n_{\text{sim}} \times \widehat{a\text{SE}}^4} + \frac{1}{2(n_{\text{sim}}-1}}$
Erreur quadratique moyenne (MSE)	$\mathrm{E}[(\hat{\theta} - \theta)^2]$	$\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} (\hat{\theta}_i - \theta)^2$	$\sqrt{\frac{\sum_{i=1}^{n_{\text{sim}}} [(\hat{\theta}_i - \theta)^2 - \widehat{\text{MSE}}]^2}{n_{\text{sim}}(n_{\text{sin}^{-1}})}}$
Couverture (%)	$\mathrm{P}(\hat{\theta}_{\text{inf}} \leq \theta \leq \hat{\theta}_{\text{sup}})$	$\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \mathbf{1}(\hat{\theta}_{\text{inf},i} \leq \theta \leq \hat{\theta}_{\text{sup},i})$	$\sqrt{\frac{\widehat{\text{Couv.}}(1-\widehat{\text{Couv.}})}{n_{\text{sim}}}}$
Rejet (%)[b]	$\mathrm{P}(p_{value} \leq \alpha)$	$\frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \mathbb{1}(p_{value,i} \leq \alpha)$	$\sqrt{\frac{\widehat{\text{Rejet}}(1-\widehat{\text{Rejet}})}{n_{\text{sim}}}}$

[a] $\widehat{\mathrm{Var}}[\widehat{\mathrm{Var}}(\hat{\theta})] \approx \frac{1}{n_{\text{sim}}-1} \sum_{i=1}^{n_{\text{sim}}} \left\{ \widehat{\mathrm{Var}}(\hat{\theta}_i) - \frac{1}{n_{\text{sim}}} \sum_{j=1}^{n_{\text{sim}}} \widehat{\mathrm{Var}}(\hat{\theta}_j) \right\}^2$

[b] Correspond respectivement à l'erreur de type I et à la puissance statistique sous les hypothèses nulle et alternative.

Annexe B

Présentation du cahier de recueil formalisé pour la cohorte AtlanRéa.

Liste des variables

NTLR0262 - RC12_0207

CRF 1

Inclusion

Page de garde 1

Données démographiques 2

Bilan biologique 3

Score IGS II 4

Score SOFA 5

Prise en charge d'une infection (1/2) 6

Prise en charge d'une infection (2/2) 7

Défaillance d'organes 8

Devenir 9

Entité nosologique 10

Traumatisés graves

Contexte 11

Bilan Lésionnel 12

Prise en charge avant l'entrée en réanimation 13

Prise en charge chirurgicale 14

Séjour en réanimation 15

Suivi sérologique 16

Valeurs Pression Intra-Cranienne 17

Ventilation mécanique 18

Examen clinique 19

Devenir 20

Cérébrolésés

Contexte 21

Prise en charge avant l'entrée en réanimation 22

Prise en charge chirurgicale 23

Séjour en réanimation 24

Suivi sérologique 25

Valeurs Pression Intra-Cranienne 26

Complications spécifiques à l'HSA 27

Examen clinique 28

Devenir 29

© Ennov Clinical

Suivi prospectif de cohorte

Centre : CENTRE (21)

LETTRE_NOM (21) LETTRE_PRENOM (21)

Patient (1ère lettre NOM /1ère lettre PRÉNOM)

DATE_NAISSANCE (21)

Date de naissance (MM/AAAA)

DATE_ENTREE_REANIMATION (21)

Date d'entrée en réanimation (JJ/MM/AAAA)

© Ennov Clinical

Centre Patient

ATLANREA

DONNÉES DÉMOGRAPHIQUES

Âge : AGE_PATIENT (21) **Sexe** : SEXE (21) *N_DOSEX*

POIDS_KG (21)
Poids réel : kg

TAILLE_CM (21)
Taille : cm

Insuffisance cardiaque (NYHA ≥ 2) : INSUFFISANCE_CARDIAQUE (21) *OUINON*

Insuffisance rénale chronique modérée (définie par Cockroft ou MDRD < 50 mL/min) :
INSUFFISANCE_RENALE (21) *OUINON*

Antécédents neurologiques :

Accident Vasculaire Cérébral AVC (21) *OUINON*
Accident Ischémique Transitoire AIT (21) *OUINON*
Démence et troubles cognitifs DEMENCE (21) *OUINON*
Traumatisme Crânien TRAUMATISME_CRANIEN (21) *OUINON*
Aucun AUCUN_ATCD (21) *OUINON*
Autre AUTRE_ATCD (21) *OUINON*

Pathologie respiratoire chronique (BPCO défini par un Tiffeneau < 70%, Asthme sous corticothérapie inhalée au long cours) : PATHOLOGIE_RESPIRATOIRE_CHRONIQU (21) *OUINON*

Diabète : DIABETE (21) *N_READIAB*

Alcoolisme chronique défini par > 30 g/jour pour un homme, > 20 g/jour pour une femme :
ALCOOLISME_CHRONIQUE (21) *OUINON*

IMMUNODEPRESSION (21)
Immunodépression (VIH, cancer évolutif, traitement immunosuppresseur) : *OUINON*

Cancer dans les 5 dernières années : CANCER (21) *OUINON*

Tabagisme actif : TABAGISME_ACTIF (21) *OUINON*

Si oui nombre de Paquets Années : NOMBRE_PAQUETS_ANNEES (21)

© Ennov Clinical

Centre Patient

AtlanRéa

BILAN BIOLOGIQUE À L'ENTRÉE

Prendre les valeurs les plus péjoratives des 24 premières heures.

En l'absence de résultat biologique, saisir ND ou NK dans le champ correspondant.

Hémoglobine (en g/dL) : HEMOGLOBINE (21)

Leucocytes (Valeur la plus élevée en G/L) : LEUCOCYTES (21)

Lymphocytes (Valeur la plus basse en G/L) : LYMPHOCYTES (21)

Neutrophiles (en G/L) : NEUTROPHILES (21)

Monocytes (en G/L) : MONOCYTES (21)

TP (en %) : TP (21)

Plaquettes (Valeur la plus basse en G/L) : PLAQUETTES (21)

Fibrinogène (en g/L) : FIBRINOGENE (21)

Lactates (en mmol/L) : LACTATES (21)

pH : PH (21)

Bicarbonatémie (en mmol/L) : BICARBONATEMIE (21)

PaO2 : PAO2_MMHG (21) en mmHg ou en kPa PAO2_KPA (21)

FiO2 (entre 0.21 et 1) : FIO2 (21)

Créatininémie (en µmol/L) : CREATININEMIE (21)

Protidémie (en g/L) : PROTIDEMIE (21)

CRP (en mg/L) : CRP (21)

Glycémie (en mmol/L) : GLYCEMIE (21)

Urémie (en mmol/L) : UREMIE (21)

Calcémie totale (Valeur la plus basse en mmol/L) : CALCEMIE_TOTALE (21)

Bilirubinémie Totale (en µmol/L) : BILIRUBINEMIE_TOTALE (21)

© Ennov Clinical

Centre Patient

AtlanRéa

Score IGS II

Prendre les valeurs les plus péjoratives dans les 24 heures suivant l'admission en réanimation

Mode d'admission : MODE_ADMISSION (21) *N_MODEA*

Maladies chroniques : MALADIES_CHRONIQUES (21) *N_REAMAL*

Score de Glasgow : SCORE_GLASGOW (21)

Age : AGE (21) *N_AGE*

Pression Art.Syst. : PRESSION_ARTERIELLE_SYSTOLIQUE (21) *N_REAPAS*

Fréquence cardiaque : FREQUENCE_CARDIAQUE (21) *N_FREQC*

Température : TEMPERATURE (21) *N_REATEMP*

PaO2/FiO2 : PAO2FIO2_IGSII (21) *N_PAFI*

Diurèse : (en L/24H) : DIURESE (21) *N_DIURESE*

Urée sanguine : UREE_SANGUINE (21) *N_REA_UREE*

Leucocytes : LEUCOCYTES1 (21) *N_REALEUCO*

Kaliémie (en mEq/L) : KALIEMIE (21) *N_KALIEM*

Natrémie (en mEq/L) : NATREMIE (21) *N_NATREM*

Bicarbonatémie (en mEq/L) : BICARBONATEMIE_IGSII (21) *N_HCO*

Bilirubine (si ictère) : BILIRUBINE (21) *N_BILIR*

Pour mettre à jour les scores, cliquer sur ENREGISTRER LA PAGE.

SCORE_IGS_II (21)

© Ennov Clinical

Centre Patient

AtlanRéa

Score SOFA (à l'admission dans le service)

- **PaO2/FiO2 :** SOFA_PAO2FIO2 (21) *N_REA_PAO*
- **Numération Plaquettaire** (en 10^3 /mm^3) : SOFA_NUMEROTATION_PLAQUETTAIRE (21) *N_REAPLAQ*
- **Bilirubinémie Totale :** SOFA_BILIRUBINE_TOTALE (21) *N_REABILI*
- **Hypotension :** SOFA_HYPOTENSION (21) *N_REAHYPO*
- **Score de Glasgow :** SOFA_SCORE_GLASGOW (21) *N_REAGLASGOW*
- **Créatininémie :** SOFA_CREATININEMIE (21) *N_REACREAT*

SOFA = SCORE_SOFA (21)

Pour mettre à jour les scores, cliquer sur ENREGISTRER LA PAGE.

© Ennov Clinical

Centre Patient

AtlanRéa

PRISE EN CHARGE

1. Prise en charge d'une infection acquise à l'hôpital : INFECTION (21) *OUINON*

(Précision : La 1ère infection au cours du séjour)

- **Poumons :** INFECTION_POUMONS (21) *OUINON*
  - Si oui : TYPE_INFECTION (21) *N_REAPOU*
  - si oui: survenue à J (J0 = entrée en réanimation, compter le jour du prélèvement bactériologique) POUMON_SURVENUE (21)
  - Symptomatologie :

    Température maximale le jour du prélèvement bactériologique : TEMP_MAX (21)

    Hyperleucocytose maximale le jour du prélèvement bactériologique : HYPERLEUCOCYTOSE_MAX (21)

    Nécessité d'augmenter la FiO2 lors de l'épisode infectieux (pour plus de 6H00) : *OUINON* AUGMENTER_FIO2 (21)

    Rapport PaO2/FiO2 le plus bas mesuré et rattaché à l'infection : RAPPORT_PAO2FIO2 (21)

    Nécessité de majorer la sédation pour assurer l'adaptation au respirateur : MAJORER_SEDATION (21) *OUINON*

    Nombre de lobes atteints à la radiographie : NB_LOBES (21)
  - si oui : codage bactériologique du prélèvement : POUMON_DECODE_1 (21) POUMON_CODAGE_BACTERIOLOGIQUE_1 (21)

    S'aider du lien ci-joint pour le POUMON_DECODE_2 (21) POUMON_CODAGE_BACTERIOLOGIQUE_2 (21)

    POUMON_DECODE_3 (21) POUMON_CODAGE_BACTERIOLOGIQUE_3 (21)
  - L'antibiothérapie probabiliste était-elle adaptée ? ANTIOBIO_ADAPTE (21) *OUINON*

- **Bactériémie :** INFECTION_BACTERIEMIE (21) *OUINON*
  - si oui: survenue à J (J0 = entrée en réanimation, compter le jour du prélèvement bactériologique) BACTERIEMIE_SURVENUE (21)
  - si oui : codage bactériologique du prélèvement : BACTERIEMIE_CODAGE_BACTERIOLOGIQ (21)

    S'aider du lien ci-joint pour le BACTERIEMIE_DECODE_1 (21) BACTERIEMIE_CODAGE_BACTERIOLOGI1 (21)

    BACTERIEMIE_DECODE_2 (21)

- **Site opératoire :** INFECTION_SITE_OPERATOIRE (21) *OUINON*
  - si oui: survenue à J (J0 = entrée en réanimation, compter le jour du prélèvement bactériologique) SITE_OPERATOIRE_SURVENUE (21)
  - si oui : codage bactériologique du prélèvement : SITE_OPERATOIRE_CODAGE_BACTERIOL (21)

    S'aider du lien ci-joint pour le SITE_OPERATOIRE_DECODE_1 (21) SITE_OPERATOIRE_CODAGE_BACTERIO1 (21)

    SITE_OPERATOIRE_DECODE_2 (21)

© Ennov Clinical

Centre Patient

AtlanRéa

- **Urine :** INFECTION_URINES (21) ***OUINON***
  - si oui: survenue à J (J0 = entrée en réanimation, compter le jour du prélèvement bactériologique) URINES_SURVENUE (21)
  - si oui : codage bactériologique du prélèvement : URINES_CODAGE_BACTERIOLOGIQUE_1 (21) URINES_CODAGE_BACTERIOLOGIQUE_2 (21)

S'aider du lien ci-joint pour le URINE_DECODE_1 (21) URINE_DECODE_2 (21)

- **Méninges :** INFECTION_MENINGES (21) ***OUINON***
  - si oui: survenue à J (J0 = entrée en réanimation, compter le jour du prélèvement bactériologique) MENINGES_SURVENUE (21)
  - si oui : codage bactériologique du prélèvement : MENINGES_CODAGE_BACTERIOLOGIQUE_ (21) MENINGES_CODAGE_BACTERIOLOGIQUE1 (21)

S'aider du lien ci-joint pour le MENINGES_DECODE_1 (21) MENINGES_DECODE_2 (21)

- **Infection due à un cathéter central :** INFECTION_CATHETER_CENTRAL (21) ***OUINON***
  - si oui: survenue à J (J0 = entrée en réanimation, compter le jour du prélèvement bactériologique) CATHETER_CENTRAL_SURVENUE (21)
  - si oui : codage bactériologique du prélèvement : CATHETER_CENTRAL_CODAGE_BACTERIO (21) CATHETER_CENTRAL_CODAGE_BACTERI1 (21)

S'aider du lien ci-joint pour le CATHETER_CENTRAL_DECODE_1 (21) CATHETER_CENTRAL_DECODE_2 (21)

- **Candidose invasive :** INFECTION_CANDIDOSE_INVASIVE (21) ***OUINON***
  - si oui: survenue à J (J0 = entrée en réanimation, compter le jour du prélèvement fongique) CANDIDOSE_INVASIVE_SURVENUE (21)
  - si oui : codage fongique du prélèvement : CANDIDOSE_INVASIVE_CODAGE_FONGIQ (21) ***N_CANDIDOSE***

Si Autre, merci de préciser : AUTRE_CODAGE_FONGIQUE (21)

© Ennov Clinical

Centre Patient

ATLANRÉA

2. Défaillance d'organe:

- **SDRA (PaO2/FiO2 < 200) :** SDRA (21) *OUINON*
  - si oui : survenu à J (J0 = entrée en réanimation) SDRA_SURVENUE (21)
  - si oui : P/F le plus bas RAPPORT_PAO2FIO21 (21) *OUINON*
- **Insuffisance rénale avec nécessité d'épuration extra-rénale :** INSUFFISANCE_RENALE1 (21)
  - si oui : survenue à J (J0 = entrée en réanimation) INSUFFISANCE_RENALE_SURVENUE (21)
  - si oui : valeur max de la Créatininémie (en µmol/L) : CREATININEMIE_MAX (21)
  - Si oui : type d'Epuration Extra Rénale : TYPE_EPURATION (21) *N_REAEPU*
  - si oui : durée d'EER : jours DUREE_EPURATION (21)

3. Durée des amines vasopressives : jours DUREE_AMINES (21)

4. Initiation d'une anticoagulation INITIATION_ANTICOAGULATION (21) *OUINON*

- si oui : à J (J0 = entrée en réanimation) ANTICOAGULATION_SURVENUE (21)
- si oui : Molécule : ANTICOAGULATION_MOLECULE (21) *N_REAMOL*
- si oui : 1ère administration à dose : ANTICOAGULATION_DOSE (21) *N_READOSE*
- Anti-agrégants en réanimation ? ANTIAGREGANTS_REANIMATION (21) *OUINON*
  Si Oui, à J ANTIAGREGANTS_SURVENUE (21)

- **Compression pneumatique intermittente :** COMPRESSION_PNEUMATIQUE (21) *OUINON*
- **Thrombose veineuse en réanimation :** THROMBOSE_VEINEUSE (21) *OUINON*
- **Embolie pulmonaire :** EMBOLIE_PULMONAIRE (21) *OUINON*
- **Hémorragie :** HEMORRAGIE (21) *OUINON*
  - si oui: survenue à J (J0 = entrée en réanimation) HEMORRAGIE_SURVENUE (21)
  - si oui : localisation : HEMORRAGIE_LOCALISATION (21) HEMORRAGIE_REPISE_CHIRURGICALE (21)
  - si oui : nécessité d'une reprise chirurgicale et/ou radio-interventionnelle : *OUINON*
  - si oui : nécessité de transfusion > ½ masse sanguine : HEMORRAGIE_TRANSFUSION (21) *OUINON*

© Ennov Clinical

Centre Patient

AtlanRéa

DEVENIR

- **Intubation trachéale :** INTUBATION_TRACHEALE (21) *OUINON*
  si oui: survenue à J (J0 = intubé à l'entrée en réanimation) INTUBATION_SURVENUE (21)
- **Durée cumulée ventilation mécanique invasive :** jours DUREE_VENTILATION_MECANIQUE (21)
- **Durée cumulée ventilation mécanique non-invasive :** jours DUREE_VENTILATION_MEC_NON_INV (21)
  Précision : VENTILATION_NON_INV_PREC (21) *N_REA_VENT*
- **Trachéotomie :** TRACHEOTOMIE (21) *OUINON*
  si oui: survenue à J TRACHEOTOME_SURVENUE (21)
- **Durée cumulée de cathétérisme central veineux :** jours DUREE_CATHETERISME_CENTRAL_VEINE (21)
- **Durée cumulée de cathétérisme artériel :** jours DUREE_CATHETERISME_ARTERIEL (21)
- **Durée cumulée de séjour en réanimation :** jours (sur l'ensemble du séjour hospitalier) DUREE_SEJOUR_REANIMATION (21)
- **Nombre de jours vivants sans antibiotiques sur les 28 premiers jours :** jours NOMBRE_JOURS_SANS_ANTIBIO (21)
- **Mortalité réanimation :** MORTALITE_REANIMATION (21) *OUINON*
- **Décision de limitation ou d'arrêt des thérapeutiques actives durant le séjour en réanimation :** DECISION_LIMITATION_THERAPEUTIQU (21) *OUINON*
  LATA_SURVENUE (21) Si oui : à J et type : LATA_TYPE (21) *N_REA_LATA*
  Si oui : motifs : - Terrain : LATA_TERRAIN (21) *OUINON*
  - Evaluation clinique : LATA_EVAL_CLINIQUE (21) *OUINON*
  - Evaluation TDM cérébrale : *OUINON* LATA_TDM_CEREBRALE (21)
  - Evaluation IRM cérébrale : *OUINON* LATA_IRM_CEREBRALE (21)

DEVENIR A J90

Devenir à J90 : DEVENIR_J90 (21) *N_MORTVIVANT*

GOS-Extended : GOS_E_J90 (21) *N_REAGOS*

© Ennov Clinical

Centre Patient

AtlanRéa

Entité nosologique

Si "Péritonites" ou "Pancréatites aigues", merci de cocher "Autre pathologie"

ENTITE_NOSOLOGIQUE (21) *N_REANOSO*

Le patient fait partie d'une étude sur l'Echographie : ECHOGRAPHIE (21) *OUINON*

Si oui, merci de préciser le type : ECHOGRAPHIE_TYPE (21) *N_REA_ECHO*

Le patient fait partie de l'étude UREA : ETUDE_UREA (21) *OUINON*

Le patient fait partie de l'étude TIPS : ETUDE_TIPS (21) *OUINON*

Le patient fait partie de l'étude Lymphocytes Neutrophiles Ratio (NLR) : ETUDE_NLR_RATIO (21) *OUINON*

Le patient fait partie de l'étude IBIS SEPSIS : ETUDE_IBIS_SEPSIS (21) *OUINON*

Le patient fait partie de l'étude BIONAGA : ETUDE_BIONAGA (21) *OUINON*

Le patient fait partie de l'étude Consultation Post-Réanimation : ETUDE_CONSULT_POST_REA (21) *OUINON*

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

ATLANREA

CONTEXTE

SMUR :

- Département : TG_SMUR_DEPARTEMENT (21) *N_READEP*
- Date : TG_SMUR_DATE (21)
- Heure d'appel ou de départ : TG_SMUR_HEURE_APPEL (21)
- Heure d'arrivée : TG_SMUR_HEURE_ARRIVEE (21)
- Heure d'arrivée au déchoquage : TG_SMUR_HEURE_ARRIVEE_DECHOQUAGE (21)

En cas de prise en charge initiale dans un autre CH, merci de saisir ci-dessus les éléments rattachés au transfert vers le CHU référent pour les traumatisés

Intervention :

Intervention primaire TG_INTERVENTION_PRIMAIRE (21) *OUINON*
Transfert d'un autre CH TG_TRANSFERT_AUTRE_CH (21) *OUINON*
Transfert du SAU TG_TRANSFERT_SAU (21) *OUINON*

Traumatisme : TG_TRAUMATISME (21) *N_REATRAUMA*

Mécanisme : TG_MECANISME (21) *N_REAMECA*

Si autre mécanisme, merci de préciser : TG_AUTRE_MECANISME (21)

Brûlure : TG_BRULURE (21) *OUINON*

Si oui : - Surface (%) : TG_SURFACE (21)

- Score USB (Unité de Brûlure Standard) : TG_SCORE_USB (21)

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

BILAN LÉSIONNEL

Score de Glasgow le plus péjoratif observé avant l'entrée en réanimation

Y = TG_SCORE_GLASGOW_Y (21) *N_GLASY*

V = TG_SCORE_GLASGOW_V (21) *N_GLASV*

M = TG_SCORE_GLASGOW_M (21) *N_GLASM*

Total = TG_SCORE_GLASGOW_TOTAL (21)

Mydriase ou anisocorie lors de la prise en charge initiale : TG_MYDRIASE (21) *OUINON*

Traumatisme crânien grave TG_TC_GRAVE (21) *OUINON*

Si oui : **Score de l'US Trauma Coma Data Bank** TG_SCORE_US (21) *N_USTCDB*

Lésion diffuse I : pas de signe d'atteinte au scanner
Lésion diffuse II : citernes de la base présentes, déviation de la ligne médiane ≤ à 5 mm et/ou absence de lésion hyperdense ou en mosaïque ≥ 25 ml
Lésion III : compression ou disparition des citernes de la base, déviation de la ligne médiane ≤ à 5 mm, absence de zone hyperdense ou en mosaïque ≥ 25 ml
Lésion diffuse IV : déviation de la ligne médiane > à 5 mm, pas de lésion hyperdense ou en mosaïque
Lésion chirurgicale : toutes les lésions chirurgicales (hématomes)
Pas de lésion chirurgicale : lésions hyperdenses ou en mosaïques > 25 ml non chirurgicales

Score AIS par région anatomique

(Sélectionner l'atteinte pour chaque zone à l'aide du menu déroulant de 1 à 6 puis cocher les 3 régions les + touchées grâce aux cases à cocher)

OUINON

TG_TETE_COU (21)	Tête et cou	TG_TETE_COU_SCORE_AIS (21) *N_AIS*
TG_FACE (21)	Face	TG_FACE_SCORE_AIS (21) *N_AIS*
TG_THORAX (21)	Thorax	TG_THORAX_SCORE_AIS (21) *N_AIS*
TG_ABDOMEN (21)	Abdomen et contenu pelvien	TG_ABDOMEN_SCORE_AIS (21) *N_AIS*
TG_MEMBRES_BASSIN (21)	Membres ou bassin	TG_MEMBRES_BASSIN_SCORE_AIS (21) *N_AIS*
TG_PEAU (21)	Peau et tissu sous cutané	TG_PEAU_SCORE_AIS (21) *N_AIS*

(AIS : 1= mineure, 2= modérée, 3= sérieux, 4= sévère, 5= critique, 6= au-delà de toute thérapeutique)
(Le rachis cervical est rattaché à tête et cou, le rachis thoracique est rattaché au thorax, le rachis lombaire est rattaché à abdomen/contenu pelvien, le sacrum est rattaché au bassin)

Score ISS = TG_SCORE_ISS (21)

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

ATLANREA

PRISE EN CHARGE AVANT L'ENTRÉE EN RÉANIMATION

Induction

Etomidate	TG_INDUCTION_ETOMIDATE (21)	*OUINON*	TG_INDUCTION_AUTRE (21)	*OUINON*
Propofol	TG_INDUCTION_PROPOFOL (21)	*OUINON*	TG_INDUCTION_NK (21)	*OUINON*
Kétamine	TG_INDUCTION_KETAMINE (21)	*OUINON*		
Pentothal	TG_INDUCTION_PENTOTHAL (21)	*OUINON*		

Intubation hors réanimation : TG_INTUBATION (21) *N_REA_INTUB*

Intubation en réanimation à J TG_INTUB_REA (21)

Hémodynamique :

Episode d'arrêt cardio-respiratoire : TG_ARRET_CARDIO_RESPIRATOIRE (21) *OUINON*

Episode avec PAS ≤ 90 mmHg : TG_PAS_90MMHG (21) *OUINON*

Dose maximale de Noradrénaline (en µg/kg/min) : TG_DOSE_MAX_NORADRENALINE (21)

Transfusion : TG_TRANSFUSION (21) *OUINON*

- Si oui : débutée en pré-hospitalier : TG_TRANSFUSION_PRE_HOSP (21) *OUINON*
- Si oui : nombre de CG : TG_NB_CG (21)
- Si oui : nombre de PFC : TG_NB_PFC (21)
- Si oui : nombre de CPA/MCP : TG_NB_CPAMCP (21)
- Si oui : dose de Fibrinogène (en g) : TG_DOSE_FIBRINOGENE (21)

Acide tranexamique : TG_ACIDE_TRANEXAMIQUE (21) *OUINON*

Novoseven : TG_NOVOSEVEN (21) *OUINON*

Neurologique :

Osmothérapie à la phase de déchoquage : TG_OSMOTHERAPIE (21) *N_APANA*

Hyperventilation contrôlée à la phase de déchoquage : TG_HYPERVENTILATION (21) *N_APANA*

Général :

CIVD biologique : TG_CIVD_BIOLOGIQUE (21) *OUINON*

Acidose sévère (pH < 7.30) : TG_ACIDOSE_SEVERE (21) *OUINON*

Hypothermie (Température < 36°) : TG_HYPOTHERMIE (21) *OUINON*

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

ATLANRÉA

PRISE EN CHARGE CHIRURGICALE

Délai entre arrivée au déchoquage et début du 1er geste (chirurgie ou pose de PIC/DVE ou embolisation) : TG_DELAI_DECHOQUAGE_1_GESTE (21) *N_REAPEC*

Neurochirurgie : TG_NEUROCHIRURGIE (21) *N_APANA*

TG_EVACUATION_HEMATOME (21) *OUINON*

- Si oui : évacuation d'un Hématome Extradural / d'un Hématome Sous-Dural :
- Si oui : craniectomie décompressive : TG_CRANIECTOMIE_DECOMPRESSIVE (21) *OUINON*
- Si oui : pose d'une Dérivation Ventriculaire Externe : TG_POSE_DVE (21) *OUINON*

  Si oui : initiation à J TG_JOUR_DVE (21) et durée : TG_DUREE_DVE (21) jours
- Si oui : évacuation d'un hématome-contusion / lobectomie : TG_LOBECTOMIE (21) *OUINON*

Chirurgie du rachis : TG_CHIRURGIE_RACHIS (21) *OUINON*

- Si oui : à J TG_JOUR_RACHIS (21) (J0 = jour d'entrée en réanimation)

Chirurgie viscérale : TG_CHIRURGIE_VISCERALE (21) *OUINON*

- Si oui : splénectomie : TG_SPLENECTOMIE (21) *OUINON*
- Si oui : perforation d'un organe creux intra-abdominal : TG_PERFORATION_ORGANE (21) *OUINON*
- Si oui : packing hépatique : TG_PACKING_HEPATIQUE (21) *OUINON*
- Si oui : autre : TG_AUTRE_CHIRURGIE (21) *OUINON*

  Préciser, champ libre TG_PRECISION_AUTRE_CHIRURGIE (21)

