

Représentation, régularisation et complexité fonctionnelle : une déconfusion expérimentale contrôlée sur données de survie

Résumé

À information d'entrée identique, deux dispositions d'un même tenseur ne produisent pas le même apprentissage. Nous partions de l'hypothèse qu'une représentation continue du temps porterait un avantage propre ; le banc la soumet à falsification. Nous soutenons, plus prudemment, que la disposition participe au biais inductif en structurant les relations, voisinages et partages de paramètres accessibles au modèle, l'invariance effective résultant du couple représentation-architecture-contrainte, non de la disposition seule. Nous testons cela sur des données de survie reconstruites depuis des courbes publiées, dans une seule famille de survie à temps discret pénalisée, par deux expériences, la fidélité de forme et l'estimation d'un effet structurel sous confusion simulée. Le plan est gelé et haché avant le run confirmatoire. Le résultat brut suggère un avantage de la base partagée en fidélité ; un plan factoriel à degrés de liberté effectifs appariés montre que cet avantage n'est pas attribuable à la continuité, le discret pénalisé égalant ou dépassant légèrement la spline continue, ce qui rattache l'effet à la régularisation. Sous confusion simulée, une fois l'incertitude quantifiée à rétention pleine, aucune différence robuste n'apparaît, la probabilité que le continu batte le plat restant proche du hasard et aucune cellule ne survivant à la correction de multiplicité. La contribution est donc une déconfusion méthodologique : dans les configurations testées, l'avantage apparent d'une représentation continue s'explique par la régularisation et les degrés de liberté effectifs, sans avantage propre détectable de la continuité, et aucune différence robuste d'estimation n'est détectée sous confusion. L'unité expérimentale demeure le pipeline représentationnel complet, non la disposition isolée.

1. Question de recherche

La question est simple à énoncer et difficile à clore. À information d'entrée identique, la disposition du tenseur change-t-elle ce qu'un modèle apprend ? La version superficielle oppose le temps porté comme axe au temps éclaté en colonnes, ordre contre absence d'ordre. La version profonde, qui est la nôtre, demande si les différences de performance attribuées à la disposition proviennent de la structure de représentation elle-même, ou de la classe fonctionnelle, du partage de paramètres et de la régularisation qu'elle induit dans un pipeline donné. C'est cette dernière question que le banc teste.

Deux expériences organisent le travail. La fidélité, la disposition change-t-elle la reconstruction de la forme de survie. Le biais causal, la disposition change-t-elle l'inférence d'un effet de traitement sous confusion. Une troisième piste, l'augmentation de données

respectant la symétrie déclarée, est présentée en perspective et non exécutée ici ; nous la sortons du cœur de l'article pour ne pas annoncer un résultat que nous ne produisons pas.

2. La tensorisation est une hypothèse d'invariance

Une représentation n'est pas un contenant neutre où l'on verserait des données déjà constituées. C'est une hypothèse sur la structure du problème. Avant qu'un modèle n'apprenne quoi que ce soit, la disposition choisie a déjà déclaré ce qui, dans le phénomène, doit rester invariant. Le format ne transporte pas l'information, il énonce sa géométrie.

Lisons chaque disposition pour ce qu'elle rend disponible, sans lui prêter plus qu'elle ne porte. Le format en colonnes indépendantes n'impose aucune métrique ni aucun ordre entre les dimensions ; il rend naturelle, mais n'impose pas, une invariance par permutation, laquelle n'existe que si les paramètres du modèle sont eux aussi permutés conjointement, ce qu'une régression à coefficient distinct par colonne ne fait pas. La disposition en axe rend disponibles un ordre, un voisinage et un partage de paramètres le long du temps ; elle n'entraîne pas pour autant une invariance à toute reparamétrisation monotone du temps, car une spline ou une base polynomiale voit ses dérivées et ses pénalités de lissage modifiées par une telle transformation. La disposition hiérarchique rend naturel un emboîtement en segments avec remise à zéro aux frontières. Trois dispositions, trois structures d'espace rendues disponibles, non trois invariances imposées.

Il faut alors séparer trois niveaux que le mot symétrie tend à confondre. Le premier est la structure de l'espace d'entrée, ordre, métrique, voisinage, hiérarchie, que la disposition fixe. Le deuxième est l'hypothèse fonctionnelle, lissage, localité, additivité, proportionnalité, que le couple représentation-modèle rend peu coûteuse ou non. Le troisième seulement est la symétrie proprement dite, invariance ou équivariance sous un groupe explicitement défini. Écrivons le modèle comme composition d'une représentation et d'une tête, $f = h \circ \varphi$, où φ envoie l'entrée dans l'espace de représentation et h apprend dessus. La symétrie effective de f dépend conjointement de φ et de h ; la disposition, qui fixe φ , ne la détermine pas seule. La formulation défendable n'est donc pas que choisir une disposition revient à choisir un groupe de symétrie, mais que la disposition rend certaines classes de symétries, de voisinages et de partages de paramètres plus naturelles que d'autres, l'invariance résultant du triple représentation, architecture, contrainte. C'est moins spectaculaire et bien plus solide. L'image condensée, la tensorisation comme moment où l'ontologie devient géométrie, garde sa valeur à cette condition de lecture : la disposition déclare la géométrie disponible, elle ne clôt pas à elle seule ce que le modèle apprendra.

Cette lecture reformule la question empirique et lui donne son nom. L'enjeu n'est pas l'axe contre la colonne, opposition de surface qui masque le fond. L'enjeu empirique est l'accord entre la structure rendue accessible par la représentation, la contrainte effectivement imposée par le modèle et sa régularisation, et la régularité du phénomène à apprendre. Quand ce triple s'accorde à la structure réelle de l'effet, le modèle hérite d'une contrainte utile et apprend mieux ; quand il impose une contrainte fausse, il apprend moins bien, à capacité et procédure de régularisation données. La disposition n'est qu'une des trois pièces de ce

pari, et le pari peut être perdu. Nous verrons d'ailleurs que, dans notre banc, l'essentiel de l'écart passe par la régularisation et non par la disposition seule.

3. État de l'art

Ce travail se tient au croisement de quatre lignées, dont aucune, seule, ne nomme son objet.

La première est celle du biais inductif et de l'invariance. Qu'un apprentissage sans biais ne généralise pas est acquis de longue date [1], et la théorie des représentations en a fait un objet explicite [2]. Sa formulation la plus aboutie est géométrique, une architecture encode une hypothèse d'invariance sous un groupe de transformations, dont découle le partage de paramètres [3-5]. Ce vocabulaire a été forgé pour les images et les graphes ; nous le transportons vers la disposition des covariables et du temps dans un tenseur de survie.

La deuxième est celle des modèles de survie profonds, de DeepSurv [6] à DeepHit et Dynamic-DeepHit [7,8], à la survie neuronale à temps discret [9] et au cadre exponentiel par morceaux qui fonde la famille employée ici [10]. Tous font varier l'architecture. Aucun ne fait varier la disposition à architecture fixée.

La troisième est celle de l'apprentissage profond tabulaire, TabTransformer, FT-Transformer, SAINT [11-13], avec le rappel que les arbres de gradient dominent souvent encore [14]. Ces travaux demandent comment encoder une variable, non quelle invariance sa disposition déclare, et ignorent la structure temps-événement.

La quatrième est celle de la comparaison indirecte ajustée sur la population, versant applicatif. La reconstruction de données individuelles depuis des courbes publiées [15] alimente la comparaison appariée et simulée, avec la distinction cardinale entre comparaison ancrée et non ancrée [16-18]. Cette lignée consomme la donnée reconstruite ; notre générateur en est un producteur, et la disposition gouverne la fidélité du bras synthétique produit.

Une cinquième lignée, que nous devons reconnaître car nos résultats de fidélité en relèvent directement, est celle de l'analyse fonctionnelle et de la régularisation temporelle : expansions de base et lissages par produit tensoriel [19], modèles à coefficients variables [20], régularisations imposant la parcimonie des différences comme le fused lasso et le trend filtering [21,22]. Le compromis que notre banc met en évidence, paramètres séparés par intervalle contre fonction lisse partagée sur le temps, y est un problème ancien et bien caractérisé. Notre apport n'est pas de le découvrir mais de le relire en termes de disposition ; l'honnêteté commande de nommer cette parenté plutôt que de présenter comme rupture ontologique une reformulation géométrique d'une régularisation fonctionnelle connue.