Saignement rétro-péritonéal : TG_SAIGNEMENT (21) *OUINON*

- Si oui : prise en charge : TG_PRISE_EN_CHARGE_SAIGNEMENT (21) *N_REASAIG*

Chirurgie thoracique : TG_CHIRURGIE_THORACIQUE (21) *OUINON*

Fracture d'un os long : TG_FRACTURE_OS_LONG (21) *OUINON*

- Si oui : réalisation à J TG_JOUR_FRACTURE (21) (J0 = jour d'entrée en réanimation)
- Si oui : fixation TG_FIXATION (21) *N_REAFIX*

Amputation d'un membre : TG_AMPUTATION_MEMBRE (21) *OUINON*

OUINON

TG_EMBOLISATION_AUCUNE (21) TG_EMBOLISATION_FOIE (21) TG_EMBOLISATION_RATE (21) TG_EMBOLISATION_BASSIN (21)

Embolisation : Pas d'embolisation Foie Rate Bassin Rein Autre

TG_EMBOLISATION_REIN (21)

Si Autre, merci de préciser : TG_PRECISION_AUTRE_EMBOLISATION (21)

TG_EMBOLISATION_AUTRE (21)

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

ATLANRÉA

SÉJOUR EN RÉANIMATION
UNIQUEMENT POUR LES PATIENTS AVEC TRAUMATISME CRÂNIEN

Neurologique : *(A saisir seulement si TC grave)*

Durée de monitorage de la Pression Intra-Crânienne : jours TG_DUREE_MONITORAGE (21)

Episode d'Hypertension Intra-Crânienne (PIC/PIV ≥ 25 mmHg) : TG_EPISIODE_HIC (21) *OUINON*

Osmothérapie : TG_TC_OSMOTHERAPIE (21) *OUINON*

- Si oui : initiation à J TG_OSMTG_OTHERAPIE_INITIATION (21) (J0 = jour du traumatisme)
- Si oui : durée : TG_OSMOTHERAPIE_DUREE (21) jours (compter un jour dès qu'un bolus a été administré)

Barbituriques : TG_BARBITURIQUES (21) *OUINON*

- Si oui : initiation à J (J0 = jour du traumatisme) TG_BARBITURIQUES_INITIATION (21)
- Si oui : durée : TG_BARBITURIQUES_DUREE (21) jours (compter un jour dès qu'un bolus a été administré)

Crâniectomie décompressive : TG_CRANIECTOMIE_DECOMPRESSIVE1 (21) *OUINON*

- Si oui : réalisée à J (J0 = jour du traumatisme) TG_CRANIECTOMIE_REALISATION (21)

Hypothermie thérapeutique : TG_HYPOTHERMIE_THERAPEUTIQUE (21) *OUINON*

- Si oui : initiation à J (J0 = jour du traumatisme) TG_HYPOTHERMIE_INITIATION (21)
- Si oui : durée : jours TG_HYPOTHERMIE_DUREE (21)

Hyperventilation contrôlée : TG_HYPERVENTILATION_CONTROLEE (21) *OUINON*

- Si oui : initiation à J (J0 = jour du traumatisme) TG_HYPERVENTILATION_INITIATION (21)
- Si oui : durée : jours TG_HYPERVENTILATION_DUREE (21)

Microdialyse cérébrale : TG_MICRODIALYSE_CEREBRALE (21) *OUINON*

Crises convulsives à l'EEG : TG_CRISES_CONVULSIVES (21) *OUINON*

Scanner cérébral de contrôle : TG_DELAI_SCANNER (21)

- Délai entre le 1er et le 2nd scanner : heures
- Evolution des lésions : TG_EVOLUTION_LESIONS (21) *N_REALES*

Sédation à l'admission : - Hypnotique : TG_HYPNOTIQUE (21) *N_REA_HYPNO*

Dose prescrite lors de la 1ère heure d'hospitalisation (dose maximale de perfusion continue, hors bolus) : TG_HYPNOTIQUE_DOSE_MAXI (21) *N_UNITE_SEDATION* TG_HYPNOTIQUE_UNITE (21)

- Morphinique : TG_MORPHINIQUE (21) *N_REA_MORPH*

Dose prescrite lors de la 1ère heure d'hospitalisation (dose maximale de perfusion continue, hors bolus) : TG_MORPHINIQUE_DOSE_MAXI (21) *N_UNITE_SEDATION* TG_MORPHINIQUE_UNITE (21)

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

Herpès Simplex Virus :

Sérologie : TG_HERPES_SEROLOGIE (21) *N_REASERO*

PCR sang **J1** : TG_HERPES_PCR_SANG_J1 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_HERPES_SANG_CHARGE_J1 (21)

PCR sang **J7** : TG_HERPES_PCR_SANG_J7 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_HERPES_SANG_CHARGE_J7 (21)

PCR sang **J15** : TG_HERPES_PCR_SANG_J15 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_HERPES_SANG_CHARGE_J15 (21)

PCR aspirations trachéales **J1** : TG_HERPES_PCR_ASPI_J1 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_HERPES_ASPI_CHARGE_J1 (21)

PCR aspirations trachéales **J7** : TG_HERPES_PCR_ASPI_J7 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_HERPES_ASPI_CHARGE_J7 (21)

PCR aspirations trachéales **J15** : TG_HERPES_PCR_ASPI_J15 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_HERPES_ASPI_CHARGE_J15 (21)

CytoMégaloVirus :

Sérologie : TG_CYTO_SEROLOGIE (21) *N_REASERO*

PCR sang **J1** : TG_CYTO_PCR_SANG_J1 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_CYTO_SANG_CHARGE_J1 (21)

PCR sang **J7** : TG_CYTO_PCR_SANG_J7 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_CYTO_SANG_CHARGE_J7 (21)

PCR sang **J15** : TG_CYTO_PCR_SANG_J15 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_CYTO_SANG_CHARGE_J15 (21)

PCR aspirations trachéales **J1** : TG_CYTO_PCR_ASPI_J1 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_CYTO_ASPI_CHARGE_J1 (21)

PCR aspirations trachéales **J7** : TG_CYTO_PCR_ASPI_J7 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_CYTO_ASPI_CHARGE_J7 (21)

PCR aspirations trachéales **J15** : TG_CYTO_PCR_ASPI_J15 (21) *N_REASERO*
Si positive : Charge (en copies/ml) : TG_CYTO_ASPI_CHARGE_J15 (21)

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

Pose d'un capteur de monitorage de la Pression Intra-Crânienne ou Intra-Ventriculaire:

TG_POSE_CAPTEUR (21) *OUINON*

- Si oui : valeur à la pose (en mmHg) : TG_VALEUR_POSE (21)

*** A compléter durant toute la période où le patient est sous ventilation contrôlée**

Jour	Heure	PIC ou PIV (en mmHg)	PPC (en mmHg)	Na la plus basse * (en mmol/L)	Vt * (en mL)	PEP * (en cmH2O)	PaO2 *	PaCO2 *	FiO2 *
J0	08h00	TG_J0_08H_PIC (21)	TG_J0_08H_PPC (21)	TG_J0_NATREMIE (21)	TG_J0_08H_VT (21)	TG_J0_08H_PEP (21)	TG_PAO2_UNITE (21)	TG_J0_08H_PACO2 (21)	TG_J0_08H_FIO2 (21
	20h00	TG_J0_20H_PIC (21)	TG_J0_20H_PPC (21)		TG_J0_20H_VT (21)	TG_J0_20H_PEP (21)		TG_J0_20H_PACO2 (21)	TG_J0_20H_FIO2 (21
J1	08h00	TG_J1_08H_PIC (21)	TG_J1_08H_PPC (21)	TG_J1_NATREMIE (21)	TG_J1_08H_VT (21)	TG_J1_08H_PEP (21)		TG_J1_08H_PACO2 (21)	TG_J1_08H_FIO2 (21
	20h00	TG_J1_20H_PIC (21)	TG_J1_20H_PPC (21)		TG_J1_20H_VT (21)	TG_J1_20H_PEP (21)		TG_J1_20H_PACO2 (21)	TG_J1_20H_FIO2 (21
J2	08h00	TG_J2_08H_PIC (21)	TG_J2_08H_PPC (21)	TG_J2_NATREMIE (21)	TG_J2_08H_VT (21)	TG_J2_08H_PEP (21)		TG_J2_08H_PACO2 (21)	TG_J2_08H_FIO2 (21
	20h00	TG_J2_20H_PIC (21)	TG_J2_20H_PPC (21)		TG_J2_20H_VT (21)	TG_J2_20H_PEP (21)		TG_J2_20H_PACO2 (21)	TG_J2_20H_FIO2 (21
J3	08h00	TG_J3_08H_PIC (21)	TG_J3_08H_PPC (21)	TG_J3_NATREMIE (21)	TG_J3_08H_VT (21)	TG_J3_08H_PEP (21)		TG_J3_08H_PACO2 (21)	TG_J3_08H_FIO2 (21
	20h00	TG_J3_20H_PIC (21)	TG_J3_20H_PPC (21)		TG_J3_20H_VT (21)	TG_J3_20H_PEP (21)		TG_J3_20H_PACO2 (21)	TG_J3_20H_FIO2 (21
J4	08h00	TG_J4_08H_PIC (21)	TG_J4_08H_PPC (21)	TG_J4_NATREMIE (21)	TG_J4_08H_VT (21)	TG_J4_08H_PEP (21)		TG_J4_08H_PACO2 (21)	TG_J4_08H_FIO2 (21
	20h00	TG_J4_20H_PIC (21)	TG_J4_20H_PPC (21)		TG_J4_20H_VT (21)	TG_J4_20H_PEP (21)		TG_J4_20H_PACO2 (21)	TG_J4_20H_FIO2 (21
J5	08h00	TG_J5_08H_PIC (21)	TG_J5_08H_PPC (21)	TG_J5_NATREMIE (21)	TG_J5_08H_VT (21)	TG_J5_08H_PEP (21)		TG_J5_08H_PACO2 (21)	TG_J5_08H_FIO2 (21
	20h00	TG_J5_20H_PIC (21)	TG_J5_20H_PPC (21)		TG_J5_20H_VT (21)	TG_J5_20H_PEP (21)		TG_J5_20H_PACO2 (21)	TG_J5_20H_FIO2 (21

© Ennov Clinical

Centre Patient

AtlanRéa

TRAUMATISÉS GRAVES

Le patient a-t-il eu au moins une tentative d'extubation ? TG_TENTATIVE_EXTUBATION (21) *OUINON*

Ventilation mécanique (pour tous les patients) TG_1ERE_EXTUBATION (21)

Jour de réalisation de la première tentative d'extubation : J (J0 = jour du traumatisme)

Avant extubation : ventilation spontanée sur pièce en T : TG_VENTILATION_PIECE_T (21) *OUINON*

Si Oui, durée : minutes TG_VENTILATION_SPONTANEE_DUREE (21)

Ré-intubation (hors chirurgie programmée) : TG_REINTUBATION (21) *OUINON*

Si Oui, jour de ré-intubation : J TG_JOUR_REINTUBATION (21)

Si Oui : TG_TYPE_REINTUBATION (21) *N_REA_TYPE_INT*

TG_DELAI_DEVENTILATION_EXTUBATIO (21)

Délai en jour entre la déventilation (définit par VS-AI avec AI<10 et FiO2≤40%) et la 1ère tentative d'extubation (ou la réalisation de la trachéotomie si jamais extubé) :

TG_ECHEC_EXTUBATION (21) *OUINON*

Echec d'extubation défini par une ré-intubation dans les 48 heures post-extubation :

Si oui : Neurologique TG_ECHEC_NEUROLOGIQUE (21) *OUINON*

Dyspnée laryngée TG_ECHEC_DYSPNEE_LARYNGEE (21) *OUINON*

Encombrement bronchique TG_ECHEC_ENCOMBREMENT (21) *OUINON*

Autre : TG_ECHEC_AUTRE (21) *OUINON* TG_AUTRE_ECHEC (21)

Auto-Extubation : TG_AUTO_EXTUBATION (21) *OUINON*

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

Le patient a-t-il eu au moins une tentative d'extubation ? TG_TENTATIVE_EXTUBATION1 (21) *OUINON*

Examen clinique le jour de la 1ère tentative d'extubation / ou de la trachéotomie	
Durée cumulée de VS sur pièce en T dans les 24 heures qui précédent l'extubation (heures)	TG_DUREE_VS_NA (21) heures
Aspiration endotrachéale ≤ 2 par heure : sécrétions endotrachéales nécessitant moins de 2 aspirations par l'IDE toutes les heures	TG_AET (21) *OUINON*
Test de fuite effectué Si oui, existence d'une fuite	TG_TEST_FUITE (21) *OUINON* TG_EXISTENCE_FUITE (21) *OUINON*
Score de Glasgow Aide à la cotation: Y: coter 4 si le patient a les yeux ouverts sans contact V: coter 1 en cas d'absence de réponses, aucune tentative de communication V: coter 3 si le patient est calme et répond aux ordres simples V: coter 4 en cas de tentative de réponse avec les lèvres, oui/non de la tête aux questions simples	Y = TG_TC_SCORE_GLASGOW_Y (21) *N_GLASY* V = TG_TC_SCORE_GLASGOW_V (21) *N_GLASV* M= TG_TC_SCORE_GLASGOW_M (21) *N_GLASM*
Réponse aux ordres simples : serrer/ouvrir la main à la demande, bouger les jambes à la demande	TG_REPONSE_ORDRES (21) *OUINON*
Suivi du regard : suit du regard le staff dans la chambre, suivi d'un objet devant les yeux	TG_SUIVI_REGARD (21) *OUINON*
Toux	TG_TOUX_SPONTANEE (21) *OUINON*
Déglutition	TG_DEGLUTITION (21) *N_READEGLU*
Stase salivaire Nécessité d'aspirations buccales ou pharyngo-laryngées par les IDE plus de 2 fois par heure ou présence de salive extériorisée sur le visage ou le cou nécessitant des soins par les AS/IDE plus de 2 fois par heure	TG_STASE_SALIVAIRE (21) *OUINON*
Vomissements extériorisés au cours des dernières 24 heures	TG_VOMISSEMENTS (21) *OUINON*
Constipation (≥ 5jours qui précèdent)	TG_CONSTIPATION (21) *OUINON*
Hypotonie : ne décolle pas de membre (supérieur ou inférieur) du plan du lit.	TG_HYPOTONIE (21) *OUINON*
Confusion : présence de mouvements coordonnés au cours de l'éveil mais non accessible aux ordres simples	TG_CONFUSION (21) *OUINON*
Agitation : mouvements auto-agressifs (sortie du lit, arrachage KT, dispositifs divers) avec nécessité d'une contention physique	TG_AGITATION (21) *OUINON*
Vient chercher la sonde d'intubation avec la main	TG_SONDE_INTUBATION (21) *OUINON*
Traitement en cours: Neuroleptiques Benzodiazépines Morphine Corticoïdes (oedème laryngé)	TG_NEUROLEPTIQUES (21) *OUINON* TG_BENZODIAZEPINE (21) *OUINON* TG_MORPHINE (21) *OUINON* TG_CORTICOTHERAPIE (21) *OUINON*

© Ennov Clinical

Centre Patient

TRAUMATISÉS GRAVES

Le patient a-t-il été extubé dans les 48 premières heures où les 3 critères suivants étaient réunis : Déventilation acquise, Score de Glasgow >= 10 et Toux : TG_EXTUBATION (21) *OUINON*

Si Non, pourquoi : Etat neurologique jugé insuffisant TG_ETAT_NEURO (21) *OUINON*

TG_ATTENTE_INTERVENTION (21) *OUINON* Dans l'attente d'une intervention / imagerie programmée

TG_TROUBLES_DEGLUTITION (21) *OUINON* Troubles de la déglution attendus

TG_AUTRE_RAISON (21) *OUINON* Autre : TG_PRECISION_NON_EXTUBATION (21)

DEVENIR À LA SORTIE DE RÉANIMATION

Décès : TG_DECES (21) *OUINON*

Si oui, cause du décès : TG_CAUSE_DECES (21) *N_REA_DECES*

Si Mort encéphalique, Prélèvement multi-organe : TG_PRELEVEMENT_ORGANES (21) *OUINON*

Score de Glasgow :

Aide à la cotation:
Y: coter 4 si le patient a les yeux ouverts sans contact
V: coter 1 en cas d'absence de réponses, aucune tentative de communication
V: coter 3 si le patient est calme et répond aux ordres simples
V: coter 4 en cas de tentative de réponse avec les lèvres, oui/non de la tête aux questions simples

Y = TG_D_SCORE_GLASGOW_Y (21) *N_GLASY*

V = TG_D_SCORE_GLASGOW_V (21) *N_GLASV*

M = TG_D_SCORE_GLASGOW_M (21) *N_GLASM*

Total = TG_D_SCORE_GLASGOW_TOTAL (21)

GOS : TG_GOS (21) *N_ATLANGOS*

STATUT A 6 MOIS

Statut à 6 mois : TG_STATUT_6MOIS (21) *N_MORTVIVANT*

© Ennov Clinical

© Ennov Clinical

Imprimé le 24/04/2017 16:04:45

Centre Patient

CÉRÉBROLÉSÉS

CONTEXTE

Hémorragie Sous-Arachnoïdienne anévrysmale : CB_HEMORRAGIE (21) *OUINON*

- Si oui : classification : CB_CLASSIFICATION_WFNS (21) *N_WFNS*
  CB_CLASSIFICATION_FISHER (21) *N_FISCHER*
- Si oui : localisation de l'anévrysme : CB_LOCALISATION_ANEVRYSME (21) *N_REALOC*

  Si Autre, merci de préciser : CB_AUTRE_LOCALISATION (21)

Rupture d'une Malformation Artério-Veineuse : CB_RUPTURE_MALFORMATION (21) *OUINON*

Hématome Intra-Parenchymateux Spontané : CB_HEMATOME (21) *OUINON*

Accident Vasculaire Cérébral ischémique : CB_AVC (21) *OUINON*

Autre : CB_AUTRE_CONTEXTE (21) *OUINON*

BILAN LÉSIONNEL

Score de Glasgow le plus péjoratif observé avant l'entrée en réanimation

Y = CB_SCORE_GLASGOW_Y (21) *N_GLASY*

V = CB_SCORE_GLASGOW_V (21) *N_GLASV*

M = CB_SCORE_GLASGOW_M (21) *N_GLASM*

Total = CB_SCORE_GLASGOW_TOTAL (21)

Mydriase lors de la prise en charge initiale : CB_MYDRIASE (21) *OUINON*

Centre Patient

CÉRÉBROLÉSÉS

ATLANRÉA

PRISE EN CHARGE AVANT L'ENTRÉE EN RÉANIMATION

Induction

Etomidate	Autre	CB_INDUCTION_ETOMIDATE (21)	OUINON	CB_INDUCTION_AUTRE (21)	OUINON
Propofol	Non Connu	CB_INDUCTION_PROPOFOL (21)	OUINON	CB_INDUCTION_NK (21)	OUINON
Kétamine		CB_INDUCTION_KETAMINE (21)	OUINON		
Pentothal		CB_INDUCTION_PENTOTHAL (21)	OUINON		

Intubation hors réanimation : CB_INTUBATION (21) *N_REA_INTUB*

Intubation en réanimation à J CB_INTUB_REA (21)

Hémodynamique :

Episode d'arrêt cardio-respiratoire : CB_ARRET_CARDIO_RESP (21) *OUINON*

Episode avec PAS ≤ 90 mmHg : CB_PAS_90MMHG (21) *OUINON*

Dose maximale de Noradrénaline (en µg/kg/min) : CB_DOSE_MAX_NORA (21)

Transfusion : CB_TRANSFUSION (21) *OUINON*

- Si oui : débutée en pré-hospitalier : CB_TRANSFUSION_PRE_HOSP (21) *OUINON*
- Si oui : nombre de CG : CB_NB_CG (21)
- Si oui : nombre de PFC : CB_NB_PFC (21)
- Si oui : nombre de CPA/MCP : CB_NB_CPAMCP (21)
- Si oui : dose de Fibrinogène (en g) : CB_DOSE_FIBRINOGENE (21)

Acide tranexamique : CB_ACIDE_TRANEXAMIQUE (21) *OUINON*

Novoseven : CB_NOVOSEVEN (21) *OUINON*

Neurologique :

Osmothérapie à la phase de déchoquage : CB_OSMOTHERAPIE (21) *OUINON*

Hyperventilation contrôlée à la phase de déchoquage : CB_HYPERVENTILATION (21) *OUINON*

© Ennov Clinical

Centre Patient

CÉRÉBROLÉSÉS

PRISE EN CHARGE CHIRURGICALE

Neurochirurgie : CB_NEUROCHIRURGIE (21) *OUINON* CB_EVACUATION_HEMATOME (21) *OUINON*

- Si oui : évacuation d'un Hématome Extradural / d'un Hématome Sous-Dural :
- Si oui : craniectomie décompressive : CB_CRANIECTOMIE_DECOMPRESSIVE (21) *OUINON*
- Si oui : pose d'une Dérivation Ventriculaire Externe : CB_POSE_DVE (21) *OUINON*

Si oui : initiation à J et durée : jours
CB_DVE_SURVENUE (21) CB_DVE_DUREE (21)

- Si oui : évacuation d'un hématome-contusion / lobectomie : CB_LOBECTOMIE (21) *OUINON*

Radiologie interventionnelle : CB_RADIOLOGIE_INTERVENTIONNELLE (21) *OUINON*

- Si oui : survenue à J (J0 = jour d'entrée en réanimation)
CB_RADIOLOGIE_SURVENUE (21)

Clip chirurgical de l'anévrysme : CB_CLIP_CHIRURGICAL (21) *OUINON*

- Si oui : survenu à J (J0 = jour d'entrée en réanimation)
CB_CLIP_CHIRURGICAL_SURVENUE (21)

© Ennov Clinical

Centre Patient

CÉRÉBROLÉSÉS

ATLANREA

SÉJOUR EN RÉANIMATION

Neurologique :

CB_DUREE_MONITORAGE (21)

Durée de monitorage de la Pression Intra-Crânienne : jours

Episode d'Hypertension Intra-Crânienne (PIC/PIV ≥ 25 mmHg) : CB_EPISIODE_HIC (21) *OUINON*

Osmothérapie : CB__NEURO_OSMOTHERAPIE (21) *OUINON*

- Si oui : initiation à J (J0 = jour du traumatisme) CB_OSMOTHERAPIE_INITIATION (21)
- Si oui : durée : jours (compter un jour dès qu'un bolus a été administré) CB_OSMOTHERAPIE_DUREE (21)

Barbituriques : CB_BARBITURIQUES (21) *OUINON*

- Si oui : initiation à J (J0 = jour du traumatisme) CB_BARBITURIQUES_INITIATION (21)
- Si oui : durée : jours (compter un jour dès qu'un bolus a été administré) CB_BARBITURIQUES_DUREE (21)

Crâniectomie décompressive : CB_CRANIECTOMIE_DECOMPRESSIVE1 (21) *OUINON*

- Si oui : réalisée à J (J0 = jour du traumatisme) CB_CRANIECTOMIE_REALISATION (21)

Hypothermie thérapeutique : CB_HYPOTHERMIE_THERAPEUTIQUE (21) *OUINON*

- Si oui : initiation à J (J0 = jour du traumatisme) CB_HYPOTHERMIE_INITIATION (21)
- Si oui : durée : jours CB_HYPOTHERMIE_DUREE (21)

Hyperventilation contrôlée : CB_HYPERVENTILATION_CONTROLEE (21) *OUINON*

- Si oui : initiation à J (J0 = jour du traumatisme) CB_HYPERVENTILATION_INITIATION (21)
- Si oui : durée : jours CB_HYPERVENTILATION_DUREE (21)

Microdialyse cérébrale : CB_MICRODIALYSE_CEREBRALE (21) *OUINON*

Crises convulsives à l'EEG : CB_CRISES_CONVULSIVES (21) *OUINON*

Scanner cérébral de contrôle :

- Délai entre le 1er et le 2nd scanner : heures CB_DELAI_SCANNER (21)
- Evolution des lésions : CB_EVOLUTION_LESIONS (21) *N_REALES*

Nouveau saignement : CB_NOUVEAUX_SAIGNEMENTS (21) *OUINON*

Hydrocéphalie : CB_HYDROCEPHALIE (21) *OUINON*

Sédation à l'admission : - Hypnotique : CB_HYPNOTIQUE (21) *N_REA_HYPNO* CB_HYPNOTIQUE_UNITE (21)
Dose prescrite lors de la 1ère heure d'hospitalisation (dose maximale de perfusion continue, hors bolus) : CB_HYPNOTIQUE_DOSE_MAXI (21) *N_UNITE_SEDATION*

CB_MORPHINIQUE_DOSE_MAXI (21) Morphinique : CB_MORPHINIQUE (21) *N_REA_MORPH* CB_MORPHINIQUE_UNITE (21)
Dose prescrite lors de la 1ère heure d'hospitalisation (dose maximale de perfusion continue, hors bolus) : *N_UNITE_SEDATION*

© Ennov Clinical

Centre Patient

CÉRÉBROLÉSÉS

Herpès Simplex Virus :

Sérologie : CB_HERPES_SEROLOGIE (21) *N_REASERO*

PCR sang **J1** : CB_HERPES_PCR_SANG_J1 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_HERPES_SANG_CHARGE_J1 (21)

PCR sang **J7** : CB_HERPES_PCR_SANG_J7 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_HERPES_SANG_CHARGE_J7 (21)

PCR sang **J15** : CB_HERPES_PCR_SANG_J15 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_HERPES_SANG_CHARGE_J15 (21)

PCR aspirations trachéales **J1** : CB_HERPES_PCR_ASPI_J1 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_HERPES_ASPI_CHARGE_J1 (21)

PCR aspirations trachéales **J7** : CB_HERPES_PCR_ASPI_J7 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_HERPES_ASPI_CHARGE_J7 (21)

PCR aspirations trachéales **J15** : CB_HERPES_PCR_ASPI_J15 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_HERPES_ASPI_CHARGE_J15 (21)

CytoMégaloVirus :

Sérologie : CB_CYTO_SEROLOGIE (21) *N_REASERO*

PCR sang **J1** : CB_CYTO_PCR_SANG_J1 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_CYTO_SANG_CHARGE_J1 (21)

PCR sang **J7** : CB_CYTO_PCR_SANG_J7 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_CYTO_SANG_CHARGE_J7 (21)

PCR sang **J15** : CB_CYTO_PCR_SANG_J15 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_CYTO_SANG_CHARGE_J15 (21)

PCR aspirations trachéales **J1** : CB_CYTO_PCR_ASPI_J1 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_CYTO_ASPI_CHARGE_J1 (21)

PCR aspirations trachéales **J7** : CB_CYTO_PCR_ASPI_J7 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_CYTO_ASPI_CHARGE_J7 (21)

PCR aspirations trachéales **J15** : CB_CYTO_PCR_ASPI_J15 (21) *N_REASERO*

Si positive : Charge (en copies/ml) : CB_CYTO_ASPI_CHARGE_J15 (21)

© Ennov Clinical

CÉRÉBROLÉSÉS

ATLANRÉA

Pose d'un capteur de monitorage de la Pression Intra-Crânienne ou Intra-Ventriculaire: CB_POSE_CAPTEUR (21) *OUINON*

Si oui : valeur à la pose (en mmHg) : CB_VALEUR_POSE (21)

*** A compléter durant toute la période où le patient est sous ventilation contrôlée**

Jour	Heure	PIC ou PIV (en mmHg)	PPC (en mmHg)	Na la plus basse * (en mmol/L)	Vt * (en mL)	PEP * (en cmH2O)	PaO2 *	PaCO2 *	FiO2 *
J0	08h00	CB_J0_08H_PIC (21)	CB_J0_08H_PPC (21)	CB_J0_NATREMIE (21)	CB_J0_08H_VT (21)	CB_J0_08H_PEEP (21)	CB_PAO2_UNITE (21)	N_REA_UNIT	N_REA_
	20h00	CB_J0_20H_PIC (21)	CB_J0_20H_PPC (21)		CB_J0_20H_VT (21)	CB_J0_20H_PEEP (21)			
J1	08h00	CB_J1_08H_PIC (21)	CB_J1_08H_PPC (21)	CB_J1_NATREMIE (21)	CB_J1_08H_VT (21)	CB_J1_08H_PEEP (21)			
	20h00	CB_J1_20H_PIC (21)	CB_J1_20H_PPC (21)		CB_J1_20H_VT (21)	CB_J1_20H_PEEP (21)			
J2	08h00	CB_J2_08H_PIC (21)	CB_J2_08H_PPC (21)	CB_J2_NATREMIE (21)	CB_J2_08H_VT (21)	CB_J2_08H_PEEP (21)			
	20h00	CB_J2_20H_PIC (21)	CB_J2_20H_PPC (21)		CB_J2_20H_VT (21)	CB_J2_20H_PEEP (21)			
J3	08h00	CB_J3_08H_PIC (21)	CB_J3_08H_PPC (21)	CB_J3_NATREMIE (21)	CB_J3_08H_VT (21)	CB_J3_08H_PEEP (21)			
	20h00	CB_J3_20H_PIC (21)	CB_J3_20H_PPC (21)		CB_J3_20H_VT (21)	CB_J3_20H_PEEP (21)			
J4	08h00	CB_J4_08H_PIC (21)	CB_J4_08H_PPC1 (21)	CB_J4_NATREMIE (21)	CB_J4_08H_VT (21)	CB_J4_08H_PEEP (21)			
	20h00	CB_J4_20H_PIC (21)	CB_J4_20H_PPC (21)		CB_J4_20H_VT (21)	CB_J4_20H_PEEP (21)			
J5	08h00	CB_J5_08H_PIC (21)	CB_J5_08H_PPC (21)	CB_J5_NATREMIE (21)	CB_J5_08H_VT (21)	CB_J5_08H_PEEP (21)			
	20h00	CB_J5_20H_PIC (21)	CB_J5_20H_PPC (21)		CB_J5_20H_VT (21)	CB_J5_20H_PEEP (21)			

Valeur de natrémie la plus haute au cours des 5 premiers jours en réanimation (en nmol/L) : CB_NATREMIE_HAUT

Le patient a-t-il eu au moins une tentative d'extubation ? CB_TENTATIVE_EXTUB (21) *OUINON*

Ventilation mécanique : CB_1ERE_TENTATIVE_EXTUB (21)

Jour de réalisation de la première tentative d'extubation : J (J0 = jour du traumatisme)

Avant extubation : ventilation spontanée sur pièce en T : CB_VENTILATION_PIECE_T (21) *OUINON*

Si Oui, durée : minutes CB_VENTILATION_SPONTANEE_DUREE (21)

Ré-intubation (hors chirurgie programmée) : CB_REINTUBATION (21) *OUINON*

Si Oui, jour de ré-intubation : J Si Oui : CB_TYPE_REINTUBATION (21) *N_REA_TYPE_INT*

CB_JOUR_REINTUBATION (21)

CB_DELAI_DEVENTILATION_EXTUBATIO (21)

Délai en jour entre la déventilation (définit par VS-AI avec AI<10 et FiO2≤40%) et la 1ère tentative d'extubation (ou la réalisation de la trachéotomie si jamais extubé) :

CB_ECHEC_EXTUBATION (21) *OUINON*

Echec d'extubation défini par une ré-intubation dans les 48 heures post-extubation :

CB_ECHEC_NEURO (21) *OUINON*

Si oui : Neurologique

CB_ECHEC_ENCOMBREMENT (21) *OUINON*

Encombrement bronchique

CB_ECHEC_DYSPNEE_LARYNGEE (21)

Dyspnée laryngée

CB_ECHEC_AUTRE (21) *OUINON*

Autre :

CB_AUTRE_ECHEC (21)

OUINON

CB_AUTO_EXTUBATION (21) *OUINON*

CB_OEDEME_PULMONAIRE (21) *OUINON*

Auto-Extubation :

Œdème pulmonaire neurogénique :

© Ennov Clinical

© Ennov Clinical

Imprimé le 24/04/2017 16:04:45

Centre Patient

CÉRÉBROLÉSÉS

AtlanRéa

Complications spécifiques à l'HSA :

Vasospasme radiologique : CB_VASOSPASME (21) *OUINON*

- Si oui, survenu à J (J0 = jour d'entrée en réanimation) CB_VASOSPASME_SURVENUE (21)

Ischémie cérébrale retardée : CB_ISCHEMIE_CEREBRALE (21) *OUINON*

Cardiopathie de stress (FEVG < 50 % en l'absence de cardiopathie sous-jacente connue) : CB_CARDIOPATHIE_STRESS (21) *OUINON*

Angioplastie : CB_ANGIOPLASTIE (21) *OUINON*

Si Oui : Type : CB_TYPE_ANGIOPLASTIE (21) *N_REA_ANGIO*

Si Chimique : CB_CHIMIQUE_PRECISIONS (21) *N_ANGIOCHIM*

Pose de stent : CB_POSE_STENT (21) *OUINON*

Centre Patient

CÉRÉBROLÉSÉS

Le patient a-t-il eu au moins une tentative d'extubation ? CB_TENTATIVE_EXTUBATION (21) *OUINON*

Examen clinique le jour de la 1ère tentative d'extubation / ou de la trachéotomie	
Durée cumulée de VS sur pièce en T dans les 24 heures qui précédent l'extubation (heures)	CB_DUREE_VS_NA (21) heures
Aspiration endotrachéale ≤ 2 par heure : sécrétions endotrachéales nécessitant moins de 2 aspirations par l'IDE toutes les heures	CB_AET (21) *OUINON*
Test de fuite effectué Si oui, existence d'une fuite	CB_TEST_FUITE (21) *OUINON* CB_EXISTENCE_FUITE (21) *OUINON*
Score de Glasgow Aide à la cotation: Y: coter 4 si le patient a les yeux ouverts sans contact V: coter 1 en cas d'absence de réponses, aucune tentative de communication M: coter 4 en cas de tentative de réponse avec les lèvres, oui/non de la tête aux questions simples	Y = CB_EC_SCORE_GLASGOW_Y (21) *N_GLASY* V = CB_EC_SCORE_GLASGOW_V (21) *N_GLASV* M= CB_EC_SCORE_GLASGOW_M (21) *N_GLASM*
Réponse aux ordres simples : serrer/ouvrir la main à la demande, bouger les jambes à la demande	CB_REPONSE_ORDRES (21) *OUINON*
Suivi du regard : suit du regard le staff dans la chambre, suivi d'un objet devant les yeux	CB_SUIVI_REGARD (21) *OUINON*
Toux	CB_TOUX_SPONTANEE (21) *OUINON*
Déglutition	CB_DEGLUTITION (21) *N_READEGLU*
Stase salivaire Nécessité d'aspirations buccales ou pharyngo-laryngées par les IDE plus de 2 fois par heure ou présence de salive extériorisée sur le visage ou le cou nécessitant des soins par les AS/IDE plus de 2 fois par heure	CB_STASE_SALIVAIRE (21) *OUINON*
Vomissements extériorisés au cours des dernières 24 heures	CB_VOMISSEMENTS (21) *OUINON*
Constipation (≥ 5jours qui précèdent)	CB_CONSTIPATION (21) *OUINON*
Hypotonie : ne décolle pas de membre (supérieur ou inférieur) du plan du lit.	CB_HYPOTONIE (21) *OUINON*
Confusion : présence de mouvements coordonnés au cours de l'éveil mais non accessible aux ordres simples	CB_CONFUSION (21) *OUINON*
Agitation : mouvements auto-agressifs (sortie du lit, arrachage KT, dispositifs divers) avec nécessité d'une contention physique	CB_AGITATION (21) *OUINON*
Vient chercher la sonde d'intubation avec la main	CB_SONDE_INTUBATION (21) *OUINON*
Traitement en cours: Neuroleptiques Benzodiazépines Morphine Corticoïdes (oedème laryngé)	CB_NEUROLEPTIQUES (21) *OUINON* CB_BENZODIAZEPINE (21) *OUINON* CB_MORPHINE (21) *OUINON* CB_CORTICOTHERAPIE (21) *OUINON*

© Ennov Clinical

Centre Patient

CÉRÉBROLÉSÉS

Le patient a-t-il été extubé dans les 48 premières heures où les 3 critères suivants étaient réunis : Déventilation acquise, Score de Glasgow >= 10 et Toux : CB_EXTUBATION (21) *OUINON*

Si Non, pourquoi : Etat neurologique jugé insuffisant CB_ETAT_NEURO (21) *OUINON*

CB_ATTENTE_INTERVENTION (21) *OUINON* Dans l'attente d'une intervention / imagerie programmée

CB_TROUBLES_DEGLUTITION (21) *OUINON* Troubles de la déglution attendus

CB_AUTRE_RAISON (21) *OUINON* Autre : CB_PRECISION_NON_EXTUBATION (21)

DEVENIR À LA SORTIE DE RÉANIMATION

Décès : CB_DECES (21) *OUINON*

Si oui, cause du décès : CB_CAUSE_DECES (21) *N_REA_DECES*

Si Mort encéphalique, Prélèvement multi-organe : CB_PRELEVEMENT_ORGANES (21) *OUINON*

Score de Glasgow :

Aide à la cotation:
Y: coter 4 si le patient a les yeux ouverts sans contact
V: coter 1 en cas d'absence de réponses, aucune tentative de communication
V: coter 3 si le patient est calme et répond aux ordres simples
V: coter 4 en cas de tentative de réponse avec les lèvres, oui/non de la tête aux questions simples

Y = CB_D_SCORE_GLASGOW_Y (21) *N_GLASY*

V = CB_D_SCORE_GLASGOW_V (21) *N_GLASV*

M = CB_D_SCORE_GLASGOW_M (21) *N_GLASM*

Total = CB_D_SCORE_GLASGOW_TOTAL (21)

GOS : CB_GOS (21) *N_ATLANGOS*

STATUT A 6 MOIS

Statut à 6 mois : CB_STATUT_6MOIS (21) *N_MORTVIVANT*

© Ennov Clinical

Annexe C

Étude de Danelian *et al.* Identification automatisée de potentielles violations de l'hypothèse de positivité

Identifying positivity violations using decision trees: introducing the P-CART algorithm

Gabriel Danelian,[1,2] Yohann Foucher,[3,4] Maxime Léger,[3,5] Florent Le Borgne,[2,3] and Arthur Chatton[2,3*]

[1] Université de Lille, Lille, France

[2] IDBC/A2COM, Pacé, France.

[3] UMR INSERM 1246 - SPHERE, Université de Nantes, Université de Tours, Nantes, France.

[4] Centre Hospitalier Universitaire de Nantes, Nantes, France.

[5] Département d'anesthésie-réanimation, Centre Hospitalier Universitaire d'Angers, Angers, France.

* Corresponding author: Arthur Chatton, UMR INSERM 1246 - SPHERE, Institut de Recherche en Santé 2, 22 Boulevard Benoni-Goullin, 44200 NANTES, France. Mail: arthur.chatton@univ-nantes.fr

Word count: 2592

Abstract

Background: The positivity assumption is crucial in observational studies where design does not guarantee ambivalence, but is often overlooked in practice. A structural (non-random) infraction of positivity is due to a subgroup of individuals being unable to receive one of the levels of exposure. To correctly estimate the causal effect, we need to identify and exclude this subgroup of individuals. We suggest a CART-based algorithm to help achieve this.

Development: Based on a succession of decision trees, the algorithm searches for combinations of covariate values that result in subgroups of individuals with a very low probability of being exposed or unexposed.

Application: To illustrate the method's usefulness, we applied it to four datasets from recently published studies. The algorithm found the two previously reported subgroups with a structural violation of the positivity, and it identified nine additional subgroups suspected of presenting a violation. According to the medical experts, three were considered false positives (probably due

to sample-to-sample fluctuation, i.e., random violation) and six were structural violations of positivity.

C*onclusions:* The suggested algorithm allows one to easily and rapidly detect potential violations of the positivity assumption. We implemented the algorithm in the R package RISCA to facilitate its use and dissemination. Easy to use and independent of the inference method, we hope that it will improve the verification of the positivity assumption in causal studies.

Keywords: causal inference; decision tree; eligibility criteria; identifiability; positivity; target population.

Key Messages

- Causal inference requires the positivity assumption: each individual must have a non-zero probability of having each level of exposure.
- Statistical approaches were commonly employed to deal with positivity violations, leading to a shift in inference population and thus in estimand.
- We suggest a CART-based method for identifying subgroups of individuals potentially responsible for random or structural violations.
- We illustrate its usefulness with the reanalysis of four recently published studies, finding that the inclusion criteria could have been better defined in three of them.
- PCART is available in the R package RISCA.

Introduction

The positivity assumption, also known as experimental treatment assignment or common support, is a cornerstone to drawing causal inferences. The positivity assumption means that each individual has a non-zero probability of being exposed and unexposed.[1] Its violation may take two forms: a structural form or a random one. A structural violation is deterministic due to the inclusion of individuals who can theoretically never receive one of the levels of the studied exposition, resulting in an incorrect definition of the eligibility criteria. Contrariwise, a random violation occurs by chance, *i*.e., without any underlying cause.[2] Such a random violation is more likely in small samples or with rare exposure.

Alongside consistency and exchangeability, positivity is essential to the identifiability of the causal effect, *i*.e., the mapping of the statistical parameter obtained from the data and the causal quantity of interest (hereafter the estimand).[3] The target population is an integral part of the estimand.[4,5] Epidemiologists are commonly interested in the population average causal effect and approximate it by the sample average causal effect.[6] Random violations of positivity can be addressed by statistical procedures, such as trimming or truncating the propensity scores if these have been used.[7] Alternatively, decision trees have been used to estimate a causal effect in a subpopulation with respect to positivity.[8,9] However, all of these approaches change the characteristics of the sample, shift the population of inference, and target a different sample average causal effect.[5,10] The difference between the population defined by the eligibility criteria and the inference population casts doubt upon the external validity of the results.[11] Therefore, Platt *et al.*[12] argued that the exclusion of the subjects unlikely to be (un)exposed, by redefining the eligibility criteria, is transparent and leads to clear statements about the generalisability of the results. This approach is closely related to a well-defined target trial.[13,14]

Rather than using decision trees to directly estimate the causal effect in an ill-defined population free of positivity violations, we suggest using classification and regression trees (CART) to identify the subgroups of individuals potentially causing such violations. Since a decision tree consists of

a set of nodes (*i.e.*, binary decision rules) themselves subdivided into other nodes (and so on, see Supplementary Figure 1), using the exposure as the outcome and one or several covariates as predictors allows one to obtain an estimation of the treatment probability in each subgroup represented by the nodes.[15,16] In the presence of a node with an extreme exposure probability, we can therefore suspect a violation of the positivity assumption.

We sought to develop a CART-based algorithm to check for potential violations of the positivity assumption and help redefine the eligibility criteria. To facilitate the use of this method, we implemented it as a function in the R package *RISCA*.[17]

The paper proceeds as follows. We present the algorithm in the next section, then apply it to several datasets from recently published studies, and present the results in the third section. The last section offers discussions and practical recommendations.

The P(ositivity)-CART algorithm

The algorithm definition.

Consider (α, β, γ) the three parameters allowing one to define the positivity violation. The parameter α is the minimal percentage of the whole sample size needed to define a problematic subgroup. A positivity violation is defined by an exposure probability lower than β% or greater than 100-β% in a subgroup.[18] The parameter γ refers to the maximal number of predictors used to define a subgroup.

Based on these parameters, we suggest using CART to identify such subgroups. Figure 1 presents an overview of the proposed algorithm, which proceeds in several steps. The first step involves estimating one tree for each predictor and memorising the leaves (*i.e.*, the final nodes) corresponding to problematic subgroups according to the parameters α and β. If γ=1, the algorithm stops. Otherwise, if at least one problematic subgroup is identified in the first step, the

corresponding predictor(s) is(are) not considered in the next step (to limit the number of identified subgroups), which estimates one tree for all possible couples of predictors and memorises the leaves corresponding to problematic subgroups according to the parameters α and β. If γ=2, the algorithm stops. Otherwise, the third step consists of one tree for all possible trio of remaining covariates not involved in the previously identified subgroups, and so on.

The default values of (α, β, γ).

To the best of our knowledge, there is no consensus on the precise definition a subgroup presenting a positivity violation. Nevertheless, one can set α=5%, as several authors suggest trimming the weights at this level in inverse probability weighing analysis.[19–21] Following D'Amour *et al.*,[18] we set β=5%. Note that this value is also consistent with the propensity score literature.[7,22] By default, the parameter γ is arbitrarily set to two since one can question the relevance of the results for higher levels.

Of note, we do not use the pruning step of CART in order to create the vastest possible number of divisions, and therefore of subgroups. Previous research also showed that using such large trees allows one to overcome a time-consuming optimisation of the CART-specific parameters.[23] Therefore, we used the default values provided in the R package *rpart* (version 4.1-15).[24]

Applications

Context.

We applied the P-CART algorithm by reanalysing datasets from four recently published observational studies.[25–28] We aimed to validate the algorithm's capacity to re-identify the positivity violations previously reported based on expert knowledge and identify potential new

ones that were previously missed. We used the adjustment set of covariates as predictors in the PCART algorithm (Supplementary Table 1) since positivity is relevant only for them.[1,2] The authors restricted the studied populations in order to respect the positivity assumption. Therefore, we did not apply these specific eligibility criteria when evaluating whether the PCART algorithm allowed us to identify the positivity violations. We conducted complete case analyses. We categorized each continuous variable according to meaningful cut-offs before running the algorithm to avoid clinically insignificant subgroups. We performed the main analysis with the default values (α=5%, β=5%, and γ=2). We considered other values (α=1%, α=10%, β=1% and β=10%) as sensitivity analyses (Supplementary Tables 2-6). In a second time, we reanalysed the data by (i) excluding the clinically plausible subgroups identified by P-CART and (ii) applying the inverse probability weighting approach performed initially by the authors (see Austin and Stuart[29] for an introduction). We performed all the analyses using R version 3.6.0.[30]

Barbiturates during intensive care for patients with traumatic brain injury.

Léger *et al.*[25] investigated the impact of barbiturates on mortality by using a cohort made up of 1088 patients admitted into intensive care units for traumatic brain injury. This treatment may be offered to reduce patients' intracranial hypertension and the corresponding consequences in terms of brain damage. The authors excluded individuals older than 70 because the therapy is contraindicated for the elderly. We applied the P-CART algorithm to the entire cohort regardless of patient age. We identified nine subgroups potentially associated with positivity violations (Table 1). First, the P-CART algorithm confirmed the patients older than 75 as problematic, a greater threshold value compared to the one suggested by the authors. Second, it indicated that patients without osmotherapy at admission had a probability of receiving barbiturates lesser than 5%. This is clinically coherent since barbiturates are a last-line therapy and should only be offered after osmotherapy. The authors had not identified this issue, presumably because these patients could have received another second-line therapy. Third, P-CART identified four

subgroups composed of patients without intracranial hypertension at admission. This characteristic was not identified in the first iteration of P-CART because the treatment probability for patients without intracranial hypertension was equal to 5.1%, just above the threshold β. Note that these individuals can receive barbiturates in a preventive way if they are identified at a high risk of the event. Fourth, the algorithm detected that a lactatemia lower than 1mmol/L, an SAPS II score from 40 to 44 without severe trauma, and an SAPS II score from 25 to 55 along with a creatinine level in between 50 and 60mmol/L were associated with a probability of barbiturates under 5%. These three subgroups do not seem to have a clinical explanation and can be qualified as random violations due to sample-to-sample fluctuation.

Figure 2 shows the results of the initial analysis performed by the authors and the reanalysis based on the restricted sample. The exclusion of individuals without intracranial hypertension at admission alone reduced the sample by almost 70%. In the final sample (N=173), the odds ratio is 1.9 (rather than 2.2 in the whole sample) and became closer to statistical insignificance with a confidence interval at 95% (CI95%) from 1.0 to 3.5.

Kidney transplantations from marginal donors.

Querard et al.[26] compared grafts from standard and marginal donors, as defined by the expanded donor criteria. The two kinds of grafts differ in their intrinsic quality. Because of the shortage of kidneys for transplantation, the grafts from standard donors are preferentially attributed to young recipients, thus implying a positivity issue. The authors did not consider the recipient age among the eligibility criteria. In contrast, P-CART detected a problematic subgroup consisting of 391 recipients younger than 30 years (8.1% of the whole sample) with a probability of receiving a marginal graft of 3.1%. The exclusion of these individuals did not change the magnitude of the effect.

The hypothermic perfusion machine for marginal donors.

Foucher et al.[27] compared the hypothermic perfusion machine to static cold storage in kidney transplantations from expanded donors. The authors reduced the studied cohort to individuals younger than 45 because of a potential structural violation: the old-to-old graft allocation policy results in a lower susceptibility of younger candidates receiving marginal grafts. By using the entire cohort with no restriction on patient age (N=1978), the P-CART algorithm did not detect this issue. Indeed, only 32 and 44 individuals were less than 45 years old in the perfusion machine and cold storage groups, respectively. In contrast, P-CART suggested that individuals transplanted before 2015 in four anonymised centres could be problematic. The centres A, B, C and D included 126, 81, 218 and 41 patients during this period, respectively. These centres respectively attributed a perfusion machine with a probability of 0.8%, 3.7%, 4.7% or 2.4%. This violation seems plausible given that some hospitals were slower to adopt hypothermic perfusion machines, which were only introduced in 2010. Similarly, patients with a cold ischemia time under 20 hours and transplanted before 2013 (N=101) had a probability of receiving a graft under static cold storage lower than 5%. This second violation also seems plausible because the perfusion machines were preferentially attributed to grafts transplanted away from the harvesting site (*i.e.*, with a high cold ischemia time). This preferential attribution became less strict with the wider availability of perfusion machines over time. The exclusion of these two subgroups shifted the HR from 0.9 (CI95% 0.7 – 1.1) in the whole sample to 0.8 (CI95% 0.6 – 1.1) in the restricted sample. Furthermore, restricting the sample halved the 10-years survival probability difference.

Induction therapy in elderly kidney transplant recipients.

Masset et al.[28] compared the risk of adverse events following a kidney transplant in elderly recipients depending on their induction therapy: anti-thymocyte globulins versus basiliximab. We did not detect any positivity violations with P-CART, confirming the authors' statement.

Discussion

By reanalysing the data of four published studies, the P-CART algorithm confirmed the two subgroups with a structural positivity violation, as stated by the authors. It also identified nine new subgroups which may be problematic. Of the nine subgroups, six were likely structural positivity violations, which were missing from the authors' results. It illustrated that the PCART algorithm would have put them on the track of potential violations of positivity, although expert knowledge will always be necessary to distinguish random and structural violations of positivity.

As illustrated by the applications, the restriction of the sample can lead to a different estimation. Although the clinical conclusions were unaffected, the effects' magnitude was smaller in the restricted samples. Interestingly, the estimate of Querard *et al.*[26] was unaffected by the restriction. These authors targeted the average causal effect on the exposed, for which a weaker positivity assumption is necessary. Indeed, only the exposed individuals should have a non-null probability of being unexposed.[31] Therefore, their inference was not affected by this subgroup composed of unexposed individuals.

The P-CART algorithm we suggest is based on three tuning parameters that define the subgroup affected by the positivity violation: α the minimum percentage of patients included in the subgroup, β the maximum probability of being exposed or unexposed in the subgroup, and γ the maximum number of predictors outlining the subgroups. In our applications, we set the parameter α at 5%, in agreement with the literature. However, it could be increased when the sample size decreases as random violations are more likely to occur. We set the parameter β at 5%, again in line with the literature. Alternative values can be considered. For instance, Petersen *et al.*[7] and Crump *et al.*[32] argued in favour of using a β = 1% and 10%, respectively. Note that the exposure prevalence can also influence this choice, a small prevalence leading to a tighter definition of the positivity, thus a smaller β. The application on the dataset of Léger *et al.*[25] outlined the advantage of increasing the parameter γ to identify structural violations when

β is close to exposure prevalence. The parameter γ may also be increased with the sample size to refine the definition of the subgroups. Nevertheless, it seems unnecessary to define structural violations with a combination of more than three variables intended to represent the exposure's allocation process.

Other approaches can be employed for diagnosing positivity violations. When a propensity score is used, Cole and Hernán warned against the presence of extreme weights.[33] Petersen *et al.* provided a bootstrap-based tool to quantify the amount of bias due to such violations.[7] Westreich and Cole suggested computing all possible contingency tables to check for the presence of empty cells.[2] Unlike P-CART, the two first approaches cannot identify the subgroups causing a positivity violation, while the last approach is only feasible in lower-dimensional settings and cannot deal with continuous variables. We stress that positivity is not solely related to propensity score-based approaches, even if a vast amount of the positivity literature is found in this context. Outcome regression-based approaches also require this assumption, albeit they can extrapolate to the problematic subgroups in the case of random violations.[7,25] Randomized clinical trials may also present a lack of positivity when the sample size is not large enough to prevent sample-to-sample fluctuations.

Some limitations need to be mentioned. First, we only considered binary treatments, and further development is needed to consider multimodal or even continuous exposure. Second, we did not perform any simulation study to illustrate the predictive performances of the algorithm. We believe that the P-CART algorithm is a practical tool to help investigators and analysts to precise the target population. Last, we did not investigate the optimal set of tuning parameters, another perspective to be addressed in future work.

In conclusion, we illustrated that P-CART represents a helpful decision-making tool for checking for the targeted population in causal studies. It enables essential discussions about the susceptible structural violations before causal analysis (e.g., propensity scores, G-computation, doubly robust estimators or multiple regressions). We recommend using the P-CART algorithm with different

tuning parameters to get an overview of the potential violations. P-CART can lead to better transparency of the research process, a better generalisability of the results, and better implementation of the findings in real life for the ultimate benefit of patients.

Author Contributions

G.D. developed the algorithm, analysed the data, and drafted the manuscript. A.C. designed the study, analysed the data, and wrote the manuscript. All authors interpreted the data, revised the manuscript, as well as read and approved the final version of the manuscript. A.C. is the guarantor of the article.

Acknowledgements

The authors would like to thank the members of AtlanREA and DIVAT groups for their involvement in the study, the physicians who helped recruit patients and all patients who participated in this study. We also thank the clinical research associates who participated in the data collection. The analysis and interpretation of these data are the responsibility of the authors.

Funding

This work was partially supported by a public grant overseen by the French National Research Agency (ANR) to create the Common Laboratory RISCA (Research in Informatic and Statistic for Cohort Analyses, www.labcom-risca.com, reference: ANR-16-LCV1-0003-01). The funder had no role in study design; analysis, and interpretation of data; writing the report; and the decision to submit the report for publication.

Conflict of interest: None declared.

References

1. Hernán M, Robins JM. Causal Inference: What if? Boca Raton: Chapman & Hall/CRC; 2020.

2. Westreich D, Cole SR. Invited Commentary: Positivity in Practice. *Am J Epidemiol*. 2010 Mar 15;**171**(6):674–677.

3. Petersen ML, Laan MJ van der. Causal Models and Learning from Data: Integrating Causal Modeling and Statistical Estimation. *Epidemiology*. 2014 May;**25**(3):418–426.

4. Maldonado G, Greenland S. Estimating causal effects. *International Journal of Epidemiology*. 2002 Apr 1;**31**(2):422–429.

5. Lundberg I, Johnson R, Stewart BM. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. *Am Sociol Rev*. 2021 Jun 1;**86**(3):532–565.

6. Balzer LB. 'All Generalizations Are Dangerous, Even This One.'-Alexandre Dumas. *Epidemiology*. 2017;**28**(4):562–566.

7. Petersen ML, Porter KE, Gruber S, Wang Y, Laan MJ van der. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res*. 2012 Feb;**21**(1):31–54.

8. Hill J, Su Y-S. Assessing lack of common support in causal inference using Bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *Ann Appl Stat*. 2013 Sep;**7**(3):1386–1420.

9. Kang J, Chan W, Kim M-O, Steiner PM. Practice of causal inference with the propensity of being zero or one: assessing the effect of arbitrary cutoffs of propensity scores. *Commun Stat Appl Methods*. 2016 Jan;**23**(1):1–20.

10. Zhu Y, Hubbard RA, Chubak J, Roy J, Mitra N. Core Concepts in Pharmacoepidemiology: Violations of the Positivity Assumption in the Causal Analysis of Observational Data: Consequences and Statistical Approaches. *Pharmacoepidemiol Drug Saf*. 2021 Nov;**30**(11):1471–1485.

11. Nethery RC, Mealli F, Dominici F. Estimating population average causal effects in the presence of non-overlap: The effect of natural gas compressor station exposure on cancer mortality. *Ann Appl Stat*. 2019 Jun;**13**(2):1242–1267.