Le point aveugle est commun. On a étudié ce qu'on met dans le tenseur et quelle architecture on fait tourner dessus. On n'a pas étudié la forme du tenseur elle-même comme biais inductif, ni relié systématiquement ce choix à la régularisation fonctionnelle qu'il induit. C'est ce lien que nous nommons, avec la prudence que nos résultats imposent.

4. Données

Le périmètre est le banc causal, deux essais de deuxième ligne en lymphome B à grandes cellules, TRANSFORM et ZUMA-7. Aucune donnée individuelle n'est mobilisée. Nous reconstruisons des jeux temps-événement-censure à partir des seules courbes de Kaplan-Meier publiées et de leurs tables d'effectifs à risque, par la méthode de Guyot et collègues [15]. Le terme juste est *compatible*, non *identique* : l'algorithme ne retrouve pas les trajectoires réelles, il régénère une distribution qui reproduit la courbe.

La validation réestime le rapport de risque sur le jeu reconstruit et vérifie qu'il retombe sur la valeur publiée à moins de trois centièmes.

Source	Critère	HR publié	HR reconstruit	Écart
TRANSFORM	survie sans événement	0,375	0,359	-0,016
TRANSFORM	survie globale	0,757	0,751	-0,006
ZUMA-7	survie sans événement	0,398	0,398	0,000
ZUMA-7	survie globale	0,730	0,729	-0,001

La structure temps-événement est donc dérivée d'essais publiés, sans donnée confidentielle (figure 3). Le recours exclusif à des sources publiques rend le banc répliquable et auditable par un tiers ; nous ne lui prêtons pas de vertu d'antériorité, qui relèverait de la date de divulgation et non de la reproductibilité. La colonne vertébrale est la survie globale, seule grandeur homogène et récupérable depuis un registre ; l'EFS composite reste intra-source. Pour l'expérience causale, s'ajoute un confondant synthétique à vérité posée, dont la loi est écrite en §5.

5. Méthodes

5.1 Les quatre dispositions

Nous décrivons quatre dispositions d'une même information, à information d'entrée identique (figure 1). Le *format large* discrétise une variable en colonnes indépendantes, sans ordre exploité. La *grille* conserve l'ordre par un indice normalisé. L'*axe continu* garde la variable comme dimension continue. La forme *hiérarchique* la structure par strate. Chaque disposition reçoit un budget d'expressivité figé et déclaré, apparié entre dispositions, degrés de liberté effectifs égaux, faute de quoi on mesurerait la capacité et non la structure. Précisons d'emblée, car la relecture l'exige à juste titre, que les expériences de résultats ne comparent empiriquement que deux de ces dispositions, les colonnes indépendantes et l'axe continu ; l'ablation du §6 ajoute une variation de degrés de liberté qui approche la grille ordonnée, mais la forme hiérarchique et une grille pleinement paramétrée ne sont pas encore évaluées. Ce comparateur intermédiaire est pourtant crucial pour séparer l'ordre de la continuité et du lissage, et son absence est une limite déclarée : en l'état, l'expérience oppose plusieurs propriétés à la fois plutôt qu'une seule.