12. Platt RW, Delaney JAC, Suissa S. The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding. *Eur J Epidemiol*. 2012 Feb;**27**(2):77–83.

13. Didelez V. Commentary: Should the analysis of observational data always be preceded by specifying a target experimental trial? *International Journal of Epidemiology*. 2016 Dec 1;**45**(6):2049–2051.

14. García-Albéniz X, Hsu J, Hernán MA. The value of explicitly emulating a target trial when using real world evidence: an application to colorectal cancer screening. *Eur J Epidemiol*. 2017 Jun 1;**32**(6):495–500.

15. Westreich D, Lessler J, Funk MJ. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology*. 2010 Aug 1;**63**(8):826–833.

16. Bi Q, Goodman KE, Kaminsky J, Lessler J. What is Machine Learning? A Primer for the Epidemiologist. *Am J Epidemiol*. 2019 Dec;**188**(12):2222–2239.

17. Foucher Y, Le Borgne, Dantan E, Gillaizeau F, Chatton A, Combescure C. RISCA: Causal Inference and Prediction in Cohort-Based Analyses. 2019. https://CRAN.R-project.org/package=RISCA (20 October 2021, date last accessed).

18. D'Amour A, Ding P, Feller A, Lei L, Sekhon J. Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics*. 2021 Apr 1;**221**(2):644–654.

19. Stürmer T, Rothman KJ, Avorn J, Glynn RJ. Treatment Effects in the Presence of Unmeasured Confounding: Dealing With Observations in the Tails of the Propensity Score Distribution—A Simulation Study. *Am J Epidemiol*. 2010 Oct 1;**172**(7):843–854.

20. Glynn RJ, Lunt M, Rothman KJ, Poole C, Schneeweiss S, Stürmer T. Comparison of alternative approaches to trim subjects in the tails of the propensity score distribution. *Pharmacoepidemiol Drug Saf*. 2019 Oct;**28**(10):1290–1298.

21. Lee BK, Lessler J, Stuart EA. Weight Trimming and Propensity Score Weighting. *PLOS ONE*. 2011 Mar 31;**6**(3):e18174.

22. Zhou Y, Matsouaka RA, Thomas L. Propensity score weighting under limited overlap and model misspecification. *Stat Methods Med Res*. 2020 Dec 1;**29**(12):3721–3756.

23. Mantovani RG, Horváth T, Cerri R, Junior SB, Vanschoren J, Leon Ferreira de Carvalho ACP de. An empirical study on hyperparameter tuning of decision trees. *arXiv:1812.02207*. 2019 Feb 12;**preprint: not peer reviewed**.

24. Therneau TM, Atkinson B. rpart: Recursive Partitioning and Regression Trees. 2019. https://CRAN.R-project.org/package=rpart (20 October 2021, date last accessed).

25. Léger M, Chatton A, Le Borgne F, Pirracchio R, Sigismond L, Foucher Y. Causal inference in case of near-violation of positivity: comparison of methods. *Biom J*. 2021:In press.

26. Querard A-H, Foucher Y, Combescure C, et al. Comparison of survival outcomes between Expanded Criteria Donor and Standard Criteria Donor kidney transplant recipients: a systematic review and meta-analysis. *Transpl Int*. 2016 Apr 1;**29**(4):403–415.

27. Foucher Y, Fournier M-C, Legendre C, et al. Comparison of machine perfusion versus cold storage in kidney transplant recipients from expanded criteria donors: a cohort-based study. *Nephrol Dial Transplant*. 2020 Jun 1;**35**(6):1043–1070.

28. Masset C, Boucquemont J, Garandeau C, et al. Induction Therapy in Elderly Kidney Transplant Recipients With Low Immunological Risk. *Transplantation*. 2020 Mar;**104**(3):613–622.

29. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. *Stat Med*. 2015 Dec 10;**34**(28):3661–3679.

30. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/

31. Pirracchio R, Carone M, Rigon MR, Caruana E, Mebazaa A, Chevret S. Propensity score estimators for the average treatment effect and the average treatment effect on the treated may yield very different estimates. *Stat Methods Med Res*. 2016 Oct;**25**(5):1968–1954.

32. Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. *Biometrika*. 2009 Mar 1;**96**(1):187–199.

33. Cole SR, Hernan MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol*. 2008 Jul 15;**168**(6):656–664.

Table 1: Subgroups of patients identified by the PCART algorithm as potential sources of non-positivity.

Authors	*Sample size*	Problematic subgroup (*n*, %)	Identified by the authors	Clinically plausible	Violation type
Léger *et al.*[25]	1088	Age ≥ 75 years (73, 6.7)	Yes	Yes	Structural
		No osmotherapy at admission (732, 67.3)	No	Yes	Structural
		25 ≤ SAPS II score < 55 & 50 ≤ Creatinine < 60 (135, 12.4)	No	No	Random
		No IH at admission nor history of head trauma (710, 65.3)	No	Yes	Structural
		No IH & severe trauma at admission (385, 35.4)	No	Yes	Structural
		No IH at admission & Creatinine < 150 (740, 68.0)	No	Yes	Structural
		No IH at admission & SAPS II score < 55 (532, 48.9)	No	Yes	Structural
		Lactatemia < 1 (197, 18.1)	No	No	Random
		40 ≤ SAPS II score < 45 & No severe trauma (69, 6.3)	No	No	Random
Querard *et al.*[26]	3422	Recipient age < 30 years (391, 8.1)	Yes	Yes	Structural
Foucher *et al.*[27]	1978	Transplant before 2014 & A - D centres[a] (376, 19.0)	No	Yes	Structural
		Transplant before 2012 & CIT ≥ 20h (101, 5.1)	No	Yes	Structural
Masset *et al.*[28]	383	None detected[b]	-	-	-

Abbreviations: BMI, Body Mass Index; CIT, Cold Ischemia Time and IH, Intracranial Hypertension. Creatinine and lactatemia were in mmol/L.

[a] The centres were anonymised. [b] Regardless of the sample and the adjustment set.

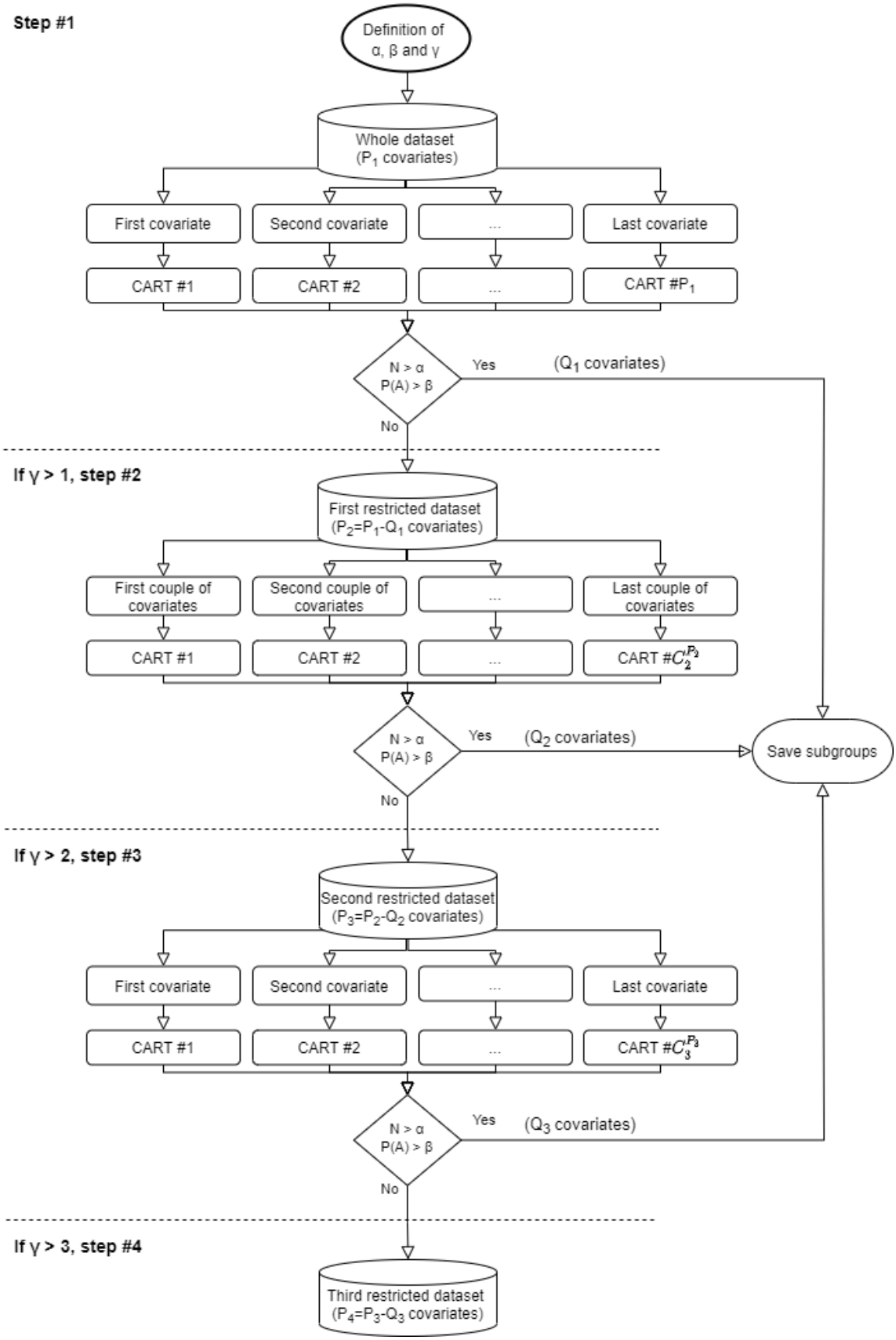


Figure 1: Flow-chart presenting the PCART algorithm. A: Treatment; N: Percentage of the whole sample contained in the subgroup; α, β and γ were user-supplied hyperparameters.

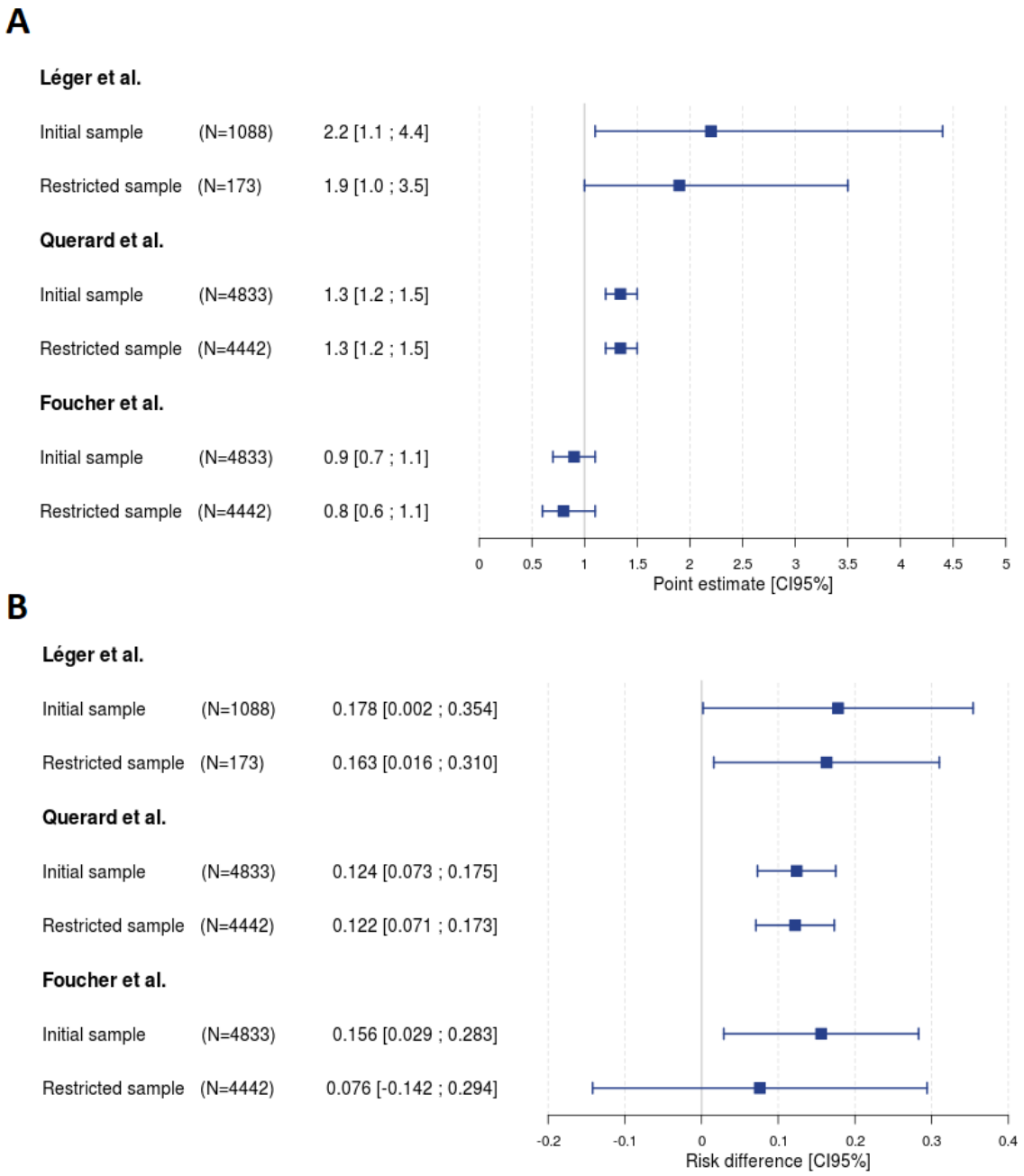


Figure 2: Results of the reanalyses considering the restricted datasets on the relative (A) and absolute scales (B). Léger et al. presents an odds-ratio and a risk difference, while the others present a hazard ratio and a 10years survival difference.

Annexe D

Étude de simulation de Chatton *et al.* Comparaison des principales méthodes d'estimation causale selon les covariables incluses

SCIENTIFIC REPORTS
nature research

Check for updates

OPEN

G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study

Arthur Chatton[1,2], Florent Le Borgne[1,2], Clémence Leyrat[1,3], Florence Gillaizeau[1,4], Chloé Rousseau[1,4,5], Laetitia Barbin[4], David Laplaud[4,6], Maxime Léger[1,7], Bruno Giraudeau[1,8] & Yohann Foucher[1,4]✉

Controlling for confounding bias is crucial in causal inference. Distinct methods are currently employed to mitigate the effects of confounding bias. Each requires the introduction of a set of covariates, which remains difficult to choose, especially regarding the different methods. We conduct a simulation study to compare the relative performance results obtained by using four different sets of covariates (those causing the outcome, those causing the treatment allocation, those causing both the outcome and the treatment allocation, and all the covariates) and four methods: g-computation, inverse probability of treatment weighting, full matching and targeted maximum likelihood estimator. Our simulations are in the context of a binary treatment, a binary outcome and baseline confounders. The simulations suggest that considering all the covariates causing the outcome led to the lowest bias and variance, particularly for g-computation. The consideration of all the covariates did not decrease the bias but significantly reduced the power. We apply these methods to two real-world examples that have clinical relevance, thereby illustrating the real-world importance of using these methods. We propose an R package *RISCA* to encourage the use of g-computation in causal inference.

The randomised controlled trial (RCT) remains the primary design for evaluating the marginal (population average) causal effect of a treatment, *i.e.*, the average treatment effect between two hypothetical worlds where: i) everyone is treated and ii) everyone is untreated[1]. Indeed, a well-designed RCT with a sufficient sample size ensures the baseline comparability between groups, thus allowing the estimation of a marginal causal effect. Nevertheless, it is well established that RCT is performed under optimal circumstances (*e.g.*, over-representation of treatment-adherent patients, low frequency of morbidity), which may be different from real-life practices[2]. Observational studies have the advantage of limiting the issue of external validity, but treated and untreated patients are often non-comparable, leading to a high risk of confounding bias.

To reduce such confounding bias, the vast majority of observational studies have been based on multivariable models (mainly linear, logistic, or Cox models), allowing for the direct estimation of conditional (subject-specific) effects, *i.e.*, the average effect across sub-populations of subjects who share the same characteristics. Several

[1]INSERM UMR 1246 - SPHERE, Université de Nantes, Université de Tours, Nantes, France. [2]A2COM-IDBC, Pacé, France. [3]Department of Medical Statistics & Cancer Survival Group, London School of Hygiene and Tropical Medicine, London, UK. [4]Centre Hospitalier Universitaire de Nantes, Nantes, France. [5]INSERM CIC1414, CHU Rennes, Rennes, France. [6]Centre de Recherche en Transplantation et Immunologie INSERM UMR1064, Université de Nantes, Nantes, France. [7]Département d'Anesthésie-Réanimation, Centre Hospitalier Universitaire d'Angers, Angers, France. [8]INSERM CIC1415, CHRU de Tours, Tours, France. ✉e-mail: Yohann.Foucher@univ-nantes.fr

methods have been proposed to estimate marginal causal effects in observational studies, amongst which propensity score (PS)-based methods are increasingly used in epidemiology and medical research[3].

Propensity score-based methods make use of the PS in four different ways to account for confounding, namely matching, stratification, conditional adjustment[4] and inverse probability of treatment weighting (IPTW)[5]. Stratification and conditional adjustment on PS are associated with the highest bias[6–8], because the two methods estimate the conditional treatment effect rather than the marginal causal effect. Matching on PS remains the most common approach with a usage rate of 83.8% in 303 surgical studies using PS-based methods[9] and 68.9% in 296 medical studies (without restriction regarding the field) also using PS-methods[10]. The IPTW appears to be less biased and associated with a lower variance than matching in several studies[8,11–14]. Nevertheless, in particular settings, full matching (FM) was associated with lower mean square error (MSE) in other studies[15–17].

Multivariable models, even non-linear ones, can also be used to indirectly estimate the marginal causal effect with g-computation (GC)[18]. This method is also called the parametric g-formula[1] or (g-)standardisation[19] in the literature. Snowden *et al.*[20] and Wang *et al.*[21] detailed the corresponding methodology for estimating the average treatment (*i.e.*, marginal causal) effect on the entire population (ATE) or only on the treated (ATT), respectively. The ATE is the average effect, at the population level, of moving an entire population from untreated to treated. The ATT is the average effect of treatment on those subjects who ultimately received the treatment[22]. Furthermore, some authors[23,24] have proposed combinations of GC and PS to improve the estimation of the marginal causal effect. These methods are known as doubly robust estimators (DRE) because they require the specification of both the outcome (for GC) and treatment allocation (for PS) mechanisms to minimise the impact of model misspecification. Indeed, these estimators are consistent as long as either the outcome model or the treatment model is estimated correctly[25].

Each of these methods carries out the adjustment in different ways, but all of these methods rely on the same condition: a correct specification of the PS or the outcome model[1]. In practice, a common issue is choosing the set of covariates to include to obtain the best performance in terms of bias and precision. Three simulation studies[7,26,27] have investigated this issue for PS-based methods. They studied four sets of covariates: those causing the outcome, those causing the treatment allocation, those are a common cause of both the treatment allocation and the outcome, and all the covariates. For the rest of this paper, we called these strategies the *outcome set*, the *treatment set*, the *common set* and the *entire set*, respectively. These studies argued in favour of the outcome or common sets for PS-based methods, but it is not immediately clear that such works will generalise to other methods of causal inference. Brookhart *et al.*[26] and Lefebvre *et al.*[27] focused on count and continuous outcomes. Austin *et al.*[7] investigated binary outcomes on matching, stratification and adjustment on PS. However, GC and DRE also require the correct specification of the outcome model with a potentially different set of covariates. Recent works have shown that efficiency losses can accompany the inclusion of unnecessary covariates[28–31]. De Luna *et al.*[32] also highlighted the variance inflation caused by the treatment set. In contrast, VanderWeele and Shpitser[33] suggested the inclusion of both the outcome and the treatment sets.

Before selecting the set of covariates, one needs to select the method to employ. Several studies have compared the performances of GC, PS-based methods and DRE in a point treatment study to estimate the ATE[13,23,25,34–36]. Half of these studies investigated a binary outcome[13,25,34]. Only Colson *et al.*[17] studied the ATT, but they focused on a continuous outcome. Except in Neugebauer and van der Laan[25], these studies only investigated the ATE (or ATT) defined as a risk difference. The CONSORT recommended the presentation of both the absolute and the relative effect sizes for a binary outcome, "*as neither the relative measure nor the absolute measure alone gives a complete picture of the effect and its implications*"[37]. None of these studies was interested in the set of covariates necessary to obtain the best performance.

In our study, we sought to compare different sets of covariates to consider to estimate a marginal causal effect. Moreover, we compared GC, PS-based methods and DRE for both the ATE and ATT, either in terms of risk difference or marginal causal OR. Three main types of outcome are used in epidemiology and medical research: continuous, binary and time-to-event outcomes. We focused on a binary outcome because i) a continuous outcome is often appealing for linear regression where the two conditional and marginal causal effects are collapsible[38], and ii) time-to-event analyses present additional methodological difficulties, such as the time-dependant covariate distribution[39]. We also limit our study to a binary treatment, as in the current literature, and the extension to three or more modalities is beyond the scope of our study.

The paper is structured as follows. In the next section, the methods are detailed. The third section presents the design and results of the simulations. In the fourth section, we consider two real data sets. Finally, we discuss our results in the last section.

Methods

Setting and notations. Let A denote the binary treatment of interest ($A = 1$ for treated patients and 0 otherwise), Y denote the binary outcome ($Y = 1$ for events and 0 otherwise), and L denote a set of baseline covariates. Consider a sample of size n in which one can observe the realisations of these random variables: a, y, and l, respectively. Define $\pi_a = E(P(Y = 1|do(A = a), L))$ or $\pi_a = E(P(Y = 1|do(A = a), L)|A = 1)$ as the expected proportions of event if the entire (ATE) or the treated (ATT) populations were treated ($do(A = 1)$) or untreated ($do(A = 0)$), respectively[40]. From these probabilities, the risk difference can be estimated as $\Delta\pi = \pi_1 - \pi_0$ and the log of the marginal causal OR estimated as $\theta = \text{logit}(\pi_1)/\text{logit}(\pi_0)$, where $\text{logit}(\bullet) = \log(\bullet/(1-\bullet))$. The methods described bellow allow for the estimation of both the ATE and the ATT effects.

Causal inference requires the three following assumptions, called *identifiability conditions*: i) The values of exposure under comparisons correspond to well-defined interventions that, in turn, correspond to the versions of treatment in the data. ii) The conditional probability of receiving every value of treatment, though not decided by the investigators, depends only on the measured covariates. iii) The conditional probability of receiving

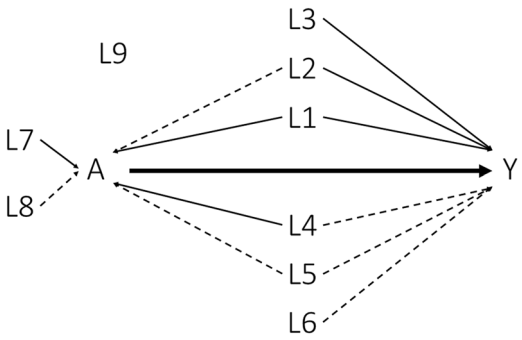


Figure 1. Causal diagram. Solid lines corresponded to a strong association (OR = 6.0) and dashed lines to a moderate one (OR = 1.5).

every value of the treatment is greater than zero, *i.e.*, is positive. These assumptions are known as *consistency*, *(conditional) exchangeability* and *positivity*, respectively[1]. However, PS-based methods rely on treatment allocation modelling to obtain a pseudo-population in which the confounders are balanced across treatment groups. Covariate balance can be checked by computing the standardised difference of the covariates included in the PS between the two treatment groups[10]. In contrast, GC relies on outcome modelling to predict hypothetical outcomes for each subject under each treatment regimen. Note that one can ignore the lack of positivity if one is willing to rely on Q-model extrapolation[1]. As is the case for standard regression models, these methods also require the assumptions of no interference, no measurement error and no model misspecification.

Weighting on the inverse of the propensity score. Formally, the PS is $p_i = P(A_i = 1|L_i)$, *i.e.* the probability that subject i ($i = 1, \ldots, n$) will be treated according to his or her characteristics L_i at the time of the treatment allocation[4]. It is often estimated using a logistic regression. The IPTW makes it possible to reduce confounding by correcting the contribution of each subject i by a weight ω_i. For ATE, Xu *et al.*[41] defined $\omega_i = A_i P(A_i = 1)/p_i + (1 - A_i)P(A_i = 0)/(1 - p_i)$. The use of stabilised weights has been shown to produce a suitable estimate of the variance even when there are subjects with extremely large weights[5,41]. For ATT, Morgan and Todd[42] defined $\omega_i = A_i + (1 - A_i)p_i/(1 - p_i)$. Based on ω_i, the following weighted univariate logistic regression can be fitted: $\text{logit}\{P(Y = 1|A)\} = \hat{\alpha}_0 + \hat{\alpha}_1 A$, resulting in $\hat{\pi}_0 = (1 + \exp(-\hat{\alpha}_0))^{-1}$, $\hat{\pi}_1 = (1 + \exp(-\hat{\alpha}_0 - \hat{\alpha}_1))^{-1}$, and $\hat{\theta} = \hat{\alpha}_1$. To obtain $\widehat{var}(\hat{\theta})$, we used a robust sandwich-type variance estimator[5] with the R package *sandwich*[43].

Full Matching on the propensity score. The FM minimises the average within-stratum differences in the PS between treated and untreated subjects[16]. Then, two weighting systems can be applied in each stratum, making it possible to estimate either the ATE or the ATT unlike other matching methods which can only estimate the ATT[44]. If t and u denote the number of treated and untreated subjects in a given stratum, one can define the weight for a subject i in this stratum as $\omega_i = A_i P(A = 1)(t + u)/u + (1 - A_i)(1 - P(A = 1))(t + u)/t$ for ATE and $\omega_i = A_i + (1 - A_i)t/u$ for ATT[16]. In the latter case, the weights of untreated subjects are rescaled such that the sum of the untreated weights across all the matched sets is equal to the number of untreated subjects: $\widetilde{\omega}_i = \omega_i \times \sum_{j=1}^{n}(1 - A_j)/\sum_{j=1}^{n}\omega_j(1 - A_j)$[45]. From the resulting paired data set, we fitted a weighted univariate logistic regression, and the rest of the data analysis is tantamount to IPTW. We used the R package *MatchIt*[45] to generate the pairs.

G-computation. Consider the following multivariable logistic regression $\text{logit}\{P(Y = 1|A, L)\} = \gamma A + \beta L$. This regression is frequently called the *Q-model*[20]. Once fitted, one can compute for all subjects $\hat{P}(Y_i = 1|do(A_i = 1), L_i)$ and $\hat{P}(Y_i = 1|do(A_i = 0), L_i)$, *i.e.* the two expected probabilities of events if they were treated or untreated[20]. For ATE, one can then obtain $\hat{\pi}_a = n^{-1}\sum_i \hat{P}(Y_i = 1|do(A_i = a), L_i)$. The same procedure can be performed amongst the treated patients for ATT[21]. For implementation in practice, consider a treated subject ($A_i = 1$) included in the fit of the Q-model. Thanks to this model, one can then compute for this subject his or her predicted probabilities of the event if he or she received the treatment ($do(A_i = 1)$) or not ($do(A_i = 0)$). Computing these predicted probabilities for all the subjects, one can obtain two vectors of probabilities if the entire sample were treated or not. The corresponding means correspond to $\hat{\pi}_1$ and $\hat{\pi}_0$, respectively. We obtained $\widehat{var}(\hat{\theta})$ by simulating the parameters of the multivariable logistic regression assuming a multinormal distribution[46]. Note that we could have used bootstrap resampling instead. However, regarding the computational burden of bootstrapping and the similar results obtained by Aalen *et al.*[46], the variance estimates in the simulation study were only based on parametric simulations. We used both bootstrap resampling and parametric simulations in the applications.