5.2 Famille de modèle

Une seule famille est employée, un modèle de survie à temps discret ajusté par régression log-linéaire pénalisée, et une seule procédure d'apprentissage. Ce choix vise à réduire les biais inductifs concurrents de celui qu'on étudie, mais il faut en refuser l'excès inverse : cette famille n'est pas sans opinion. Elle impose une forme additive, une fonction de lien, une pénalisation, une approximation par intervalles et une manière particulière de partager ou non les paramètres. L'unité expérimentale n'est donc pas la disposition seule mais le pipeline représentationnel complet, disposition, paramétrisation, pénalisation, tête. En robustesse, un perceptron minimal à une couche cachée, capacité appariée, vérifie que l'effet n'est pas un pur artefact de linéarité. Nous ne faisons pas varier l'architecture, si bien que l'effet observé mêle la disposition et son interaction avec le modèle et la régularisation ; cette réserve est centrale, non marginale. Sur l'appariement des budgets, nous égalisons le nombre de paramètres bruts et la force de pénalité entre dispositions, la base continue recevant quelques degrés polynomiaux et le format large un nombre de colonnes équivalent. Cet appariement est paramétrique, non fonctionnel : deux représentations de même dimension n'ont ni la même classe fonctionnelle ni le même biais de lissage, et cette égalité imparfaite est une limite déclarée, non une garantie. C'est du reste le cœur même de ce que le banc étudie, et il serait naïf de prétendre l'avoir neutralisé par un simple comptage de paramètres.

5.3 Conception gelée avant le run confirmatoire

La confusion de l'expérience causale est fabriquée par nous, avec des représentations conçues par nous. L'objection est la bonne, et la parade n'est pas de la nier mais de retirer le choix des mains de l'auteur. Trois dispositifs. Les budgets d'expressivité sont gelés et hachés avant tout tirage, aucune représentation ne reçoit après coup le terme qui la sauve. La forme du confondant n'est pas choisie mais échantillonnée sur un axe de régularité, du rugueux au lisse, en incluant des formes qu'aucune disposition ne peut exprimer, de sorte que personne ne gagne par construction. Un contrôle oracle vérifie configuration par configuration que la vérité est récupérable, et les configurations où elle ne l'est pas sont écartées par une règle déclarée d'avance. Nous parlons de conception gelée avant le run confirmatoire, non de préenregistrement : un fichier haché localement avant exécution n'équivaut pas à un dépôt horodaté par un tiers, et seul un tel dépôt, avec historique de versions et séparation explicite des analyses exploratoires et confirmatoires, rendrait la garantie pleinement opposable.

5.4 Le processus générateur du confondant

La distinction qui gouverne tout est celle-ci : l'estimateur ne voit qu'une représentation du confondant, l'oracle voit le confondant. Le biais mesure l'écart entre les deux (figure 2).

Pour chaque sujet i , un confondant latent $U_i \sim \mathcal{N}(0, \Sigma_U)$, $U_i \in \mathbb{R}^d$. Le traitement dépend de U_i , ce qui crée la confusion :

$$\Pr(A_i = 1 \mid U_i) = \sigma(\alpha_0 + \gamma w_A^\top U_i), \quad \sigma(x) = (1 + e^{-x})^{-1},$$

le scalaire γ réglant la force de confusion, balayé sur une grille. La survie est posée en temps discret, dans la langue de la famille employée :

$$\lambda_i(t) = \Pr(T_i = t \mid T_i \geq t, A_i, U_i) = \sigma(\beta_0(t) + \tau A_i + g_s(U_i, t)),$$

où $\beta_0(t)$ est le log-odds de base par intervalle, calé sur la forme reconstruite, τ le coefficient de traitement du modèle à temps discret, et g_s l'effet du confondant. Précisons l'estimand, car le rapport de risque prête à confusion : τ est ici un effet conditionnel structurel, le coefficient du mécanisme générateur, non un effet marginal. Le rapport de risque n'étant pas colligeable, la comparaison $|\hat{\tau} - \tau|$ n'est licite que parce que estimand et estimateur vivent dans le même espace conditionnel. Un estimand marginal, différence de survie moyenne restreinte ou risque à horizon fixe obtenu par g-formule, serait plus robuste et fera l'objet d'une extension ; nous nous en tenons ici au coefficient structurel, cohérent avec la famille employée.