Targeted Maximum Likelihood Estimator. Amongst the several existing DREs, we focused on the targeted maximum likelihood estimator (TMLE)[24], for which estimators of ATE and ATT have been proposed[47]. The TMLE begins by fitting the Q-model to estimate the two expected hypothetical probabilities of events $\hat{\pi}_1$ and $\hat{\pi}_0$. An additional "targeting" step involves estimation of the treatment allocation mechanism, *i.e.*, the PS

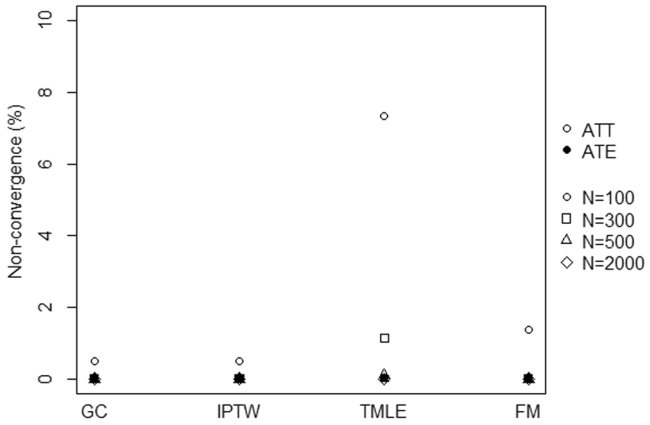


Figure 2. Percentage of simulation iterations which did not converge according to the methods.

$P(A_i = 1|L_i)$, which is then used to update the initial estimates obtained by GC. In the presence of residual confounding, the PS provides additional information to improve the initial estimates. Finally, the updated estimates of $\hat{\pi}_1$ and $\hat{\pi}_0$ are used to generate $\widehat{\Delta\pi}$ or $\hat{\theta}$. We used the efficient influence curve to obtain standard errors[47,48]. A recent tutorial provides a step-by-step guided implementation of TMLE[49].

Simulation study

Design. We used a close data generating procedure from previous studies on PS models[7,50]. We generated the data in three steps. i) Nine covariates ($L_1, \ldots, L_9$) were independently simulated from a Bernoulli distribution with a parameter equal to 0.5 for all covariates. ii) We generated the treatment A according to a Bernoulli distribution with a probability obtained by the logistic model with the following linear predictor: $\gamma_0 + \gamma_1 L_1 + \cdots + \gamma_9 L_9$. We fixed the parameter γ_0 at −3.3 or −5.2 to obtain a percentage of treated patients equal to 50% for scenarios related to ATE and 20% for ATT, respectively. iii) We simulated the event Y using a Bernoulli distribution with a probability obtained by the logistic model with the following linear predictor: $\beta_0 + \beta_1 A + \beta_2 L_1 + \cdots + \beta_{10} L_9$. We set the parameter β_1 for a conditional OR at 0 (the null hypothesis is no treatment effect) or 2 (the alternative hypothesis is a negative impact of treatment). We also fixed the parameter β_0 at −3.65 and −3.5 to obtain a percentage of the event close to 50% in ATE and ATT, respectively. Figure 1 presents the values of the regression coefficients γ_1 to γ_9 and β_1 to β_{10}. We considered four covariates sets as explained in the introduction: the outcome set included the covariates L_1 to L_6, the treatment set included the covariates L_1, L_2, L_4, L_5, L_7, L_8, the common set included the covariates L_1, L_2, L_4, L_5, and the entire set included the covariates L_1 to L_9. For each of the four methods and the four covariate sets, we studied the performance under different sample sizes: $n = 100$, 300, 500 and 2000. For each scenario, we randomly generated 10 000 data sets. We computed the theoretical values of π_1 and π_0 by averaging the values of π_1 and π_0 obtained from univariate logistic models (treatment as the only covariate) fitted from data sets simulated as above, except that the treatment A was simulated independently of the covariates L[50]. We reported the following criteria: i) the percentage of non-convergence, ii) the mean absolute bias (*e.g.*, $E(\hat{\theta}) - \theta$), iii) the MSE ($E[(\hat{\theta} - \theta)^2]$), the variance estimation bias $\left[\text{VEB} = 100 \times \left(\sqrt{E[\widehat{V}ar(\hat{\theta})]}/\sqrt{Var(\hat{\theta})} - 1\right)\right]$[51], the empirical coverage rate of the nominal 95% confidence intervals (CIs), defined as the percentage of 95% CI including the theoretical value, the type I error, defined as the percentage of rejection of the null hypothesis under the null hypothesis, and the statistical power, defined as the percentage of rejections of the null hypothesis under the alternative hypothesis. The MSE was our primary performance measure of interest because it combines bias and variance. We assumed that the identifiability conditions hold in these scenarios. We further performed the same simulations by omitting L_1 in the PS or in the Q-model to evaluate the impact of an unmeasured confounder. We performed all the analyses using R version 3.6.0[52].

Results

Convergence. Non-convergence only occurred for ATT estimation when sample sizes were lower or equal to 300 subjects (see Fig. 2). The GC, IPTW and FM had a minimal convergence percentage higher than 98%, even under small sample size (n = 100). Similarly, TMLE experienced some difficulty in converging for ATT estimation in the medium-sized sample (n = 300). However, they experienced severe difficulty in converging in the small sample with a convergence percentage of approximately 92%.

Mean bias. As expected with the common set, the mean absolute bias of θ was close to zero for GC, IPTW and TMLE when the three identifiability assumptions hold with a maximum at −0.028 given moderate sample size (n = 300) under the alternative hypothesis for ATT estimation (Table 1). Note that the three other covariate sets led to a bias close to zero with a maximum of 0.053 for TMLE with the entire set given small sample size (n = 100) under the alternative hypothesis for ATE estimation (Table 2). Furthermore, FM was also associated with a similar bias with a maximum of 0.082 given a small sample size (n = 100), with the treatment set under the alternative hypothesis for the ATE estimation. With an unmeasured confounder, the bias increased in all scenarios with a

n	method	selection strategy	mean bias				log OR				
			π_0	π_1	$\Delta\pi$	log OR	MSE	MSE*	VEB (%)	coverage (%)	power (%)
100	GC	outcome	0.000	−0.001	−0.001	0.012	0.526	0.716	−6.2	94.1	17.7
		treatment	0.002	−0.001	−0.003	0.006	0.580	0.786	−5.7	94.1	14.0
		common	0.002	−0.001	−0.003	0.006	0.552	0.735	−4.2	94.8	15.1
		entire	−0.001	−0.001	−0.001	0.013	0.558	0.768	−8.8	93.3	16.9
	IPTW	outcome	0.000	−0.001	−0.001	0.008	0.578	0.727	10.8	97.3	7.8
		treatment	−0.000	−0.001	−0.001	0.000	0.716	0.837	−1.2	95.1	9.8
		common	0.002	−0.001	−0.003	0.003	0.587	0.743	6.6	96.8	8.8
		entire	−0.003	−0.001	0.002	0.005	0.741	0.838	−1.5	95.2	9.6
	TMLE	outcome	−0.001	−0.001	0.000	0.002	0.694	0.794	30.0	95.7	5.8
		treatment	0.000	−0.001	−0.001	−0.020	0.876	0.955	183.3	98.8	1.0
		common	−0.000	−0.001	−0.001	−0.001	0.702	0.794	10.4	95.3	7.3
		entire	−0.003	−0.001	0.001	−0.013	0.886	0.953	412.2	98.8	0.5
	FM	outcome	−0.004	−0.001	0.003	0.022	0.665	0.787	−16.7	90.1	18.9
		treatment	−0.006	−0.001	0.004	0.017	0.822	0.911	−32.3	81.3	25.2
		common	−0.001	−0.001	−0.000	0.010	0.653	0.795	−15.3	91.0	17.5
		entire	−0.008	−0.001	0.006	0.022	0.842	0.921	−33.8	80.3	26.7
300	GC	outcome	0.001	−0.001	−0.002	−0.021	0.283	0.555	−1.6	94.5	43.6
		treatment	0.002	−0.001	−0.003	−0.024	0.319	0.606	−2.3	94.3	35.2
		common	0.002	−0.001	−0.003	−0.023	0.304	0.561	−1.5	94.8	38.5
		entire	0.001	−0.001	−0.002	−0.022	0.297	0.600	−2.6	94.0	39.9
	IPTW	outcome	0.002	−0.001	−0.003	−0.027	0.301	0.556	16.4	97.9	24.0
		treatment	0.001	−0.001	−0.002	−0.026	0.372	0.628	6.6	96.2	21.4
		common	0.003	−0.001	−0.004	−0.028	0.318	0.563	9.1	96.8	26.1
		entire	0.001	−0.001	−0.002	−0.025	0.361	0.622	11.7	97.2	20.0
	TMLE	outcome	0.000	−0.001	−0.001	−0.023	0.358	0.577	−2.3	93.6	29.0
		treatment	0.002	−0.001	−0.003	−0.035	0.454	0.683	51.2	99.1	6.8
		common	0.001	−0.001	−0.002	−0.023	0.378	0.582	−3.5	93.0	26.5
		entire	0.002	−0.001	−0.003	−0.035	0.432	0.674	81.8	99.3	4.4
	FM	outcome	−0.000	−0.001	−0.001	−0.020	0.351	0.579	−11.7	91.9	37.2
		treatment	−0.001	−0.001	−0.000	−0.022	0.444	0.656	−30.2	82.7	38.9
		common	0.001	−0.001	−0.002	−0.024	0.363	0.587	−14.6	90.4	36.9
		entire	−0.001	−0.001	0.000	−0.020	0.439	0.662	−29.3	83.2	39.1
500	GC	outcome	0.001	−0.001	−0.002	−0.014	0.217	0.509	−1.1	94.7	64.5
		treatment	0.001	−0.001	−0.002	−0.014	0.245	0.556	−1.5	94.4	53.6
		common	0.001	−0.001	−0.002	−0.015	0.233	0.618	−0.8	94.8	57.6
		entire	0.001	−0.001	−0.002	−0.014	0.228	0.552	−2.0	94.2	60.5
	IPTW	outcome	0.002	−0.001	−0.003	−0.019	0.230	0.509	16.5	97.9	43.3
		treatment	0.000	−0.001	−0.001	−0.013	0.285	0.574	6.8	96.6	35.4
		common	0.002	−0.001	−0.003	−0.018	0.244	0.514	9.2	96.8	43.7
		entire	0.000	−0.001	−0.001	−0.014	0.274	0.571	12.3	97.2	33.9
	TMLE	outcome	0.001	−0.001	−0.002	−0.015	0.272	0.521	−4.7	93.4	48.5
		treatment	0.001	−0.001	−0.002	−0.018	0.347	0.618	35.0	99.1	15.9
		common	0.000	−0.001	−0.001	−0.013	0.289	0.527	−4.8	93.1	43.7
		entire	0.001	−0.001	−0.002	−0.019	0.328	0.611	51.1	99.3	12.9
	FM	outcome	0.001	−0.001	−0.002	−0.015	0.265	0.525	−9.9	92.4	53.0
		treatment	−0.001	−0.001	−0.000	−0.011	0.346	0.597	−31.0	82.7	51.7
		common	0.001	−0.001	−0.001	−0.014	0.283	0.530	−15.8	90.1	52.3
		entire	−0.002	−0.001	0.001	−0.008	0.340	0.596	−29.8	83.2	52.6
2000	GC	outcome	0.000	0.000	−0.000	−0.002	0.108	0.479	−1.7	94.7	99.6
		treatment	0.001	0.000	−0.000	−0.003	0.122	0.524	−1.2	94.8	98.6
		common	0.001	0.000	−0.000	−0.003	0.116	0.480	−0.9	94.7	99.1
		entire	0.000	0.000	−0.000	−0.002	0.113	0.523	−1.8	94.5	99.4
	IPTW	outcome	0.002	0.000	−0.001	−0.006	0.113	0.478	16.3	97.6	98.1
		treatment	0.000	0.000	−0.000	−0.002	0.138	0.539	7.9	96.4	93.0
		common	0.002	0.000	−0.001	−0.006	0.120	0.480	9.4	97.0	97.7
		entire	0.000	0.000	−0.000	−0.002	0.131	0.537	13.9	97.4	93.6

Continued

n	method	selection strategy	mean bias				log OR				
			π_0	π_1	$\Delta\pi$	log OR	MSE	MSE*	VEB (%)	coverage (%)	power (%)
2000	TMLE	outcome	0.001	0.000	−0.000	−0.002	0.132	0.483	−5.9	93.3	97.5
		treatment	0.000	0.000	0.000	−0.002	0.169	0.568	18.2	98.2	71.8
		common	−0.000	0.000	0.000	−0.000	0.142	0.486	−5.6	93.6	95.5
		entire	0.001	0.000	−0.000	−0.004	0.158	0.565	23.5	98.6	75.3
	FM	outcome	0.000	0.000	−0.000	−0.002	0.134	0.484	−12.0	91.6	97.7
		treatment	0.001	0.000	−0.000	−0.005	0.203	0.548	−41.6	74.6	89.9
		common	0.001	0.000	−0.000	−0.003	0.149	0.485	−20.5	88.5	96.7
		entire	0.000	0.000	0.000	−0.002	0.162	0.543	−26.9	84.5	94.8

Table 1. Simulation results comparing the ATT estimation under the alternative hypothesis. *MSE in the presence of an unmeasured confounder. Theoretical values: $\pi_1 = 0.701$, $\pi_0 = 0.589$, $\theta = 0.492$.

minimum of 0.456 for GC with the common set given a large sample size for the ATT estimation (see Online Supporting Information (OSI) for complete results). The results were similar under the null hypothesis (see OSI).

Variance. For all methods, the outcome set led to the lowest MSE, followed closely by the common set. G-computation led to the lowest MSE and FM to the highest. In ATT, IPTW had lower MSE than TMLE. Note that the VEB was particularly high for FM in all ATE scenarios with a minimum of −17.5% (n = 500 with the outcome set). For the ATT, FM also had a higher VEB than other methods, apart from TMLE with the treatment or entire sets in sample sizes of fewer than 2000 subjects. In the presence of an unmeasured confounder, the MSE increased in all scenarios in agreement with the increase in bias. The VEBs did not change notably with an unmeasured confounder.

Coverage and error rates. G-computation produced coverage rates close to 95%, except for ATE in a small sample size leading to an anti-conservative 95% CIs with a minimum of 91.7% with the entire set under the null hypothesis. Anti-conservatives 95% CIs were also produced by FM in all scenarios, and by TMLE given a small sample size. Conversely, conservative 95% CIs were obtained when using TMLE for the ATT with the entire or the treatment sets, and when using IPTW for ATT or ATE with the outcome or the common sets.

Lending confidence to these results, the type I error was close to 5% for GC in all scenarios and may vary for other methods. The power was more impacted by the choice of the covariate set. The outcome set led to the highest power for GC.

Applications

We illustrated our findings by using two real data sets. First, we compared the efficiency of two treatments, *i.e.*, Natalizumab and Fingolimod, sharing the same indication for active relapsing-remitting multiple sclerosis. Physicians preferentially use Natalizumab in practice for more active disease, indicating possible confounders. Given the absence of a clinical trial with a direct comparison of their efficacy, Barbin *et al.*[53] recently conducted an observational study. We reused their data. Second, we sought to study barbiturates that can lead to a reduction of the patient functional status. Indeed, barbiturates are suggested in Intensive Care Units (ICU) for the treatment of refractory intracranial pressure increases. However, the use of barbiturates is associated with haemodynamic repercussions that can lead to brain ischaemia and immunodeficiency, which may contribute to the occurrence of infection. These applications were conducted in accordance with the French law relative to clinical noninterventional research. According to the French law on Bioethics (July 29, 1994; August 6, 2004; and July 7, 2011, Public Health Code), the patients' written informed consent was collected. Moreover, data confidentiality was ensured in accordance with the recommendations of the French commission for data protection (Commission Nationale Informatique et Liberté, CNIL decisions DR-2014-558 and DR-2013-047 for the first and the second application, respectively).

To define the four sets of covariates, we asked experts (D.L. for multiple sclerosis and M.L. for ICU) which covariates were causes of the treatment allocation and which were causes of the outcome, as proposed by VanderWeele and Shpitser[33]. We checked the positivity assumption and the covariate balance (see OSI). We applied B-spline transformations for continuous variables when the log-linearity assumption did not hold.

Natalizumab versus Fingolimod to prevent relapse in multiple sclerosis patients. The outcome was at least one relapse within one year of treatment initiation. Six hundred and twenty-nine patients from the French national cohort OFSEP were included (www.ofsep.org). The first part of Table 3 presents a description of their baseline characteristics.

All included patients could have received either treatment. Therefore, we sought to estimate the ATE. The first part of Table 4 presents the results according to the different possible methods and covariate sets. The GC, IPTW and TMLE yield similar results regardless of the covariate sets considered. Thus, Fingolimod exhibits lower efficacy than Natalizumab with an OR [95% CI] ranging from 1.50 [1.02; 2.21] for IPTW with the entire set to 1.55 [1.06; 2.28] for GC with the common set. When using FM, the OR ranged from 1.73 [1.19; 2.51] with the outcome set to 1.78 [1.23; 2.56] with the common set. Note that, unlike IPTW, FM does not to balance all covariates in the outcome set with standardised differences higher than 10%.

Overall, the confounder-adjusted proportion of patients with at least one relapse within the first year of treatment was lower in the hypothetical world where all patients received Natalizumab (approximately 20% and

n	method	set	mean bias				log OR				
			π_0	π_1	$\Delta\pi$	log OR	MSE	MSE*	VEB (%)	coverage (%)	power (%)
100	GC	outcome	−0.001	−0.002	−0.001	−0.003	0.404	0.634	−7.3	93.2	24.7
		treatment	−0.002	−0.001	0.000	0.004	0.477	0.727	−9.5	92.4	19.9
		common	−0.001	−0.002	−0.001	−0.002	0.434	0.650	−6.6	93.5	22.1
		entire	−0.002	−0.001	0.001	0.003	0.450	0.714	−11.4	91.8	22.6
	IPTW	outcome	−0.003	−0.001	0.001	0.011	0.464	0.646	12.1	97.4	12.1
		treatment	−0.006	0.002	0.008	0.046	0.633	0.769	−7.6	93.8	16.7
		common	−0.002	−0.001	0.001	0.010	0.480	0.657	6.3	96.3	13.5
		entire	−0.006	0.003	0.009	0.053	0.647	0.773	−7.2	94.7	16.4
	TMLE	outcome	−0.001	−0.002	−0.000	0.003	0.438	0.642	−14.3	89.5	26.9
		treatment	−0.004	0.002	0.006	0.039	0.572	0.757	−24.9	84.3	27.5
		common	−0.001	−0.002	−0.001	0.002	0.469	0.657	−10.7	90.9	21.2
		entire	−0.005	0.003	0.007	0.043	0.544	0.748	−30.7	80.9	34.3
	FM	outcome	−0.005	0.002	0.006	0.039	0.549	0.710	−24.3	87.1	28.5
		treatment	−0.009	0.005	0.014	0.082	0.677	0.832	−37.7	78.0	35.1
		common	−0.005	0.001	0.006	0.038	0.563	0.713	−26.3	85.8	29.1
		entire	−0.007	0.006	0.014	0.082	0.674	0.830	−37.3	78.1	34.8
300	GC	outcome	−0.000	−0.000	0.000	0.001	0.221	0.532	−1.9	94.5	59.8
		treatment	−0.000	−0.000	0.000	0.001	0.259	0.608	−2.8	94.3	47.4
		common	−0.000	−0.000	0.000	0.001	0.237	0.539	−1.2	94.8	53.5
		entire	−0.000	−0.000	0.000	0.001	0.241	0.600	−3.4	94.0	53.0
	IPTW	outcome	−0.001	−0.000	0.001	0.006	0.239	0.533	20.2	98.0	34.7
		treatment	−0.002	0.000	0.003	0.014	0.330	0.615	4.6	96.0	29.5
		common	−0.001	−0.000	0.001	0.006	0.252	0.541	13.3	97.4	36.5
		entire	−0.002	0.000	0.002	0.013	0.326	0.607	7.9	96.6	28.5
	TMLE	outcome	−0.000	−0.001	−0.000	0.000	0.233	0.532	−3.0	93.9	54.2
		treatment	−0.001	0.000	0.002	0.009	0.310	0.612	−10.4	90.6	40.2
		common	−0.001	−0.001	0.000	0.001	0.249	0.540	−1.5	94.6	48.1
		entire	−0.001	0.000	0.001	0.008	0.290	0.603	−13.2	89.6	46.1
	FM	outcome	−0.002	0.000	0.002	0.010	0.294	0.552	−20.2	88.7	51.6
		treatment	−0.003	0.003	0.006	0.032	0.389	0.652	−39.3	77.0	53.3
		common	−0.001	−0.000	0.001	0.008	0.315	0.588	−25.5	86.2	51.3
		entire	−0.003	0.003	0.006	0.032	0.377	0.644	−37.4	77.8	52.2
500	GC	outcome	−0.000	0.000	0.001	0.003	0.168	0.501	−0.4	94.8	81.1
		treatment	−0.000	0.000	0.001	0.002	0.198	0.573	−1.0	94.8	69.0
		common	−0.000	0.000	0.000	0.002	0.183	0.505	−0.7	94.9	75.0
		entire	−0.000	0.000	0.001	0.004	0.183	0.569	−1.0	94.8	75.3
	IPTW	outcome	−0.001	0.000	0.001	0.005	0.180	0.501	22.2	98.3	58.5
		treatment	−0.001	0.001	0.001	0.007	0.248	0.573	8.1	96.5	42.3
		common	−0.001	0.000	0.001	0.005	0.193	0.505	13.8	97.3	58.6
		entire	−0.001	0.000	0.001	0.006	0.239	0.569	13.1	97.2	41.3
	TMLE	outcome	−0.000	0.000	0.000	0.002	0.177	0.501	−0.8	94.7	76.8
		treatment	−0.000	0.000	0.000	0.003	0.234	0.571	−5.9	92.7	56.1
		common	−0.000	0.000	0.000	0.002	0.190	0.505	−0.5	94.7	69.7
		entire	−0.000	0.000	0.000	0.003	0.218	0.566	−7.5	91.8	63.1
	FM	outcome	−0.001	0.000	0.001	0.005	0.219	0.518	−17.5	89.8	70.1
		treatment	−0.002	0.002	0.003	0.018	0.302	0.598	−39.8	76.2	65.5
		common	−0.001	−0.000	0.001	0.005	0.266	0.555	−31.8	82.3	66.4
		entire	−0.002	0.002	0.004	0.019	0.289	0.592	−37.1	78.3	66.2
2000	GC	outcome	−0.000	−0.000	−0.000	−0.001	0.085	0.482	−0.6	94.6	100.0
		treatment	0.000	−0.001	−0.001	−0.003	0.099	0.550	−0.6	94.7	99.8
		common	0.000	−0.001	−0.001	−0.003	0.092	0.483	−0.8	94.7	99.9
		entire	−0.000	−0.000	−0.000	−0.001	0.091	0.550	−0.6	94.7	99.9
	IPTW	outcome	−0.000	−0.000	0.000	0.002	0.090	0.482	21.2	98.2	99.8
		treatment	0.000	−0.001	−0.001	−0.002	0.122	0.547	9.3	96.7	95.1
		common	−0.000	−0.000	0.000	0.001	0.096	0.483	13.5	97.3	99.7
		entire	0.000	−0.000	−0.001	−0.002	0.117	0.546	14.3	97.5	95.6

Continued

n	method	set	mean bias				log OR				
			π_0	π_1	$\Delta\pi$	log OR	MSE	MSE*	VEB (%)	coverage (%)	power (%)
2000	TMLE	outcome	−0.000	−0.000	−0.000	−0.001	0.088	0.482	−0.6	94.8	100.0
		treatment	0.000	−0.001	−0.001	−0.003	0.116	0.545	−2.2	94.4	98.7
		common	0.000	−0.000	−0.001	−0.002	0.095	0.483	−0.3	94.8	99.9
		entire	0.000	−0.000	−0.001	−0.002	0.108	0.544	−2.6	94.1	99.4
	FM	outcome	−0.000	−0.000	−0.000	0.000	0.129	0.497	−29.9	82.9	99.0
		treatment	−0.001	−0.000	0.000	0.003	0.169	0.569	−46.6	70.6	96.2
		common	0.000	−0.000	−0.001	−0.001	0.205	0.534	−55.9	61.1	92.7
		entire	−0.000	−0.000	0.000	0.002	0.145	0.549	−37.7	77.9	98.2

Table 2. Simulation results comparing the ATE estimation under the alternative hypothesis. *MSE in the presence of an unmeasured confounder. Theoretical values: $\pi_1 = 0.557$, $\pi_0 = 0.441$, $\theta = 0.466$.

varying slightly depending on method and set of covariates) than one in which all patients received Fingolimod (approximately 28%). This difference of approximately 8% is clinically meaningful and suggests the superiority of Natalizumab over Fingolimod to prevent relapses at one year. This result was concordant with the recent clinical literature[53,54].

Impact of barbiturates in the ICU on the functional status at three months. We define an unfavourable functional outcome by a 3-month Glasgow Outcome Scale (GOS) lower than or equal to 3. We used the data from the French observational cohort AtlanREA (www.atlanrea.org) to estimate the ATT of barbiturates because physicians recommended these drugs to a minority of severe patients. The second part of Table 3 presents the baseline characteristics of the 252 included patients.

The second part of Table 4 presents the results according to the different possible methods and covariate sets. G-computation and TMLE lead to the conclusion of a significant negative effect of barbiturates regardless of the covariate set considered with an OR [95% CI] ranging from 0.43 [0.25; 0.76] for GC with the common set to 0.51 [0.29; 0.90] for TMLE with the entire set. By contrast, the results were discordant when using different covariate sets for IPTW and FM. We report, for instance, OR estimates obtained by FM ranging from 1.520 with the outcome set to 2.300 with the common set. In line with the simulation study, the estimated standard errors were higher for these methods (0.294 and 0.293 for GC and TMLE when the outcome set was considered, respectively) leading to lower power. Note also that standardised differences were higher than 10% for the IPTW with the entire set (see OSI) and for FM with the outcome, the treatment and the entire sets.

Depending on the methods and sets of covariates included, we estimated that from 18% to 20% of patients treated with barbiturates had an unfavourable GOS at three months. If these patients had not received barbiturates, the methods estimate that from 30% to 35% would have had an unfavourable GOS at three months. For the patients, this difference is meaningful but full clinical relevance depends also on the effect of barbiturates on other clinically relevant outcomes, such as death or ventilator-associated pneumonia. However, the results obtained by GC or TMLE differ with those obtained by Majdan *et al.*[55], who did not find any significant effect of barbiturates on the GOS at six months. Two main methodological reasons can explain this difference: the GOS was at six months rather than three months post-initiation, and the authors used multivariate logistic regression leading to a different estimand.

Discussion

The aim of this study was to better understand the different sets of covariates to consider when estimating the marginal causal effect.

The results of our simulation study, limited to the studied scenarios, highlight that the use of the outcome set was associated with the lower bias and variance, principally when associated with GC, for both ATE and ATT. As expected, an unmeasured confounder led to increased bias, regardless of method employed. Although we do not report an impact on the variance, the effect's over- or under-estimation leads to the corresponding over- or under-estimation of power and compromises the validity of the causal inference.

The performance of FM is lower than that of the other studied methods, especially for the variance. Our results were in line with King and Nielsen[56], who argued for halting the use of PS matching for many reasons such as covariate imbalance, inefficiency, model dependence and bias. Nonetheless, Colson *et al.*[17] found slightly higher MSE for GC than FM. Their more simplistic scenario, with only two simulated confounders leading to little covariate imbalance, could explain the difference with our results. Moreover, is unclear whether they accounted for the matched nature of the data, as recommended by Austin and Stuart[16] or Gayat *et al.*[50].

While DRE offers protection against model misspecification[23,34,36], our simulation study resulted in the finding that GC was more robust to the choice of the covariate set than the other methods, TMLE included. This result was particularly important when the treatment set was taken into account, which fits with the results of Kang and Schafer[35]: when both the PS and the Q-model were misspecified, DRE had lower performance than GC. Furthermore, GC was associated with lower variance than DRE in several simulation studies[13,17,35], which accords with our results.

The first application to multiple sclerosis (ATE) illustrated similar results between the studied methods. In contrast, the second application (ATT) to severe trauma or brain-damaged patients showed different results between the methods. In agreement with simulations, the estimations obtained with GC or TMLE were similar

A - Multiple sclerosis	Overall (n = 629)		First line treatment					Relapse at 1 year				
			Ntz (n = 326)		Fng (n = 303)		p	No (n = 478)		Yes (n = 151)		p
Patient age, years (mean, sd)	37.0	9.6	36.8	9.9	37.2	9.2	0.6505	37.1	9.7	36.6	9.2	0.5849
Female patient (n, %)	479.0	76.2	254.0	77.9	225.0	74.3	0.2822	367.0	76.8	112.0	74.2	0.5124
Disease duration, years (mean, sd)	8.5	6.4	8.0	6.1	9.0	6.8	0.0505	8.6	6.6	8.2	6.0	0.4809
At least one relapse (n, %)	526.0	83.6	293.0	89.9	233.0	76.9	<0.0001	391.0	81.8	135.0	89.4	0.0277
Gd-enhancing lesion on MRI (n, %)	311.0	49.4	185.0	56.7	126.0	41.6	0.0001	240.0	50.2	71.0	47.0	0.4944
EDSS score >3 (n, %)	288.0	45.8	166.0	50.9	122.0	40.3	0.0074	212.0	44.4	76.0	50.3	0.1986
Previous immunomodulatory treatment (n, %)	556.0	88.4	293.0	89.9	263.0	86.8	0.2284	424.0	88.7	132.0	87.4	0.6672
B – ICU	**Overall (n = 252)**		**Barbiturates treatment**					**Favourable GOS at 3 months**				
			No (n = 178)		Yes (n = 74)		p	No (n = 180)		Yes (n = 72)		p
Patient age, years (mean, sd)	47.4	17.4	48.7	17.9	44.1	15.7	0.0565	50.8	16.4	38.7	16.9	<0.0001
Female patient (n, %)	89.0	35.3	58.0	32.6	31.0	41.9	0.1592	68.0	37.8	21.0	29.2	0.1963
Diabetes (n, %)	17.0	6.7	15.0	8.4	2.0	2.7	0.0989	15.0	8.3	2.0	2.8	0.1122
Nosological entity: Severe trauma (n, %)	124.0	49.2	95.0	53.4	29.0	39.2	0.0403	77.0	42.8	47.0	65.3	0.0012
SAP ≤90 mmHg before admission (n, %)	56.0	22.2	36.0	20.2	20.0	27.0	0.2368	46.0	25.6	10.0	13.9	0.0442
Evacuation of subdural or extradural hematoma (n, %)	41.0	16.3	33.0	18.5	8.0	10.8	0.1301	27.0	15.0	14.0	19.4	0.3878
External ventricular drain (n, %)	64.0	25.4	39.0	21.9	25.0	33.8	0.0486	48.0	26.7	16.0	22.2	0.4640
Evacuation of cerebral hematoma or lobectomy (n, %)	42.0	16.7	28.0	15.7	14.0	18.9	0.5362	34.0	18.9	8.0	11.1	0.1345
Decompressive craniectomy (n, %)	27.0	10.7	15.0	8.4	12.0	16.2	0.0686	21.0	11.7	6.0	8.3	0.4396
Blood transfusion before admission (n, %)	34.0	13.5	25.0	14.0	9.0	12.2	0.6903	26.0	14.4	8.0	11.1	0.4841
Pneumonia before increased ICP (n, %)	29.0	11.5	16.0	9.0	13.0	17.6	0.0519	19.0	10.6	10.0	13.9	0.4538
Osmotherapy (n, %)	112.0	44.4	75.0	42.1	37.0	50.0	0.2525	89.0	49.4	23.0	31.9	0.0115
GCS score ≥8	62.0	24.6	39.0	21.9	23.0	31.1	0.1237	37.0	20.6	25.0	34.7	0.0183
Hemoglobin, g/dL (mean, sd)	11.8	2.3	11.7	2.2	12.1	2.5	0.1824	11.8	2.4	11.9	1.9	0.7373
Platelets, counts/mm^3 (mean, sd)	206.7	78.0	207.4	79.7	205.1	74.2	0.8312	209.0	83.8	200.9	61.1	0.4589
Serum creatinine, mmol/L (mean, sd)	71.1	29.3	71.1	27.6	71.1	33.3	0.9853	72.4	32.6	67.9	18.7	0.2732
Arterial pH (mean, sd)	7.3	0.1	7.3	0.1	7.3	0.1	0.0978	7.3	0.1	7.3	0.1	0.6317
Serum proteins, g/L (mean, sd)	58.2	10.4	57.7	10.6	59.6	9.7	0.1662	58.0	10.7	58.8	9.7	0.5963
Serum urea, mmol/L (mean, sd)	5.0	2.5	5.2	2.7	4.7	1.8	0.1827	5.2	2.3	4.5	2.9	0.0505
PaO_2/FiO_2 ratio (mean, sd)	302.7	174.0	292.7	154.7	326.6	212.9	0.1595	282.1	172.4	354.2	168.4	0.0028
SAPS II score (mean, sd)	47.6	11.4	47.6	10.7	47.6	12.9	0.9847	49.9	10.8	41.8	10.7	<0.0001

Table 3. Baseline characteristics of patients of the two studied cohorts. Ntz: Natalizumab, Fng: Fingolimod, Gd: Gadolinium, MRI: Magnetic Resonance Imaging, EDSS: Expanded Disability Status Scale, SAP: Systolic Arterial Pressure, ICP: Intra-Cranial Pressure, GCS: Glasgow Coma Scale, PaO_2/FiO_2: arterial partial Pressure of Oxygen/Fraction of Inspired Oxygen, SAPS II: Simplified Acute Physiology Score II.

in terms of logOR estimation and variance regardless of the covariate set considered. Estimations obtained with IPTW or FM were highly variable, depending on the covariate set employed: some indicated a negative impact of barbiturates and others did not. These results also tended to demonstrate that GC or TMLE had the highest statistical power. Variances obtained by parametric simulations or by bootstrap resampling were similar (results not displayed).

One can, therefore, question the relative predominance of the PS-based approach compared to GC, although there are several potential explanations. First, there appears to be a pre-conceived notion according to which multivariable non-linear regression cannot be used to estimate marginal absolute and relative effects[57]. Indeed, under logistic regression, the mean sample probability of an event is different from the event probability of a subject with the mean sample characteristics. Second, while there is an explicit variance formula for the IPTW[58], the equivalent is missing for the GC. The variance must be obtained by bootstrapping, simulation or the delta method. Third, several didactic tutorials on PS-based methods can be found, for instance[59–61].

We still believe that PS-based methods may have value when multivariate modelling is complex, for instance, for multi-state models[62]. In future research, it would be interesting to examine whether the use of potentially better settings would provide equivalent results, such as the Williamson estimator for IPTW[58], the Abadie-Imbens estimator for PS matching[63], or bounded the estimation of TMLE, which can also be updated several times[36]. We also emphasise that we did not investigate these methods when the positivity assumption does not hold. Several authors have studied this problem[13,25,35,36,64]. G-computation was less biased than IPTW or DRE except in Porter *et al.*[36], where the violation of the positivity assumption was also associated with model misspecifications. The robustness of GC to non-positivity could be due to a correct extrapolation into the missing sub-population, which is not feasible with PS[1]. Other perspectives of this work are to extend the problem to i) time-to-event, continuous or multinomial outcomes and ii) multinomial treatment. However, implementing GC using continuous treatment raises many important considerations concerning the research question and resulting inference[64].

application	method	set	$\hat{\pi}_0$	$\hat{\pi}_1$	$\hat{\theta}$	SE	95% CI
A - Multiple sclerosis	GC	outcome	20.3	28.2	0.432	0.189	[0.062, 0.802]
		treatment*	20.3	28.3	0.436	0.195	[0.054, 0.819]
		common*	20.3	28.3	0.436	0.195	[0.054, 0.819]
		entire	20.3	28.2	0.431	0.191	[0.056, 0.806]
	IPTW	outcome	21.2	28.8	0.406	0.195	[0.023, 0.789]
		treatment*	20.3	28.2	0.433	0.191	[0.059, 0.808]
		common*	20.3	28.2	0.433	0.191	[0.059, 0.808]
		entire	21.3	28.9	0.406	0.196	[0.022, 0.791]
	TMLE	outcome	21.2	28.8	0.407	0.195	[0.025, 0.790]
		treatment*	20.3	28.2	0.433	0.190	[0.061, 0.806]
		common*	20.3	28.2	0.433	0.190	[0.061, 0.806]
		entire	21.1	28.9	0.410	0.196	[0.026, 0.794]
	FM	outcome	19.1	29.0	0.549	0.189	[0.178, 0.921]
		treatment*	19.9	30.6	0.575	0.187	[0.210, 0.941]
		common*	19.9	30.6	0.575	0.187	[0.210, 0.941]
		entire	21.1	31.9	0.561	0.183	[0.201, 0.920]
B - ICU	GC	outcome	66.3	81.1	0.778	0.294	[0.201, 1.354]
		treatment	65.3	81.1	0.824	0.298	[0.240, 1.407]
		common	65.0	81.1	0.836	0.289	[0.270, 1.402]
		entire	66.5	81.1	0.769	0.295	[0.191, 1.347]
	IPTW	outcome	31.0	81.1	0.656	0.356	[−0.042, 1.354]
		treatment	68.2	81.1	0.693	0.355	[−0.002, 1.388]
		common	67.4	81.1	0.729	0.353	[0.038, 1.421]
		entire	69.2	81.1	0.645	0.362	[−0.064, 1.354]
	TMLE	outcome	66.2	79.6	0.692	0.293	[0.118, 1.266]
		treatment	65.4	80.2	0.758	0.288	[0.194, 1.322]
		common	64.8	79.9	0.769	0.298	[0.185, 1.354]
		entire	66.4	79.4	0.668	0.285	[0.109, 1.228]
	FM	outcome	73.8	81.1	0.419	0.342	[−0.252, 1.090]
		treatment	67.2	81.1	0.739	0.337	[0.078, 1.399]
		common	65.1	81.1	0.831	0.336	[0.173, 1.490]
		entire	66.2	81.1	0.782	0.336	[0.123, 1.442]

Table 4. Results of the two applications. *Treatment and common sets contain same covariates. π_0: Percentage of event in the Natalizumab (or control) group, π_1: Percentage of event in the Fingolimod (or Barbiturates) group, SE: standard error.

To facilitate its use in practice, we have implemented the estimation of both ATE and ATT, and their 95% CI, from a logistic model in the existing R package entitled *RISCA* (available at cran.r-project.org/web/packages/RISCA). We provide an example of R code in the appendix. Note that the package did not consider the inflation of the type I error rate due to the modelling steps of the Q-model. Users also have to consider novel strategies for post-model selection inference.

In the applications, we classified covariates into sets based on experts knowledge[33]. However, several statistical methods can be useful when no clinical knowledge is available. Heinze *et al.*[65] proposed a review of the most used, while Witte and Didelez[66] reviewed strategies specific to causal inference. Alternatively, data-adaptive methods have recently been developed, such as the outcome-adaptive LASSO[67] to select covariates associated with both the outcome and the treatment allocation. Nevertheless, according to our results, it may be preferable to focus on constructing the best outcome model based on the outcome set. For instance, the consideration of a super learner[68,69], merging models and modelling machine learning algorithms may represent an exciting perspective[70].

Finally, we emphasise that the conclusions from our simulation study cannot be generalised to all situations. They are consistent with the current literature on causal inference, but theoretical arguments are missing for generalisation. Notably, our results must be considered in situations where both the PS and the Q-model are correctly specified and where positivity holds.

To conclude, we demonstrate in a simulation study that adjusting for all the covariates causing the outcome improves the estimation of the marginal causal effect (ATE or ATT) of a binary treatment in a binary outcome. Considering only the covariates that are a common cause of both the outcome and the treatment is possible when the number of potential confounders is large. The strategy consisting of considering all available covariates, *i.e.*, no selection, did not decrease the bias but significantly decreased the power. Amongst the different studied methods, GC had the lowest bias and variance regardless of covariate set considered. Consequently, we recommend that the use of the GC with the outcome set, because of its highest power in all the simulated scenarios. For

instance, at least 500 individuals were necessary to achieve a power higher than 80% in ATE, with a theoretical OR at 2, and a percentage of treated subjects at 50%. In ATT, we needed larger sample size to reach a power of 80% because the estimation considers only the treated patients. With 2000 individuals, all the studied methods with the outcome set led to a bias close to zero and a statistical power superior to 95%.

Received: 9 July 2019; Accepted: 26 April 2020;
Published online: 08 June 2020

References

1. Hernan, M. A. & Robins, J. M. *Causal Inference: What if*? (Chapman & Hall/CRC, 2020).
2. Zwarenstein, M. & Treweek, S. What kind of randomized trials do we need? *Journal of Clinical Epidemiology* **62**, 461–463, https://doi.org/10.1016/j.jclinepi.2009.01.011 (2009).
3. Gayat, E. *et al*. Propensity scores in intensive care and anaesthesiology literature: a systematic review. *Intensive Care Medicine* **36**, 1993–2003, https://doi.org/10.1007/s00134-010-1991-5 (2010).
4. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55, https://doi.org/10.2307/2335942 (1983).
5. Robins, J. M., Hernán, M. A. & Brumback, B. Marginal structural models and causal inference in epidemiology. *Epidemiology* **11**, 550–560, https://doi.org/10.1097/00001648-200009000-00011 (2000).
6. Lunceford, J. K. & Davidian, M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in medicine* **23**, 2937–2960, https://doi.org/10.1002/sim.1903 (2004).
7. Austin, P. C., Grootendorst, P., Normand, S.-L. T. & Anderson, G. M. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Statistics in Medicine* **26**, 754–768, https://doi.org/10.1002/sim.2618 (2007).
8. Abdia, Y., Kulasekera, K. B., Datta, S., Boakye, M. & Kong, M. Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study. *Biometrical Journal* **59**, 967–985, https://doi.org/10.1002/bimj.201600094 (2017).
9. Grose, E. *et al*. Use of propensity score methodology in contemporary high-impact surgical literature. *Journal of the American College of Surgeons* **230**, 101–112.e2, https://doi.org/10.1016/j.jamcollsurg.2019.10.003 (2020).
10. Ali, M. S. *et al*. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. *Journal of Clinical Epidemiology* **68**, 112–121, https://doi.org/10.1016/j.jclinepi.2014.08.011 (2015).
11. Le Borgne, F., Giraudeau, B., Querard, A. H., Giral, M. & Foucher, Y. Comparisons of the performance of different statistical tests for time-to-event analysis with confounding factors: practical illustrations in kidney transplantation. *Statistics in Medicine* **35**, 1103–1116, https://doi.org/10.1002/sim.6777 (2016).
12. Hajage, D., Tubach, F., Steg, P. G., Bhatt, D. L. & De Rycke, Y. On the use of propensity scores in case of rare exposure. *BMC Medical Research Methodology* **16**, https://doi.org/10.1186/s12874-016-0135-1 (2016).
13. Lendle, S. D., Fireman, B. & van der Laan, M. J. Targeted maximum likelihood estimation in safety analysis. *Journal of Clinical Epidemiology* **66**, S91–S98, https://doi.org/10.1016/j.jclinepi.2013.02.017 (2013).
14. Austin, P. C. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. *Statistics in Medicine* **29**, 2137–2148, https://doi.org/10.1002/sim.3854 (2010).
15. Austin, P. C. & Stuart, E. A. Estimating the effect of treatment on binary outcomes using full matching on the propensity score. *Statistical Methods in Medical Research* **26**, 2505–2525, https://doi.org/10.1177/0962280215601134 (2017).
16. Austin, P. C. & Stuart, E. A. The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. *Statistical Methods in Medical Research* **26**, 1654–1670, https://doi.org/10.1177/0962280215584401 (2017).
17. Colson, K. E. *et al*. Optimizing matching and analysis combinations for estimating causal effects. *Scientific Reports* **6**, https://doi.org/10.1038/srep23222 (2016).
18. Robins, J. M. A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. *Mathematical Modelling* **7**, 1393–1512, https://doi.org/10.1016/0270-0255(86)90088-6 (1986).
19. Vansteelandt, S. & Keiding, N. Invited commentary: G-computation-lost in translation? *American Journal of Epidemiology* **173**, 739–742, https://doi.org/10.1093/aje/kwq474 (2011).
20. Snowden, J. M., Rose, S. & Mortimer, K. M. Implementation of g-computation on a simulated data set: Demonstration of a causal inference technique. *American Journal of Epidemiology* **173**, 731–738, https://doi.org/10.1093/aje/kwq472 (2011).
21. Wang, A., Nianogo, R. A. & Arah, O. A. G-computation of average treatment effects on the treated and the untreated. *BMC Medical Research Methodology* **17**, https://doi.org/10.1186/s12874-016-0282-4 (2017).
22. Imbens, G. W. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics* **86**, 4–29, https://doi.org/10.1162/003465304323023651 (2004).
23. Bang, H. & Robins, J. M. Doubly robust estimation in missing data and causal inference models. *Biometrics* **61**, 962–973, https://doi.org/10.1111/j.1541-0420.2005.00377.x (2005).
24. van der Laan, M. J. & Rubin, D. B. Targeted maximum likelihood learning. *The International Journal of Biostatistics* **2**, https://doi.org/10.2202/1557-4679.1043 (2006).
25. Neugebauer, R. & van der Laan, M. J. Why prefer double robust estimators in causal inference? *Journal of Statistical Planning and Inference* **129**, 405–426, https://doi.org/10.1016/j.jspi.2004.06.060 (2005).
26. Brookhart, M. A. *et al*. Variable Selection for Propensity Score Models. *American Journal of Epidemiology* **163**, 1149–1156, https://doi.org/10.1093/aje/kwj149 (2006).
27. Lefebvre, G., Delaney, J. A. C. & Platt, R. W. Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Statistics in Medicine* **27**, 3629–3642, https://doi.org/10.1002/sim.3200 (2008).
28. Schisterman, E. F., Cole, S. R. & Platt, R. W. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology* **20**, 488–495, https://doi.org/10.1097/EDE.0b013e3181a819a1 (2009).
29. Rotnitzky, A., Li, L. & Li, X. A note on overadjustment in inverse probability weighted estimation. *Biometrika* **97**, 997–1001, https://doi.org/10.1093/biomet/asq049 (2010).
30. Schnitzer, M. E., Lok, J. J. & Gruber, S. Variable selection for confounder control, flexible modeling and collaborative targeted minimum loss-based estimation in causal inference. *The International Journal of Biostatistics* **12**, 97–115, https://doi.org/10.1515/ijb-2015-0017 (2016).
31. Myers, J. A. *et al*. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology* **174**, 1213–1222, https://doi.org/10.1093/aje/kwr364 (2011).
32. De Luna, X., Waernbaum, I. & Richardson, T. S. Covariate selection for the nonparametric estimation of an average treatment effect. *Biometrika* **98**, 861–875, https://doi.org/10.1093/biomet/asr041 (2011).
33. VanderWeele, T. J. & Shpitser, I. A new criterion for confounder selection. *Biometrics* **67**, 1406–1413, https://doi.org/10.1111/j.1541-0420.2011.01619.x (2011).

34. Schuler, M. S. & Rose, S. Targeted maximum likelihood estimation for causal inference in observational studies. *American Journal of Epidemiology* **185**, 65–73, https://doi.org/10.1093/aje/kww165 (2017).
35. Kang, J. D. Y. & Schafer, J. L. Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22**, 523–539, https://doi.org/10.1214/07-STS227 (2007).
36. Porter, K. E., Gruber, S., van der Laan, M. J. & Sekhon, J. S. The relative performance of targeted maximum likelihood estimators. *The International Journal of Biostatistics* **7**, https://doi.org/10.2202/1557-4679.1308 (2011).
37. Moher, D. *et al.* Consort 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* **340**, c869, https://doi.org/10.1136/bmj.c869 (2010).
38. Greenland, S., Robins, J. M. & Pearl, J. Confounding and collapsibility in causal inference. *Statistical Science* **14**, 29–46, https://doi.org/10.1214/ss/1009211805 (1999).
39. Aalen, O. O., Cook, R. J. & Røysland, K. Does cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Analysis* **21**, 579–593, https://doi.org/10.1007/s10985-015-9335-y (2015).
40. Pearl, J., Glymour, M. & Jewell, N. P. *Causal Inference in Statistics: A Primer* (John Wiley & Sons, 2016).
41. Xu, S. *et al*. Use of Stabilized Inverse Propensity Scores as Weights to Directly Estimate Relative Risk and Its Confidence Intervals. *Value in Health* **13**, 273–277, https://doi.org/10.1111/j.1524-4733.2009.00671.x (2010).
42. Morgan, S. L. & Todd, J. J. A diagnostic routine for the detection of consequential heterogeneity of causal effects. *Sociological Methodology* **38**, 231–282, https://doi.org/10.1111/j.1467-9531.2008.00204.x (2008).
43. Zeileis, A. Object-oriented computation of sandwich estimators. *Journal of Statistical Software* **16**, 1–16, https://doi.org/10.18637/jss.v016.i09 (2006).
44. Austin, P. C. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments: Propensity scores and survival analysis. *Statistics in Medicine* **33**, 1242–1258, https://doi.org/10.1002/sim.5984 (2014).
45. Ho, D., Imai, K., King, G. & Stuart, E. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software* **42**, 1–28, https://doi.org/10.18637/jss.v042.i08 (2011).
46. Aalen, O. O., Farewell, V. T., De Angelis, D., Day, N. E. & Gill, O. N. A markov model for hiv disease progression including the effect of hiv diagnosis and treatment: application to aids prediction in england and wales. *Statistics in Medicine* **16**, 2191–2210, https://doi.org/10.1002/(sici)1097-0258(19971015)16:19<2191::aid-sim645>3.0.co;2-5 (1997).
47. van der Laan, M. J. & Rose, S. *Targeted learning: causal inference for observational and experimental data*. Springer series in statistics (Springer, 2011).
48. Hampel, F. R. The influence curve and its role in robust estimation. *Journal of the American Statistical Association* **69**, 383–393, https://doi.org/10.2307/2285666 (1974).
49. Luque-Fernandez, M. A., Schomaker, M., Rachet, B. & Schnitzer, M. E. Targeted maximum likelihood estimation for a binary treatment: A tutorial. *Statistics in Medicine* **37**, 2530–2546, https://doi.org/10.1002/sim.7628 (2018).
50. Gayat, E., Resche-Rigon, M., Mary, J.-Y. & Porcher, R. Propensity score applied to survival data analysis through proportional hazards models: a Monte Carlo study. *Pharmaceutical Statistics* **11**, 222–229, https://doi.org/10.1002/pst.537 (2012).
51. Morris, T. P., White, I. R. & Crowther, M. J. Using simulation studies to evaluate statistical methods. *Statistics in Medicine* **38**, 2074–2102, https://doi.org/10.1002/sim.8086 (2019).
52. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2014).
53. Barbin, L. *et al*. Comparative efficacy of fingolimod vs natalizumab. *Neurology* **86**, 771–778, https://doi.org/10.1212/WNL.0000000000002395 (2016).
54. Kalincik, T. *et al*. Switch to natalizumab versus fingolimod in active relapsing-remitting multiple sclerosis. *Annals of Neurology* **77**, 425–435, https://doi.org/10.1002/ana.24339 (2015).
55. Majdan, M. *et al*. Barbiturates Use and Its Effects in Patients with Severe Traumatic Brain Injury in Five European Countries. *Journal of Neurotrauma* **30**, 23–29, https://doi.org/10.1089/neu.2012.2554 (2012).
56. King, G. & Nielsen, R. Why propensity scores should not be used for matching. *Political Analysis* **27**, 435–454, https://doi.org/10.1017/pan.2019.11 (2019).
57. Nieto, F. J. & Coresh, J. Adjusting survival curves for confounders: a review and a new method. *American Journal of Epidemiology* **143**, 1059–1068, https://doi.org/10.1093/oxfordjournals.aje.a008670 (1996).
58. Williamson, E. J., Forbes, A. & White, I. R. Variance reduction in randomised trials by inverse probability weighting using the propensity score. *Statistics in Medicine* **33**, 721–737, https://doi.org/10.1002/sim.5991 (2014).
59. Williamson, E. J., Morley, R., Lucas, A. & Carpenter, J. Propensity scores: From naïve enthusiasm to intuitive understanding. *Statistical Methods in Medical Research* **21**, 273–293, https://doi.org/10.1177/0962280210394483 (2012).
60. Austin, P. C. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research* **46**, 399–424, https://doi.org/10.1080/00273171.2011.568786 (2011).
61. Haukoos, J. S. & Lewis, R. J. The propensity score. *JAMA* **314**, 1637–1638, https://doi.org/10.1001/jama.2015.13480 (2015).
62. Gillaizeau, F. *et al*. Inverse probability weighting to control confounding in an illness-death model for interval-censored data. *Statistics in Medicine* **37**, 1245–1258, https://doi.org/10.1002/sim.7550 (2018).
63. Abadie, A. & Imbens, G. W. Large sample properties of matching estimators for average treatment effects. *Econometrica* **74**, 235–267, https://doi.org/10.1111/j.1468-0262.2006.00655.x (2006).
64. Moore, K. L., Neugebauer, R., van der Laan, M. J. & Tager, I. B. Causal inference in epidemiological studies with strong confounding. *Statistics in Medicine* **31**, 1380–1404, https://doi.org/10.1002/sim.4469 (2012).
65. Heinze, G., Wallisch, C. & Dunkler, D. Variable selection - a review and recommendations for the practicing statistician. *Biometrical Journal* **60**, 431–449, https://doi.org/10.1002/bimj.201700067 (2018).
66. Witte, J. & Didelez, V. Covariate selection strategies for causal inference: Classification and comparison. *Biometrical Journal* **61**, 1270–1289, https://doi.org/10.1002/bimj.201700294 (2019).
67. Shortreed, S. M. & Ertefaie, A. Outcome-adaptive lasso: variable selection for causal inference. *Biometrics* **73**, 1111–1122, https://doi.org/10.1111/biom.12679 (2017).
68. van der Laan, M. J., Polley, E. C. & Hubbard, A. E. Super learner. *Statistical Applications in Genetics and Molecular Biology* **6**, Article 25, https://doi.org/10.2202/1544-6115.1309 (2007).
69. Naimi, A. I. & Balzer, L. B. Stacked generalization: An introduction to super learning. *European journal of epidemiology* **33**, 459–464, https://doi.org/10.1007/s10654-018-0390-z (2018).
70. Pirracchio, R. & Carone, M. The balance super learner: A robust adaptation of the super learner to improve estimation of the average treatment effect in the treated based on propensity score matching. *Statistical Methods in Medical Research* **27**, 2504–2518, https://doi.org/10.1177/0962280216682055 (2018).

Acknowledgements

The authors would like to thank the members of AtlanREA and OFSEP Groups for their involvement in the study, the physicians who helped recruit patients and all patients who participated in this study. We also thank the clinical research associates who participated in the data collection. The analysis and interpretation of these data are the responsibility of the authors. This work was partially supported by a public grant overseen by the

French National Research Agency (ANR) to create the Common Laboratory RISCA (Research in Informatic and Statistic for Cohort Analyses, www.labcom-risca.com, reference: ANR-16-LCV1-0003-01). The funder had no role in study design; analysis, and interpretation of data; writing the report; and the decision to submit the report for publication.

Author contributions

A.C. and Y.F. designed and conceptualised the study, conducted statistical analyses, analysed the data and drafted the manuscript for intellectual content, F.L.B., F.G. and B.G. designed and conceptualised the study, analysed the data and revised the manuscript for intellectual content, C.L. and C.R. analysed the data and revised the manuscript for intellectual content, L.B., D.L. and M.L. had a significant role in the acquisition of data and revised the manuscript for intellectual content. All authors approved the final version of the manuscript.

Competing interests

Dr. Y. Foucher has received speaking honoraria from Biogen and Sanofi. Pr. D. Laplaud has received Funding for travel or speaker honoraria from Biogen, Novartis, and Genzyme. He has participated in advisory boards in the past years Biogen-Idec, TEVA Pharma, Novartis, and Genzyme. The other authors declared no conflict of interest.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-65917-x.