L'effet g_s n'est pas fixé mais tiré d'une famille indexée par la régularité s :

$$g_s(u, t) = \sum_{k=1}^K a_k \cos(\omega_k^\top u + \nu_k t + b_k), \quad a_k \propto (\|\omega_k\|^2 + \nu_k^2)^{-s/2}.$$

Un s grand donne une surface lisse, un s petit une surface rugueuse. Les trois régimes en sont des instances : confusion multiple ($d > 1$, Σ_U non diagonale), confusion temporelle ($\nu_k \neq 0$), non-linéarité (harmoniques élevées en u , dont la quadratique est le cas simple). Ces trois régimes couvrent trois sources fréquentes de mauvaise spécification, la dépendance entre composantes, la variation temporelle et la non-linéarité ; ils ne constituent pas une taxonomie exhaustive des écarts structurels, qui incluent aussi discontinuités, seuils, interactions de haut ordre, effets à support local ou changements de régime. De même, le balayage de régularité s parcourt un continuum de lissage, mais il ne subsume pas les différences de topologie fonctionnelle : une fonction à rupture n'est pas une spline plus ou moins lisse au sens pertinent pour tout estimateur. Nous testons donc un continuum de lissage sur trois familles de mauvaise spécification, non l'ensemble des écarts possibles.

La censure est non informative, $C_i \perp T_i \mid A_i, U_i, Y_i = \min(T_i, C_i)$, $\delta_i = \mathbb{1}(T_i \leq C_i)$, ce qui garde τ identifiable. L'estimand est τ , connu, la grandeur rapportée le biais $\hat{\tau} - \tau$.

L'oracle ajuste sur le vrai confondant dans sa classe, et une configuration n'est retenue que si

$$|\hat{\tau}_{\text{oracle}} - \tau| < \epsilon.$$

Les deux générateurs écartés à la construction en sont des instances, celui dont l'effet n'était pas identifiable et celui dont la censure devenait informative, rejetés par la règle, non au jugé. Précisons ce que récupérable recouvre, car le terme mêle plusieurs choses distinctes : identifiabilité théorique, estimabilité à effectif fini, bonne spécification de la classe oracle, convergence numérique. L'oracle ajustant dans la classe exacte du générateur, son échec signale une non-identifiabilité ou une censure informative plutôt qu'une mauvaise spécification ; nous parlons donc plus exactement d'échec du critère oracle, et le §5.6 sépare explicitement la part structurelle de la part d'échantillonnage fini.

Chaque disposition ne donne à l'estimateur qu'une représentation $R(U_i)$ sous budget gelé, et l'estimateur ajuste $\hat{\tau}_R$ dans la classe $\mathcal{H}_R$ que la disposition autorise. Le biais résiduel est alors une fonction de l'accord :

$$|\hat{\tau}_R - \tau| \text{ petit} \Leftrightarrow g_s \in \overline{\mathcal{H}_R} \text{ (à la résolution du budget).}$$

C'est la prédiction a priori du banc, non un fait établi : la disposition ne devrait réduire le biais causal que lorsque la classe fonctionnelle qu'elle autorise contient la structure du confondant. Les résultats du §6 disent dans quelle mesure cette prédiction se vérifie, et la réponse, on le verra, est plus prudente que la prédiction.

5.5 Métriques

Pour la fidélité, l'erreur quadratique intégrée à la courbe, complétée de l'erreur de survie moyenne restreinte et de la calibration intégrée, toutes rapportées et cohérentes entre elles. Pour l'estimation sous confusion simulée, nous rapportons le biais signé, l'erreur absolue et surtout la différence appariée d'erreur entre représentations pour le coefficient conditionnel τ . La concordance et le score de Brier sont écartés, sans objet pour un coefficient structurel sur une reconstruction à deux bras.

5.6 Filtre de récupérabilité, grille structurelle et régime d'événements

Le filtre oracle de la conception initiale confondait deux choses distinctes, la récupérabilité de la vérité, qui est une propriété du générateur, et le bruit d'échantillonnage de l'estimateur oracle en effectif fini. Un diagnostic le montre : sous le régime initial l'oracle est sans biais en moyenne, de l'ordre de zéro, mais son écart-type d'échantillonnage vaut environ 0,10, très supérieur au seuil ϵ de 0,05. Le filtre par graine rejetait donc les deux tiers des configurations pour du bruit et non pour une non-récupérabilité, sous-alimentant les cellules causales. Nous corrigeons par deux gestes déclarés et rehachés, motivés par ce diagnostic.