Correspondence and requests for materials should be addressed to Y.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

Annexe E

Étude de simulation de Le Borgne *et al.* Apport de l'apprentissage automatique en g-computation

scientific reports

Check for updates

OPEN

G-computation and machine learning for estimating the causal effects of binary exposure statuses on binary outcomes

Florent Le Borgne[1,2,6], Arthur Chatton[1,2,6], Maxime Léger[1,3], Rémi Lenain[1,4] & Yohann Foucher[1,5]✉

In clinical research, there is a growing interest in the use of propensity score-based methods to estimate causal effects. G-computation is an alternative because of its high statistical power. Machine learning is also increasingly used because of its possible robustness to model misspecification. In this paper, we aimed to propose an approach that combines machine learning and G-computation when both the outcome and the exposure status are binary and is able to deal with small samples. We evaluated the performances of several methods, including penalized logistic regressions, a neural network, a support vector machine, boosted classification and regression trees, and a super learner through simulations. We proposed six different scenarios characterised by various sample sizes, numbers of covariates and relationships between covariates, exposure statuses, and outcomes. We have also illustrated the application of these methods, in which they were used to estimate the efficacy of barbiturates prescribed during the first 24 h of an episode of intracranial hypertension. In the context of GC, for estimating the individual outcome probabilities in two counterfactual worlds, we reported that the super learner tended to outperform the other approaches in terms of both bias and variance, especially for small sample sizes. The support vector machine performed well, but its mean bias was slightly higher than that of the super learner. In the investigated scenarios, G-computation associated with the super learner was a performant method for drawing causal inferences, even from small sample sizes.

Machine learning (ML) is a set of mathematical and statistical methods that computer systems use to perform tasks without specific instructions. In medical research, there is an increasing interest in these methods for prediction and, more recently, for causality[1] There is a large intersection between these fields since the first step of causal modelling consists of predicting the exposure for propensity score (PS)-based methods[2,3] or the outcome for G-computation (GC)[4,5].

Several recent methodological studies have therefore studied the potential applicability of ML for causal inference. A large number simulation-based studies have compared several ML methods to obtain PSs[1,6–10]. While the corresponding PS-based results were very encouraging, GC was compared to PS-based methods in the context of classical regression models and showed several advantages in terms of statistical power[11–14] and robustness of the estimates regardless of the set of included covariates[11]. However, simulation-based studies related to the use of ML for predicting outcomes in GC are infrequent. Austin examined the use of ensemble-based methods (bagged classification and regression trees (CART), random forests, and boosted CART (BCART)) and concluded that BCART was the highest performing algorithm[15]. He also concluded that BCART had a lower bias when it was used to impute potential outcomes than when it was used to estimate the PS for inverse probability treatment weighting.

In this paper, we studied the performances of GC in combination with different ML algorithms, including a super learner (SL), through simulations to estimate causal effects. Many of the previous studies were based on

[1]INSERM UMR 1246 - SPHERE, Nantes University, Tours University, 22 Boulevard Bénoni Goullin, 44200 Nantes, France. [2]IDBC-A2COM, Pacé, France. [3]Département D'Anesthésie Réanimation, Centre Hospitalier Universitaire D'Angers, Angers, France. [4]Lille University Hospital, Lille, France. [5]Nantes University Hospital, Nantes, France. [6]These authors contributed equally: Florent Le Borgne and Arthur Chatton. ✉email: yohann.foucher@univ-nantes.fr

large samples. Therefore, we made sure to include scenarios with small sample sizes. We limited our study to case where both the exposure and outcome were binary and to small-medium sample sizes. We also focused on ML techniques that are applicable in daily practice, i.e., with reasonable computation times on modern laptops or workstations.

Methods

G-computation. Let $Y(1)$ and $Y(0)$ be the two potential outcomes under the exposure and the non-exposure, respectively[16]. Let (Z, X) denote the random variables related to the exposure statuses of individuals ($Z = 1$ for exposed individuals and 0 otherwise) and the k covariates ($X = X_1, \ldots, X_k$) measured before exposure, respectively. The average causal effect is $ACE = E[Y(1) - Y(0)]$. It represents the mean difference between the outcomes of individuals if they had been exposed or unexposed[17].

Suppose (Y_i, Z_i, X_i) a dataset for analysis consists of n independent realisations of (Y, Z, X). The first step of GC is to fit $f(Y|Z, X)$, and this outcome model is frequently referred to as the Q-model[5]. Once estimated, the Q-model aims to predict, for each individual i ($i = 1, \ldots, n$), the two potential outcomes under each exposure status by maintaining her/his covariates X_i at the observed values and setting Z_i to 1 and 0: $\widehat{Y}_i(1) = \widehat{f}(Y|1, X_i)$ and $\widehat{Y}_i(0) = \widehat{f}(Y|0, X_i)$. The average causal effect is then estimated by $\widehat{ACE} = n^{-1}\sum_{i=1}^{n}\left[\widehat{Y}_i(1) - \widehat{Y}_i(0)\right]$.

Covariates selection. One of the main differences between prediction and causality is the selection of covariates. Knowledge of the causal relationship structure is essential for conducting causal inference[18]. This knowledge consists of excluding the mediators, colliders[19], and instrumental variables[20,21]. Note that a benefit of GC over PS-based methods is that it more effectively prevents instrumental variables, which are often included in the PS. In this context, the advantages and limits of ML algorithms have been well described[22,23]. As noted by VanderWeele and Shpitser[24], investigators can identify the causes of exposure statuses or outcomes as potential covariates.

Unfortunately, full knowledge of causal relationships is often unavailable. There is a growing literature about the best set of covariates to consider, and it recommends including all the covariates that cause the outcome[11,21,25]. The corresponding data-driven selection procedure for GC is straightforward since it corresponds to the predictors of the Q-model.

ML techniques. In contrast with PS-based methods, which consist of predicting exposure statuses, the Q-model must keep the exposure status as one of the predictors. This is not possible for several ML techniques, such as random forests, except by estimating $f(.)$ separately for the exposed and unexposed individuals. Nevertheless, this solution is not reasonable for small sample sizes (we have tested it, and the results confirm its deficient performances for $n < 1000$; data not shown). Below, we briefly describe the ML methods that we included in our simulations. For more details on these ML techniques, see McNeish for the penalized methods[26], and Bi et al. for the other methods[27]. We performed all the analyses using R version 3.6.1.

Lasso logistic regression (LLR). L1 regularisation allows for the selection of the predictors. To obtain a flexible model, we considered all the possible interactions between the exposure status Z and covariates X. Moreover, we used b-splines for the quantitative variables of the vector X. We used the *glmnet* function included in the *glmnet* package.

Elasticnet logistic regression (ELR). We used the same flexible logistic regression as previously defined, but with both the L1 and L2 regularisations (two tuning parameters).

Neural network (NN). We chose a neural network with one hidden layer, as this is probably the most common network architecture[3]. Its size constitutes the single tuning parameter. We used the *nnet* function of the *nnet* package.

Support vector machine (SVM). We chose the radial basis function kernel to flex the linear assumption. We used the *svmRadial* function of the *kernlab* package with two tuning parameters: the cost penalty of misclassification and the flexibility of the classification.

Boosted CART (BCART). This ML technique is an ensemble method, that is, a method that averages the percentages of events in the terminal nodes of several tree partitions. Four tuning parameters must be chosen: the number of trees, the highest level of covariate interactions, the learning rate, and the minimum number of observations in the terminal nodes. We used the *gbm* function included in the *gbm* package.

For the five methods listed above (LLR, ELR, NN, SVM, and BCART), we chose their respective tuning parameters by maximising the average area under the receiver operating characteristic curve (AUC) of tenfold cross-validation. We used the *caret* package with a tuning grid of length equals 20.

Super learner (SL). We included the previous ML techniques in the SL, with the exception of BCART due to the resulting computational burden. The SL consists of averaging the predictions obtained from the four approaches by using a weighted linear predictor[28]. In agreement with our previous choice, we estimated the weights by maximising the average AUC of tenfold cross-validation. We used the *SuperLearner* package.

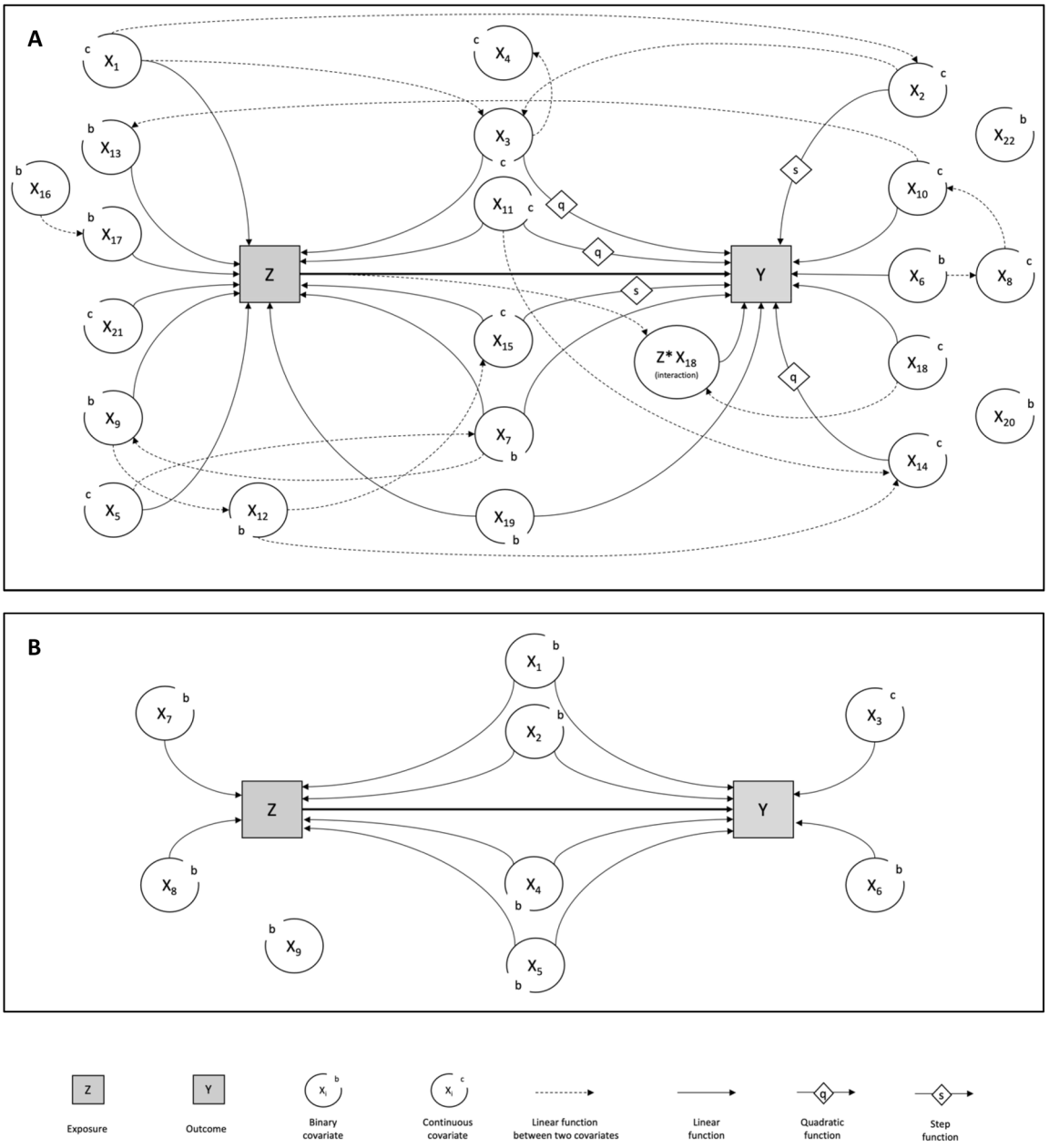


Figure 1. Directed acyclic graphs associated with the two simulated scenarios. (**A**) The realistic scenario with 22 covariates, linear and nonlinear relationships, and one interaction. (**B**) The simplistic scenario with nine covariates, linear relationships, and no interaction.

Variance estimation. By bootstrapping the entire procedure[29], one can obtain the standard error and the confidence interval of the *ACE*. Regarding the corresponding computational burden, a compromise consists of choosing the tuning parameters based on the entire sample and then using these values in the subsequent bootstrap samples[30,31]. Moreover, to consider the possible overfitting associated with such ML techniques, we performed a bootstrap cross-validation procedure. We trained the ML algorithms from the bootstrap sample, while we estimated the *ACE* from the individuals not included in the bootstrap sample. In this paper, we performed 500 iterations.

Simulation-based study

Data generation. We considered two main scenarios, as illustrated in Fig. 1 (the related models are in Supplementary Tables S1 and S2). First, we simulated the continuous and binary covariates from X_1 to X_k, allowing for dependences between the simulated covariate and those already generated. Second, we obtained Z and Y with Bernouilli distributions. The logit of the corresponding probabilities equaled the linear functions of X and (X, Z).

We choose two contrasting scenarios. We defined a realistic situation (Fig. 1A, Supplementary Table S1) with 22 correlated covariates at baseline. Nine covariates were included in the outcome model, among which one

covariate interacted with the exposure effect, two effects were step functions, three were quadratic functions, and four were linear. In contrast, we defined a simplistic situation (Fig. 1B, Supplementary Table S2) with nine independent covariates. Six covariates were included in the outcome model with linear effects and no interaction.

We simulated all the covariates X as variables measured before exposure. We did not consider mediators and colliders. As previously stated, the investigator must exclude these variables from the set of covariates. We studied different sample sizes: $n = 100$, 500, and 1000. For each scenario, we randomly generated 10,000 datasets.

Performance criteria. We computed the theoretical *ACE* by averaging the *ACE* estimations obtained from the univariate logistic models (with Z as the only explanatory variable) fitted based on datasets that were simulated as above, except that Z was generated independently of X[11,32]. We reported the following criteria (the formulae can be found in the Supplementary Materials): the mean bias (MB), the root mean square error (RMSE), the empirical standard deviation (ESD), the asymptotic standard deviation (ASD), the variance estimation bias (VEB), the empirical coverage rate of the nominal 95% confidence interval (95% CI), and the statistical power. We compared the performances of the previous ML techniques. In addition, we examined the results and compared them with those obtained by a perfectly specified LR, i.e., a LR with the same linear predictor as the one defined in the last lines of Supplementary Tables S1 and S2, in which we only estimated the corresponding regression coefficients.

Comparison of the ML techniques in terms of bias. *Overall results.* To evaluate the calibration of the ML methods for the simulated data, we added calibration plots of 10 simulated datasets for each combination of methods (LLR, ELR, NN, SVM, SL), complexity (simplistic, realistic), and sample size ($n = 100, 500, 1000$) to the Supplementary Materials (Figures S1-10). One can observe an overfitting of the ELR, SVM, and SL when $n = 100$, and this can be explained by the fact that the number of parameters was too large compared to the sample size.

We report the simulation results in Figs. 2, 3 and 4 for the realistic and simplistic scenarios (the numerical details can be found in Supplementary Tables S3 and S4). Independent of the sample size and the complexity of the relationships between the covariates and the outcome, BCART was associated with a significant level of bias, with the MB being higher than 3%.

The impact of the sample size in the realistic situation. To differentiate between the other methods, one can compare the MBs obtained when the relationships between the covariates and outcome are difficult for the analyst to manage, i.e., a realistic situation. When the learning support is small ($n = 100$), the penalized methods (ELR and LLR) and the NN resulted in unacceptable MBs higher than 4%. In contrast, the two remaining methods (SVM and SL) were associated with values lower than 1%. With large sample sizes ($n \geq 500$), the four methods performed correctly with MBs less than 3%, and the lowest MB was obtained with the SL (MB < 1% for all sample sizes). To further discriminate between the SVM and SL in this realistic situation, one can notice that the MB remained negligible for the SL regardless of the sample size, while for the SVM, the MB increased with the sample size (values between 1 and 2% when $n \geq 500$).

The impact of the sample size in the simplistic situation. Except when $n = 1000$, for which they were outperformed by the SL (MB < 1%), the penalized methods were associated with the smallest biases in the simplistic situation, with MBs less than 1% regardless of the sample size. The penalized methods were even the only methods such low values when $n = 100$. The NN was the only method with no significant variations according to the sample size (i.e., MBs between 1 and 2% for all three sample sizes).

Comparison of the ML techniques in terms of variance. *Overall results.* Regardless of the scenarios and the sample sizes used, one can observe an underestimation of the variance using BCART. Its VEB ranged from − 2 to − 56%.

The impact of the sample size in the realistic situation. To differentiate between the other methods, one can first consider the smallest sample size (n =100). The penalized approaches (LLR and ELR) resulted in the highest estimations of the variance, with ASDs close to 0.10. The SVM and NN were associated with the smallest variances, with ASDs close to 0.6 (the VEBs were − 6.4% and 8.8%, respectively). Compared with the two previous ML techniques, the SL resulted in a slightly higher ASD at 0.7, but a lower VEB at − 3.7%. For larger sample sizes ($n \geq 500$), the results in terms of variance were close for the four following approaches: LLR, ELR, SVM and SL. The NN was associated with an unacceptable overestimation of the variance (VEB = 19.0% and 31.1% for $n = 500$ and 1000, respectively).

The differences between the realistic and simplistic situations. The results were similar when the relationships between the covariates and the outcome were easier for the analyst to model (i.e., the simplistic situation). However, one can underline an exception: when $n = 100$, the NN resulted in an ASD close to those of the penalized approaches.

Synthesis of bias and variance in terms of the root mean square error and coverage. Even if BCART resulted in a critical level of bias, its RMSEs were reasonable, and this is mainly because of the previously reported underestimation of the variance. This bias associated with an underestimated variance resulted in coverage ranging from 57.2 to 82.2%, and the upper bound of this range is considerably lower than the nominal

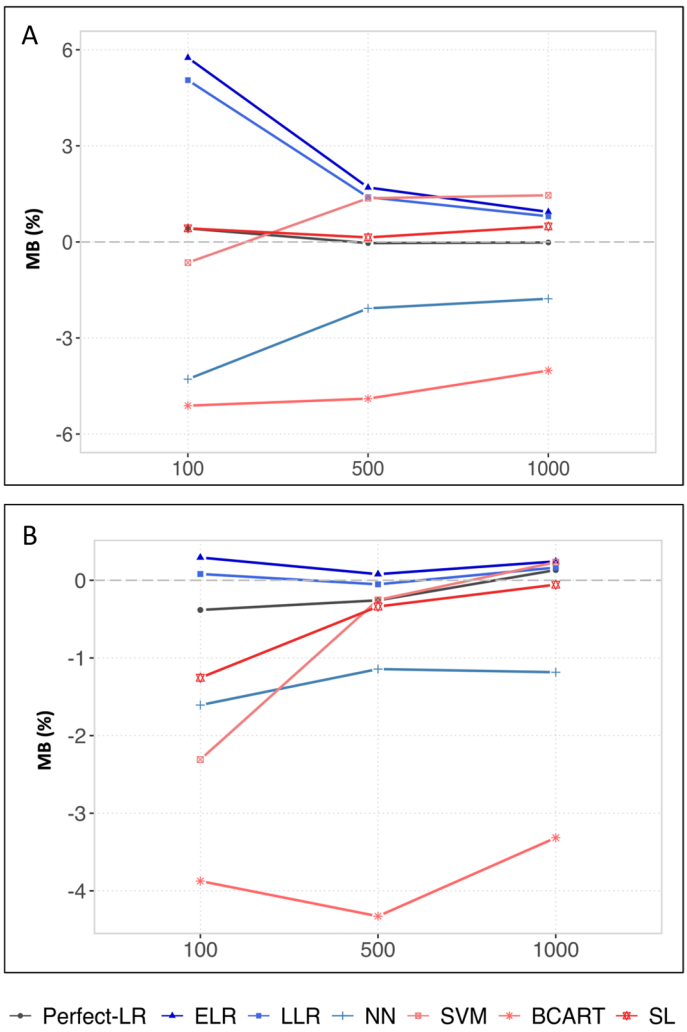


Figure 2. Mean biases (MBs) of G-computation in realistic (**A**) and simplistic (**B**) situations with the following Q-models: the theoretical logistic regression, elasticnet logistic regression, lasso logistic regression, neural network, support vector machine, boosted CART and super learner.

value of 95%. For the smallest sample size, in both the realistic and simplistic situations, the RMSEs of the penalized methods were among the highest because of their high-level of variance (simplistic situation) or high levels of bias (realistic situation).

When $n \geq 500$, the RMSEs of the penalized methods were close to those observed for the ML-based methods (NN, SVM and SL). However, for these two approaches, one can observe slightly anti-conservative 95% CIs in the realistic situation, because of their slight biases. For the remaining ML-based methods, the RMSEs were comparable for the three sample sizes and in the two situations, but the results of the NN should be interpreted with caution. Indeed, for $n = 100$, the NN was associated with a significant bias, but a low variance estimation, resulting in a CI of 86.6%, lower than the nominal value of 95%.

As previously reported, the two remaining methods (SL and SVM) were the two ML techniques associated with the smallest MBs. For each scenario, the MB of the SL was even lower than the value of SVM. This explains why the nominal coverage was slightly higher when using the SL. For instance in the realistic scenario, the coverage values associated with the SVM were 92.6%, 93.7% and 91.4% for n =100, 500 and 1000, respectively, while they were 93.1%, 95.2% and 94.6% for the SL.

Power of the unbiased methods. We only consider the methods and the scenarios in which the MB were lower than 1% due to the problems encountered when interpreting the power in the presence of bias.

The realistic situation. When $n = 100$, the SVM and SL had MBs lower than 1%. Of the two methods, the best power was achieved by the SVM (36.5% vs 30.8% for the SL). When $n = 1000$, the ELR, LLR and SL had MBs lower than 1%, and the best power values were achieved by the penalized methods (92.4% for the ELR, 91.5% for the LLR and 89.3% for the SL).

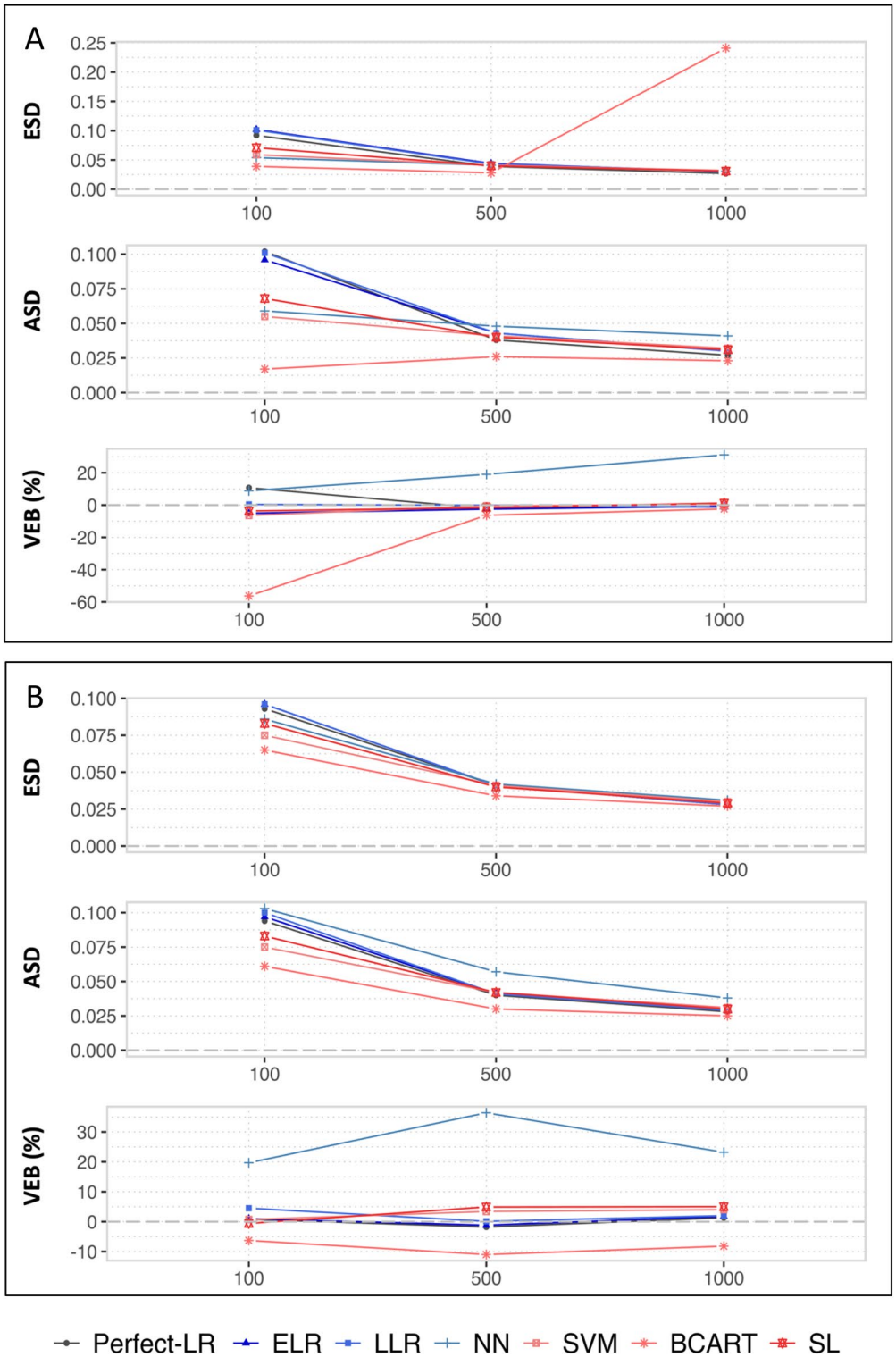


Figure 3. Empirical and asymptotic standard deviations (ESDs and ASDs, respectively) and variance estimation biases (VEBs) of G-computation in realistic (**A**) and simplistic (**B**) situations with the following Q-models: the theoretical logistic regression, elasticnet logistic regression, lasso logistic regression, neural network, support vector machine, boosted CART and super learner.

The simplistic situation. When n =100, only the penalized methods had MBs lower than 1%. The best power was obtained by the ELR (20.2% versus 18.0% for the LLR). When $n \geq 500$, we additionally observed MBs lower than 1% for the SVM and SL. The penalized methods were always associated with the best powers when compared with those of the two ML techniques with a gain between 1 and 4% depending on the scenarios.

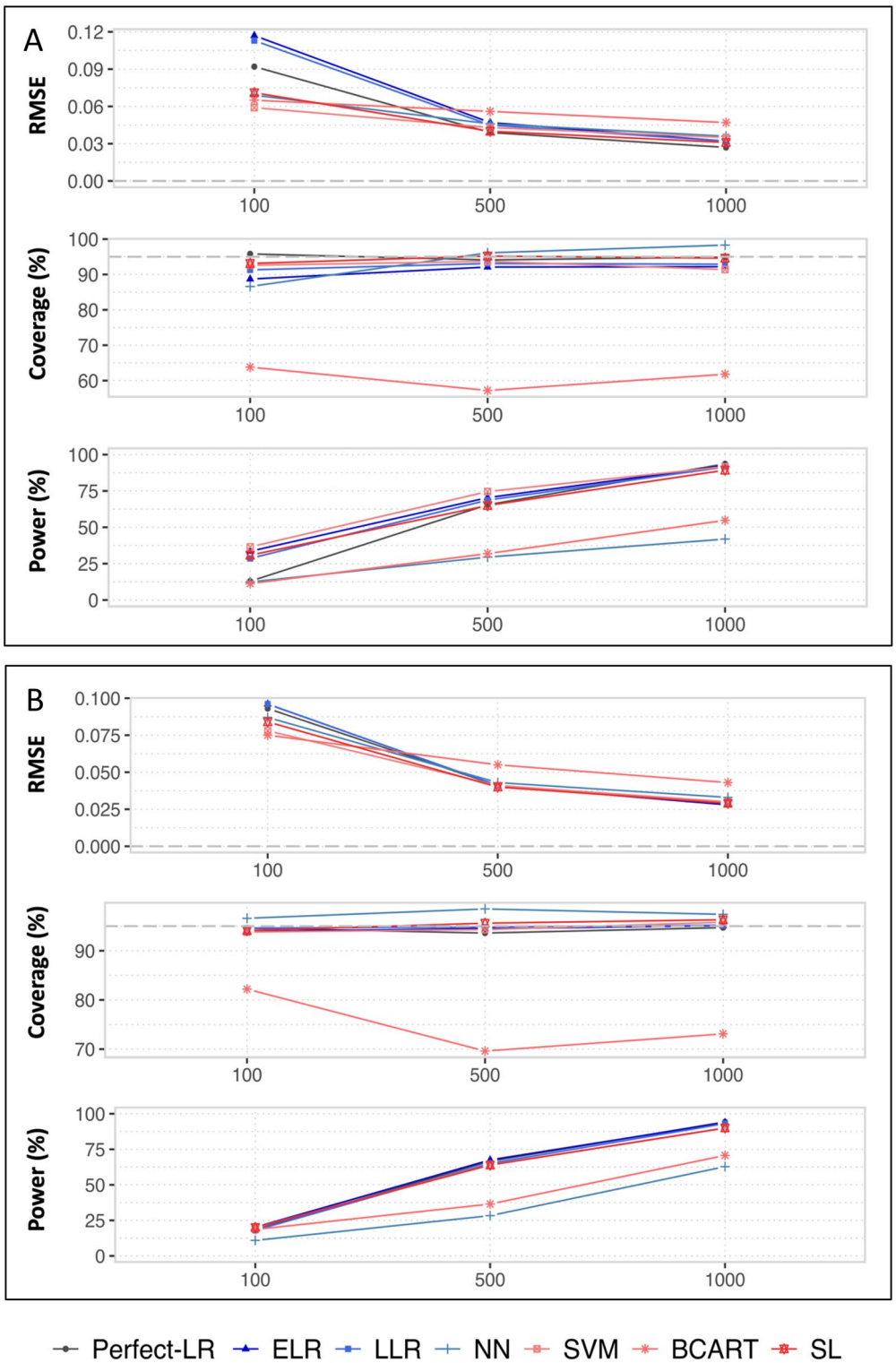


Figure 4. Root mean square errors (RMSEs), coverages and powers of G-computation in realistic (**A**) and simplistic (**B**) situations with the following Q-models: the theoretical logistic regression, elasticnet logistic regression, lasso logistic regression, neural network, support vector machine, boosted CART and super learner.