D'abord, la grille structurelle est découplée de l'estimation. La récupérabilité est jugée une seule fois par configuration de générateur, sur un grand échantillon où l'écart-type de l'oracle devient petit, et son seuil est fixé à trois fois cet écart-type résiduel, de sorte qu'elle n'écarte que les configurations dont le biais oracle persiste à grand échantillon, les structurellement non identifiables, et non celles qui ne s'écartent que par échantillonnage. L'expérience à effectif fini tourne ensuite sur toutes les graines des configurations admises, sans refiltrage. La grille ne regarde que le générateur, jamais laquelle des dispositions l'emporte, ce qui garantit qu'elle relève la puissance sans biaiser le contraste, c'est la ligne entre corriger la puissance et régler l'issue.

Ensuite, le régime est calé sur davantage d'événements, hasard de base relevé et censure allégée, la puissance en survie tenant au nombre d'événements et non à l'effectif brut. L'écart-type de l'oracle passe ainsi de 0,10 à 0,07. La combinaison rétablit une rétention pleine, dix-huit à vingt graines sur vingt, et les estimations causales cessent d'être bruitées.

Pour la traçabilité, le tableau des déviations au protocole initial.

Élément	Protocole initial	Amendement	Justification
Filtre oracle	par graine	par configuration à grand N	à séparer bruit et non-récupérabilité

Élément	Protocole initial	Amendement	Justification
Régime d'événements	initial	hasard relevé, censure allégée	abaisser la variance de l'estimateur
Seuil de récupérabilité	de fixe (0,05)	3σ résiduel	contrôler le faux rejet

L'amendement a été diagnostiqué après une première exécution jugée trop instable ; il est donc informé par cette exécution, tout en restant aveugle à la disposition gagnante. Ce statut d'amendement confirmatoire, et non de protocole initial, est déclaré comme tel. Le régime enrichi en événements modifie le monde simulé, aussi avons-nous vérifié que le verdict n'en dépend pas. En faisant varier le hasard de base sous grille oracle et mêmes graines que l'analyse principale, du régime riche, environ quinze pour cent d'événements par intervalle, au régime pauvre, environ huit pour cent, la probabilité que le continu batte le plat reste entre 0,46 et 0,61 et la différence appariée demeure dans une marge de six millièmes, tous les intervalles traversant zéro. Le résultat nul n'est donc pas un artefact du régime d'événements amendé. Une grille complète croisant événements, censure et force de confusion resterait souhaitable pour une revue exigeante.

5.7 Détail du plan factoriel

Le plan factoriel est décrit ici pour être reconstruit sans le code. La représentation continue est une spline pénalisée, base B-spline cubique à support local assortie d'une pénalité sur les différences secondes de ses coefficients, un P-spline ; un polynôme global de degré cinq, initialement employé et mécaniquement défavorable sur les ruptures, a été écarté au profit de cette vraie spline. La représentation discrète est une indicatrice par intervalle assortie d'une pénalité sur les différences secondes des log-odds voisins, sans base continue ; la grille est un indice ordonné linéaire. Les degrés de liberté effectifs sont la trace de l'opérateur de lissage, $\text{tr}[(X^T W X + \lambda P)^{-1} X^T W X]$, évaluée à la solution pénalisée obtenue par moindres carrés repondérés itérés. L'appariement fixe une cible de degrés de liberté et ajuste λ par recherche dichotomique pour l'atteindre. L'estimation se fait sur un ensemble d'apprentissage, l'évaluation sur un ensemble de test disjoint à parts égales, sur vingt-cinq graines ; la vérité lisse est un hasard en sinus carré, la vérité à rupture un plateau suivi d'un saut. Le contraste discret contre spline est apparié par graine et son intervalle à quatre-vingt-quinze pour cent est un bootstrap apparié à deux mille répliques, cohérent avec le bras sous confusion. L'appariement par la trace égalise une complexité effective globale, non la géométrie complète des opérateurs, deux opérateurs de même trace pouvant différer par leur spectre ou leur comportement aux frontières ; nous parlons donc de complexité effective globale comparable, non de flexibilité strictement égale. Le design est ainsi une ablation contrôlée à complexité effective appariée, non un plan factoriel complètement croisé, et nous ne le surqualifions pas.

6. Résultats

La logique d'ensemble du banc, de l'effet brut au principe méthodologique, est résumée en figure R0, que le lecteur pressé peut lire seule. Les résultats sont rapportés avec leur incertitude, et volontairement lus à la baisse. Le run confirmatoire tourne à rétention pleine, quatre-vingt-dix à quatre-vingt-dix-huit pour cent des configurations admises. Cette forte rétention réduit le risque d'une sélection substantielle, sans l'exclure, et nous en rapportons donc la distribution : les taux de rejet oracle vont de zéro à cinq pour cent selon la régularité et le régime, à une exception près, la confusion temporelle, où ils atteignent quinze pour cent. C'est la seule cellule où une sélection résiduelle mérite prudence, et nous l'y signalons ; ailleurs le faible taux de rejet limite, sans l'exclure, le risque d'orientation.

Fidélité (figures R1, R4 et R5). Représenter le temps par une base partagée plutôt que par des colonnes indépendantes réduit modestement l'erreur de reconstruction, l'écart augmentant avec le nombre d'intervalles, là où les colonnes indépendantes, une par intervalle, surajustent, ce qu'objective un écart entre déviance d'apprentissage et de test. Une ablation contrôlée à complexité effective globale appariée en précise la nature. Nous comparons, à même complexité effective mesurée par la trace de l'opérateur de lissage, une représentation discrète à pénalité de différences secondes, qui partage l'information entre intervalles voisins sans base continue, et une vraie spline pénalisée à support local. Le choix du comparateur importe, et nous le corrigeons : un polynôme global, initialement employé, est mécaniquement désavantagé sur une rupture et exagérerait l'écart. Avec une spline pénalisée, les différences d'ISE tombent d'un ordre de grandeur, autour du millionième, et le contraste apparié par bootstrap ne montre aucune supériorité de la représentation continue : sur vérité lisse à six degrés de liberté la différence est nulle, ailleurs elle favorise le discret de façon infime. La survie moyenne restreinte, la calibration intégrée et la déviance test sont identiques entre les deux. La formulation exacte est donc plus prudente que « se valent » : à complexité effective globale comparable, aucune supériorité propre de la continuité n'est observée, et les écarts résiduels, faibles et souvent orientés vers le discret, sont sans pertinence pratique. L'effet de fidélité relève du compromis biais-variance et de la régularisation, non d'une géométrie de la continuité. C'est le résultat le plus net du banc, et il faut se garder de remplacer une doctrine pro-continu par un miroir pro-discret : les deux se comportent, à complexité appariée, de la même manière.

Le tableau donne les quatre métriques préspecifiées, cohérentes entre elles, par représentation et vérité, la représentation continue étant la spline pénalisée.

Représentation (vérité)	df eff	ISE	RMST	calibration	déviance test
colonnes indép. (lisse)	48,0	$4,2 \cdot 10^{-4}$	$9,3 \cdot 10^{-3}$	$1,5 \cdot 10^{-2}$	0,259
grille df=2 (lisse)	2,0	$1,3 \cdot 10^{-3}$	$1,3 \cdot 10^{-2}$	$2,8 \cdot 10^{-2}$	0,238
discret fused df=6 (lisse)	6,0	$3,5 \cdot 10^{-4}$	$9,6 \cdot 10^{-3}$	$1,4 \cdot 10^{-2}$	0,233
spline pénalisée df=6 (lisse)	6,0	$3,5 \cdot 10^{-4}$	$9,8 \cdot 10^{-3}$	$1,4 \cdot 10^{-2}$	0,233
colonnes indép. (rupture)	47,0	$5,8 \cdot 10^{-4}$	$1,1 \cdot 10^{-2}$	$1,9 \cdot 10^{-2}$	0,230
grille df=2 (rupture)	2,0	$1,6 \cdot 10^{-3}$	$1,4 \cdot 10^{-2}$	$2,9 \cdot 10^{-2}$	0,205
discret fused df=6 (rupture)	6,0	$5,5 \cdot 10^{-4}$	$1,2 \cdot 10^{-2}$	$1,8 \cdot 10^{-2}$	0,205