ML techniques versus the perfectly specified LR. The performances of the perfectly specified LR were better than those of the ML techniques for large sample sizes (n =1000). One can observe mean bias values close to 0%, and variance bias values close to 1%. Nevertheless, when the sample size decreased in the realistic situation, the performances of the perfectly specified LR decreased more than those of several ML techniques. When n =500, the variance bias associated with the perfectly specified LR was − 2.1% versus − 0.1% for the LLR,

− 0.5% for SVM and − 1.6% for the SL. When n =100, the variance bias associated with the perfectly specified LR was 10.7% versus 0.4% for the LLR, − 3.7% for the SL, and − 6.4% for the SVM. In this latter scenario, these three ML techniques resulted in higher statistical powers than the one obtained with the perfectly specified LR.

Application

Context. We applied the methods to evaluate the efficacy of barbiturates prescribed during the first 24 h of an episode of intracranial hypertension. The control group included patients without barbiturates at 24 h. One can use this treatment to decrease refractory intracranial pressure, but its effectiveness remains debated due to the associated adverse events (*e.g.*, haemodynamic impacts or infectious complications).

We used data from the French prospective cohort AtlanREA. We considered patients with intracranial pressures higher than 20 mmHg. We conducted this study following French law relative to non-interventional clinical research. Written informed consent was collected. Moreover, the French commission for data protection approved the collection (CNIL DR-2013-047). The study was approved by the AtlanREA scientific council (www.atlanrea.org) and the ethics committee of the French Society of Anesthesia and Intensive Care (SFAR, https://sfar.org/).

Implementation of the methods. We reduced the set of covariates to the possible causes of the outcome without considering the consequences of barbiturate use. We described this selection in detail in Supplementary Table S5. For the ML-based methods, we considered all the covariates before exposure and the corresponding interactions with the exposure status. As in the previous simulations, we used b-splines for the continuous covariates in the penalized methods. For the investigator-based method, all the outcome causes previously listed were included (Supplementary Table S5). The log-linearity assumption for continuous covariates seemed to be satisfied. We assumed that there was no interaction because of the absence of clinical relevance.

Results

Table 1 describes the 252 patients. Seventy-four patients were in the treatment group. The outcome was the proportion of patients with a favourable Glasgow Outcome Scale (GOS ≤ 3) at three months after admission to the intensive care unit. Figure 5 presents the confounder-adjusted estimates. The investigator-based approach resulted in a 17.5% decrease in the percentage of patients with favourable 3-month GOS due to barbiturates (95% CI from 6.6 to 28.4%). We observed similar results for the ELR and LLR, in terms of both the estimates and the 95% CIs. The other ML techniques resulted in lower associations, and the one fpr the NN was even nonsignificant (ACE =0.4%, 95% CI from − 3.1 to 2.4%). The SL resulted in a small but significant association (ACE =6.2%, 95% CI from 0.6% to 11.8%).

For a MacBook pro with a 2.6 GHz Intel Core i7 processor (6 cores), the results were available in 6.5 min for the ELR, 16.3 min for the LLR, 7.1 min for the NN, 2.3 min for the BCART, 2.6 min for the SVM, and 7 min for the SL.

Discussion

When modelling the outcome model for the GC in the presence of small to medium sample sizes, the results of our simulations tended to demonstrate that ML techniques allow for accurate estimations of causal effects. Overall, the SL remained robust in all situations and achieved a relevant compromise between both bias reduction and variance estimation. In contrast, the performances of the other methods tended to vary more significantly according to the complexity of the relationships between the covariates and the outcome (simplistic versus realistic situations) and the sample size. Nevertheless, in some situations, the other methods obtained better performances than those of the SL. When the sample size was small (n = 100) in the realistic scenario, the SVM had a larger MB but a smaller ASD, with an overall smaller RMSE. In this situation, the two ML techniques (SL and SVM) were even associated with lower variances than that of the perfectly specified LR. For instance, the variance bias was − 3.7% for the SL versus 10.7% for the perfectly specified LR. One can explain this result by the sample-to-sample fluctuation, which can lead to an observed structure that is different from the theoretical one. When the sample size was small in the simplistic scenario, the penalized methods (ELR and LLR) had lower MBs and similar RMSEs.

The use of ML techniques for causal inference does not preclude human intervention. In addition to the choice of the Q-model, we need to exclude the mediators, colliders and instrumental variables by considering the underlying causal structure. The use of directed acyclic graphs can help with this task[33]. We also emphasise that ML techniques do not serve as a cure-all for poor study designs or poor data quality. It is of primary importance to investigate the identifiability conditions: the exposure levels correspond to well-defined interventions, the corresponding conditional probabilities depend only on the measured covariates, and must be higher than zero. These assumptions are consistency, exchangeability, and positivity, respectively[34]. In this paper, we focused on the estimation of a causal effect given that the identifiability conditions were satisfied. In practice, the predictive performance of the Q-model is not sufficient to ensure the absence of bias in the estimation of the causal effect, which requires a precise conceptual knowledge of the causal model[35].

Perfect knowledge of the causal structure is impossible to obtain in practice. Therefore, the analyst and the investigator construct the Q-model to approximate the causal structure as closely as possible. This may involve different steps such as the transformation of the continuous covariates to respect the log-linearity assumption, the selection of the covariates, or the choice of relevant interaction(s). While the steps performed by the analyst are data-driven and stochastic, they are systematically ignored in the estimation of the effect variance[36]. The widespread interest in (human-free) ML stems from the possibility of considering a valid post-selection inference by bootstrapping the entire estimation procedure[29].

	Overall ($n = 252$)		Barbiturates treatment					Favourable GOS at three months				
			No ($n = 178$)		Yes ($n = 74$)		p	No ($n = 180$)		Yes ($n = 72$)		p
Female patient (n, %)	89	35.3	58	32.6	31	41.9	0.1592	68	37.8	21	29.2	0.1963
Diabetes (n, %)	17	6.7	15	8.4	2	2.7	0.0989	15	8.3	2	2.8	0.1122
No sological entity: severe trauma (n, %)	124	49.2	95	53.4	29	39.2	0.0403	77	42.8	47	65.3	0.0012
SAP ≤ 90 mmHg before admission (n, %)	56	22.2	36	20.2	20	27.0	0.2368	46	25.6	10	13.9	0.0442
Evacuation of subdural or extradural hematoma (n, %) (*)	41	16.3	33	18.5	8	10.8	0.1301	27	15.0	14	19.4	0.3878
External ventricular drain (n, %)	64	25.4	39	21.9	25	33.8	0.0486	48	26.7	16	22.2	0.4640
Evacuation of cerebral hematoma or lobectomy (n, %) (*)	42	16.7	28	15.7	14	18.9	0.5362	34	18.9	8	11.1	0.1345
Decompressive craniectomy (n, %) (*)	27	10.7	15	8.4	12	16.2	0.0686	21	11.7	6	8.3	0.4396
Blood transfusion before admission (n, %)	34	13.5	25	14.0	9	12.2	0.6903	26	14.4	8	11.1	0.4841
Pneumonia (n, %) (*)	29	11.5	16	9.0	13	17.6	0.0519	19	10.6	10	13.9	0.4538
Osmotherapy (n, %) (*)	112	44.4	75	42.1	37	50.0	0.2525	89	49.4	23	31.9	0.0115
GCS score ≥ 8 (n, %)	62	24.6	39	21.9	23	31.1	0.1237	37	20.6	25	34.7	0.0183
Patient age, years (mean, sd)	47.4	17.4	48.7	17.9	44.1	15.7	0.0565	50.8	16.4	38.7	16.9	0.0000
Haemoglobin, g/dL (mean, sd)	11.8	2.3	11.7	2.2	12.1	2.5	0.1824	11.8	2.4	11.9	1.9	0.7373
Platelets, counts/mm^3 (mean, sd)	206.7	78.0	207.4	79.7	205.1	74.2	0.8312	209.0	83.8	200.9	61.1	0.4589
Serum creatinine, mmol/L (mean, sd)	71.1	29.3	71.1	27.6	71.1	33.3	0.9853	72.4	32.6	67.9	18.7	0.2732
Arterial pH (mean, sd)	7.3	0.1	7.3	0.1	7.3	0.1	0.0978	7.3	0.1	7.3	0.1	0.6317
Serum proteins, g/L (mean, sd)	58.2	10.4	57.7	10.6	59.6	9.7	0.1662	58.0	10.7	58.8	9.7	0.5963
Serum urea, mmol/L (mean, sd)	5.0	2.5	5.2	2.7	4.7	1.8	0.1827	5.2	2.3	4.5	2.9	0.0505
PaO_2/FiO_2 ratio (mean, sd)	302.7	174.0	292.7	154.7	326.6	212.9	0.1595	282.1	172.4	354.2	168.4	0.0028
SAPS II score (mean, sd)	47.6	11.4	47.6	10.7	47.6	12.9	0.9847	49.9	10.8	41.8	10.7	0.0000

Table 1. Baseline characteristics of patients according to the treatment group ($n = 252$) and the GOS at three months after the treatment initiation. GOS score was dichotomised into favourable outcomes (good recovery or moderate disability) or unfavourable outcomes (severe disability, vegetative state or death). GOS, Glasgow outcome Scale; SAP, systolic arterial pressure; HICP, high intracranial pressure; GCS, Glasgow Coma Scale; PaO_2, partial arterial pressure of oxygen; FiO_2, fraction of inspired oxygen; SAPS, Simplified Acute Physiology Score. *Before HICP.

ML techniques are often associated with big data, especially in the field of causal inference[8,37,38]. Nevertheless, we described the acceptable properties of the SL used in a GC framework to provide causal inference conclusions from databases including several hundred subjects. To obtain this result, we first selected several simple ML techniques. We excluded deep learning techniques, such as neural networks with multiple hidden layers. Second, we retained the ML techniques that allow for maintaining the exposure as one of the predictors. Third, we included two parametric models. Fourth, we used bootstrap cross-validation to prevent overfitting. Fifth, we used two ML techniques (NN and SVM) for which there was no selection of predictors. Consequently, all covariates were also included in the SL, even those with low contributions due to having no association. The removal of confounders in GC can result in confounding bias, which can explain the poor performances of the penalized methods in realistic situations. These choices participated in the lower bias of the SL versus that of BCART. Our GC results are in agreement with the conclusions of Gruber et al., which concerned PS-based analyses[8]. Indeed, BCART is an ensemble learning method that avoids cross-validation by a single partitioning of the data into training and validation sets. It allows us to reduce the computational time, but it should be used with caution for small sample sizes.

Our study suffered from limitations. First, the results from the simulations cannot be generalised to all situations. Even if they are consistent with the current literature related to the use of ML in PS-based analyses, theoretical arguments are missing for generalisation purposes. Second, one perspective of our work is to improve the proposed SL with additional ML techniques or differently tuned techniques. For instance, we fixed the length of the tuning grid at 20; a lower value may be acceptable for reducing the computational time. The V-fold cross-validation is also an important parameter. We fixed $V = 10$, as conventionally used. A more appropriate choice could also be studied. For example, Naimi and Balzer recommended increasing V as the sample size decreases[22]. Third, we focused on the comparison of the ML techniques used in GC. We did not perform comparisons with other methods used for causal inference, such as the influence function-based or doubly robust estimators. In particular, the double/debiased machine learning and targeted maximum likelihood estimator allow for the unrestricted use of data-adaptive methods[38]. The principle is to combine the modelling of the outcome and exposure mechanisms to obtain an unbiased estimate when at least one of the two models is well-specified. However, such doubly robust estimators also have several drawbacks. If both models are misspecified, the estimation is more biased than that of a single-robust estimator such as GC[14]. The inclusion of a mediator also leads to more bias than that of GC[39]. Several studies have additionally reported that GC has a lower variance than those of doubly robust estimators[11–14]. As previously stated, the use of GC also represents a partial solution for preventing the selection of instrumental variables since it is independent of the exposure modelling. Fourth, our study focused

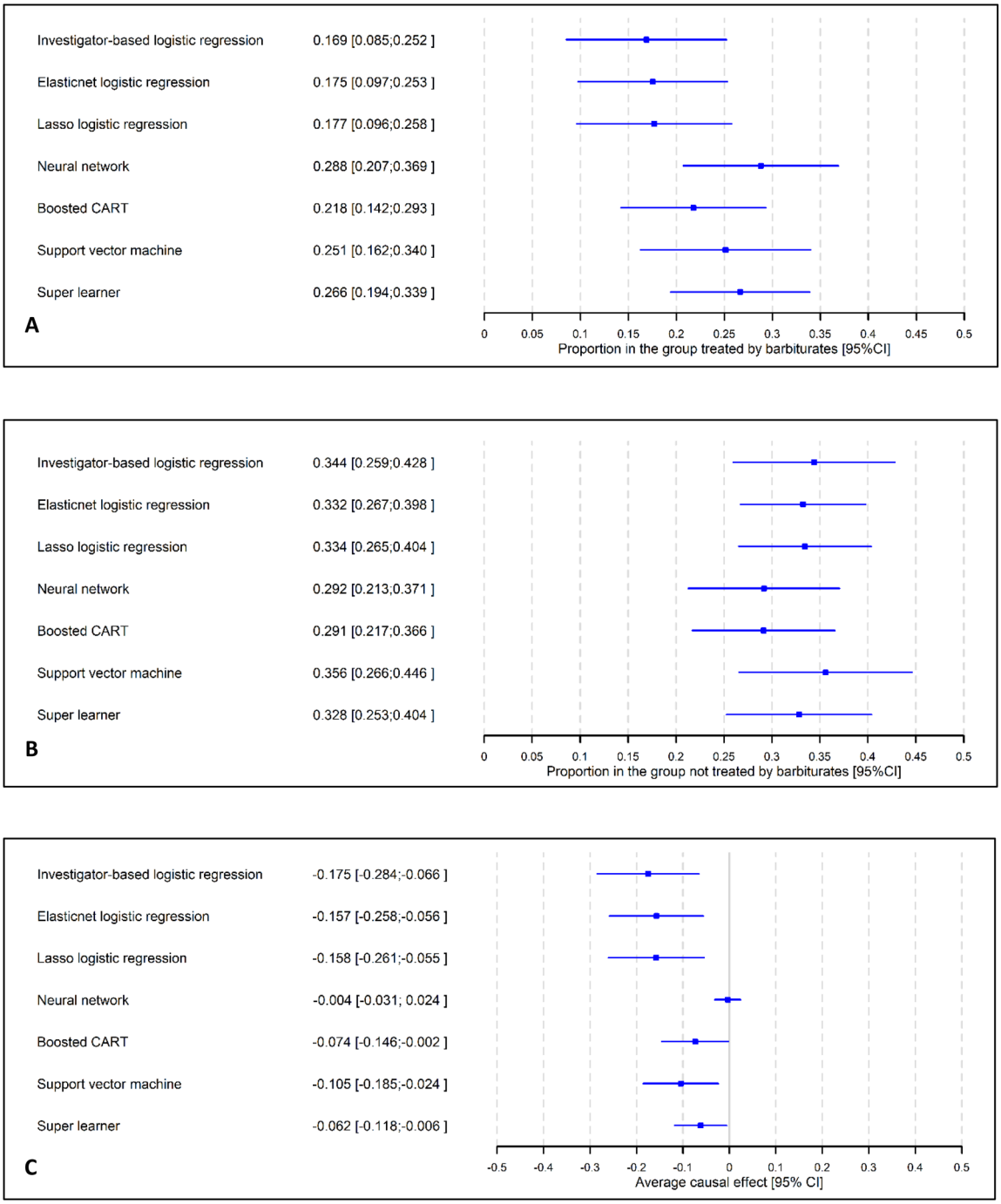


Figure 5. Estimations of the confounder-adjusted proportions of patients with favourable GOS among the patients treated with barbiturates (**A**), patients not treated with barbiturates during the first 24 h postadmission (**B**), and the corresponding average causal effects (**C**).

on the situation where both the exposure status and the outcome are binary. The generalisation of our approach to other contexts, especially for time-to-event outcomes, represents a short-term goal. Finally, we focused on the *ACE* if the entire sample had been exposed and if it had not been exposed. Additional analyses are needed to confirm these results to estimate the average causal effect only for the exposed individuals[40].

In conclusion, the super-learned G-computation is a promising method for causal inference, even with only several hundred subjects. The SVM represents an interesting alternative for small sample sizes with one hundred subjects when the relationships between the covariates and the outcome are complex. For such a small sample size, penalized methods appeared to be the best alternatives when the relationships were simplistic (few covariates with linear relationships and without interactions). The computation times of these ML techniques associated with GC were reasonable. Note that GC with the SL as the Q-model is implemented in the *RISCA* package (cran.r-project.org, version ≥ 0.82). The user can set the number of splits for cross-validation and the number of

parameter combinations to be evaluated. This is a particular solution, but it is not recommended for analysing any type of data using the same algorithm. We believe that such ML techniques constitute an opportunity for analysts to save some of their time used for repetitive modelling steps and use it for applying prior knowledge of the medical field and improving their comprehension of the given data structure.

Received: 6 July 2020; Accepted: 24 December 2020
Published online: 14 January 2021

References

1. Blakely, T., Lynch, J., Simons, K., Bentley, R. & Rose, S. Reflection on modern methods: when worlds collide: prediction, machine learning and causal inference. *Int. J. Epidemiol.* https://doi.org/10.1093/ije/dyz132.
2. Rosenbaum, P. R. & Rubin, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55 (1983).
3. Westreich, D., Lessler, J. & Funk, M. J. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J. Clin. Epidemiol.* **63**, 826–833 (2010).
4. Robins, J. M. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math. Model.* **7**, 1393–1512 (1986).
5. Snowden, J. M., Rose, S. & Mortimer, K. M. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am. J. Epidemiol.* **173**, 731–738 (2011).
6. Setoguchi, S., Schneeweiss, S., Brookhart, M. A., Glynn, R. J. & Cook, E. F. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiol. Drug Saf.* **17**, 546–555 (2008).
7. Lee, B. K., Lessler, J. & Stuart, E. A. Improving propensity score weighting using machine learning. *Stat. Med.* **29**, 337–346 (2010).
8. Gruber, S., Logan, R. W., Jarrín, I., Monge, S. & Hernán, M. A. Ensemble learning of inverse probability weights for marginal structural modeling in large observational datasets. *Stat. Med.* **34**, 106–117 (2015).
9. Pirracchio, R., Petersen, M. L. & van der Laan, M. Improving propensity score estimators' robustness to model misspecification using super learner. *Am. J. Epidemiol.* **181**, 108–119 (2015).
10. Cannas, M. & Arpino, B. A comparison of machine learning algorithms and covariate balance measures for propensity score matching and weighting. *Biom. J.* **61**, 1049–1072 (2019).
11. Chatton, A. *et al.* G-computation, propensity score-based methods, and targeted maximum likelihood estimator for causal inference with different covariates sets: a comparative simulation study. *Sci. Rep.* (i**n press**) (2020).
12. Lendle, S. D., Fireman, B. & van der Laan, M. J. Targeted maximum likelihood estimation in safety analysis. *J. Clin. Epidemiol.* **66**, S91-98 (2013).
13. Colson, K. E. *et al.* Optimizing matching and analysis combinations for estimating causal effects. *Sci. Rep.* **6**, 23222 (2016).
14. Kang, J. D. Y. & Schafer, J. L. Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. *Stat. Sci.* **22**, 523–539 (2007).
15. Austin, P. C. Using ensemble-based methods for directly estimating causal effects: an investigation of tree-based G-computation. *Multivar. Behav. Res.* **47**, 115–135 (2012).
16. Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.* **66**, 688–701 (1974).
17. Hernán, M. A. A definition of causal effect for epidemiological research. *J. Epidemiol. Commun. Health* **58**, 265–271 (2004).
18. Lin, S.-H. & Ikram, M. A. On the relationship of machine learning with causal inference. *Eur. J. Epidemiol.* https://doi.org/10.1007/s10654-019-00564-9 (2019).
19. VanderWeele, T. J. Principles of confounder selection. *Eur. J. Epidemiol.* **34**, 211–219 (2019).
20. Myers, J. A. *et al.* Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am. J. Epidemiol.* **174**, 1213–1222 (2011).
21. Brookhart, M. A. *et al.* Variable selection for propensity score models. *Am. J. Epidemiol.* **163**, 1149–1156 (2006).
22. Naimi, A. I. & Balzer, L. B. Stacked generalization: an introduction to super learning. *Eur. J. Epidemiol.* **33**, 459–464 (2018).
23. Keil, A. P. & Edwards, J. K. You are smarter than you think: (super) machine learning in context. *Eur. J. Epidemiol.* **33**, 437–440 (2018).
24. VanderWeele, T. J. & Shpitser, I. A new criterion for confounder selection. *Biometrics* **67**, 1406–1413 (2011).
25. Lefebvre, G., Delaney, J. A. C. & Platt, R. W. Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Stat. Med.* **27**, 3629–3642 (2008).
26. McNeish, D. M. Using lasso for predictor selection and to assuage overfitting: a method long overlooked in behavioral sciences. *Multivar. Behav. Res.* **50**, 471–484 (2015).
27. Bi, Q., Goodman, K. E., Kaminsky, J. & Lessler, J. What is machine learning? A primer for the epidemiologist. *Am. J. Epidemiol.* https://doi.org/10.1093/aje/kwz189 (2019).
28. van der Laan, M. J., Polley, E. C. & Hubbard, A. E. Super learner. *Stat. Appl. Genet. Mol. Biol.* **6**, Article25 (2007).
29. Efron, B. Estimation and accuracy after model selection. *J. Am. Stat. Assoc.* **109**, 991–1007 (2014).
30. Schumacher, M., Binder, H. & Gerds, T. Assessment of survival prediction models based on microarray data. *Bioinformatics* **23**, 1768–1774 (2007).
31. Foucher, Y. & Danger, R. Time dependent ROC curves for the estimation of true prognostic capacity of microarray data. *Stat. Appl. Genet. Mol. Biol.* **11**, Article 1 (2012).
32. Gayat, E., Resche-Rigon, M., Mary, J.-Y. & Porcher, R. Propensity score applied to survival data analysis through proportional hazards models: a Monte Carlo study. *Pharm. Stat.* **11**, 222–229 (2012).
33. Hernán, M. A., Hsu, J. & Healy, B. A second chance to get causal inference right: a classification of data science tasks. *Chance* **32**, 42–49 (2019).
34. Hernán, M. A. & Taubman, S. L. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int. J. Obes.* **32**, S8–S14 (2008).
35. Díaz, I. Machine learning in the estimation of causal effects: targeted minimum loss-based estimation and double/debiased machine learning. *Biostatistics* **21**, 353–358 (2020).
36. Berk, R., Brown, L., Buja, A., Zhang, K. & Zhao, L. Valid post-selection inference. *Ann. Stat.* **41**, 802–837 (2013).
37. Wyss, R. *et al.* Using super learner prediction modeling to improve high-dimensional propensity score estimation. *Epidemiology* **29**, 96–106 (2018).
38. Karim, M. E., Pang, M. & Platt, R. W. Can we train machine learning methods to outperform the high-dimensional propensity score algorithm?. *Epidemiology* **29**, 191–198 (2018).
39. Keil, A. P. *et al.* Resolving an apparent paradox in doubly robust estimators. *Am. J. Epidemiol.* **187**, 891–892 (2018).
40. Pirracchio, R. *et al.* Propensity score estimators for the average treatment effect and the average treatment effect on the treated may yield very different estimates. *Stat. Methods Med. Res.* **25**, 1938–1954 (2016).

Acknowledgements
The authors would like to thank the members of AtlanREA group for their involvement in the study, the physicians who helped recruit patients and all patients who participated in this study. We also thank the clinical research associates who participated in the data collection. The analysis and interpretation of these data are the responsibility of the authors.

Author contributions
Y.F. supervised this work and performed the simulations and other statistical analyses. F.L.B. and A.C. participated in the design of the simulation-based study. All the authors were engaged in the writing of the final proposal.

Funding
The French National Research Agency (ANR) partially supported this work (Research in Informatic and Statistic for Cohort Analyses, www.labcom-risca.com, reference: ANR-16-LCV1-0003-01). The funder had no role in study design; analysis, and interpretation of data; writing the report; and the decision to submit the report for publication.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information is available for this paper at https://doi.org/10.1038/s41598-021-81110-0.

Correspondence and requests for materials should be addressed to Y.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021

DOCTORAT BIOLOGIE BRETAGNE SANTE LOIRE

Nantes Université

Titre : Inférence causale sur données observationnelles : développement et application pour les soins critiques

Mots clés : Inférence causale, scores de propension, G-computation, soins critiques

Résumé :

L'augmentation croissante des données observationnelles, notamment dans les services de soins critiques amènent à considérer l'utilisation des méthodes statistiques d'inférence causale.

Cette thèse présente deux travaux permettant de mettre en lumière les défis actuels de ces méthodes statistiques appliquées sur des données observationnelles de santé.

Le premier travail évalue l'impact des barbituriques dans une population de patients traumatisés crâniens, inclus de manière prospective dans la cohorte ouverte de soins critiques AtlanRéa. L'évaluation de l'impact des barbituriques a été permis par le respect des hypothèses de l'inférence causale et l'utilisation d'une méthode basée sur les scores de propension : la pondération.

Au-delà du résultat de cette analyse, ayant notamment mis en évidence une augmentation de la mortalité dans le groupe traité par barbituriques, nous avons été confrontés à la problématique de l'infraction de l'hypothèse de positivité.

Nous avons ensuite comparé différentes méthodes statistiques d'inférence causale dans un contexte d'infraction de l'hypothèse de positivité, pouvant être associée à la problématique d'extrapolation. Les méthodes prédisant la survenue de l'évènement sont les plus robustes dans ces situations.

Dans ce contexte d'accumulation de données de santé, une perspective d'optimisation de l'utilisation des méthodes statistiques dans le cadre de l'inférence résidera dans le recours aux algorithmes d'apprentissage automatisé (*machine learning*) pour éviter les problèmes de spécification des modèles.

Title : Causal inference from observational data: development and applications for critical care

Keywords : Causal inference, propensity score, G-computation, critical care

Abstract :

The increasing amount of observational data, especially in critical care, leads to the consideration of causal inference statistical methods.

This thesis presents two works that highlight the current challenges of these statistical methods used on observational health data.

The first analysis evaluates the impact of barbiturates in a population of trauma brain injured patients prospectively included in the open critical care cohort AtlanRéa. The evaluation of the impact of barbiturates was possible by respecting the assumptions of causal inference and by using a method based on propensity scores: inverse probability weighting.

Beyond the results of this analysis, which showed an increase in mortality in the group treated with barbiturates, we were faced with the problem of the violation of the positivity assumption.

We then compared different statistical methods of causal inference in a context of violation of the positivity assumption, which can be associated with an extrapolation issue. The methods predicting the occurrence of the outcome are the most robust in these situations.

In this context of accumulation of health data, a perspective of optimization of the use of statistical methods in the framework of causal inference will reside in the use of machine learning algorithms to avoid the problems of model specification.