Représentation (vérité)	df eff	ISE	RMST	calibration	déviance test
spline pénalisée df=6 (rupture)	6,0	$5,6 \cdot 10^{-4}$	$1,2 \cdot 10^{-2}$	$1,8 \cdot 10^{-2}$	0,205

Le tableau porte aussi la leçon substantielle : les écarts d'ISE entre discret et spline, statistiquement résolus mais de l'ordre du millionième, sont invisibles sur la RMST, la calibration et la déviance. La significativité statistique ne vaut pas pertinence pratique, et aucune marge de pertinence clinique préspecifiée ne serait franchie.

Erreur d'estimation de l'effet structurel sous confusion simulée (figures R2 et R3). Nous renommons ce bras, car il n'estime pas un effet causal clinique mais l'erreur de récupération d'un coefficient conditionnel dans un monde simulé. L'unité rapportée est la différence appariée entre dispositions sur les mêmes graines, rééchantillonnée par un bootstrap apparié par graine, l'unité de rééchantillonnage étant la graine et non le sujet, deux mille réplifications, intervalles percentile. Le verdict est nul, et il est désormais quantifié plutôt qu'affirmé. Aucune des huit cellules ne survit à une correction de Holm, la plus petite valeur p valant 0,038 pour un seuil corrigé de 0,006 ; les tests des signes appariés sont tous non significatifs ; la probabilité que le continu batte le plat reste entre 0,43 et 0,59. La cellule $s = 2$, marginale avant correction, ne résiste ni à Holm ni au test des signes : c'est un faux positif de multiplicité. Nous caractérisons aussi ce que ce nul recouvre, faute de quoi il serait surinterprété. La différence minimale détectable à quatre-vingts pour cent de puissance est d'environ 0,022 en log-odds, et les données sont compatibles avec l'absence de tout écart supérieur à 0,033, cette valeur étant la plus large borne d'intervalle de confiance à quatre-vingt-quinze pour cent observée sur les cellules, exprimée en log-odds par intervalle, dont l'interprétation pratique est un décalage négligeable du hasard conditionnel, soit, sur un hasard de base de cinq à vingt pour cent, un écart absolu de l'ordre de deux à cinq dixièmes de point de pourcentage. C'est donc un nul dans une marge, non une équivalence prouvée : un effet plus fin que la différence détectable ne peut être exclu, et l'incertitude rapportée est celle du Monte-Carlo conditionnel au banc, non celle du transport à d'autres essais. Ce nul est stable au régime d'événements, la sensibilité du §5.6 le confirmant du hasard riche au pauvre, mais son niveau ponctuel reste sensible au tirage des configurations, si bien qu'il se lit au sens de la correction de multiplicité et non comme point stable. Un contrôle sous générateur externe, un temps de vie accéléré de Weibull hors de la famille d'estimation, en éclaire la portée : quand les représentations ne sont pas appariées en complexité, la représentation riche retire une confusion résiduelle que la représentation grossière laisse, et l'écart est alors net et systématique ; mais c'est un effet de classe fonctionnelle et de capacité, non de continuité, qui confirme la déconfusion plutôt qu'un avantage propre du format. Un contrôle externe à complexité appariée reste à conduire. Le plan d'hypothèses est déclaré et hiérarchique : contraste principal intégré sur s , puis test global représentation contre régularité, puis contrastes secondaires par régime, enfin analyses exploratoires cellule par cellule, la correction de Holm ne s'appliquant qu'à ce dernier niveau exploratoire.

Réplication sur générateurs indépendants (figure R6). Pour répondre au reproche du générateur endogène, nous appliquons le même protocole à trois familles de générateurs structurellement indépendantes : un modèle de Cox à risques proportionnels, un temps de vie accéléré de Weibull, et le modèle logistique discret. Dans chacune, un confondant lisse agit sur le traitement et sur l'issue, et nous mesurons l'erreur de récupération du coefficient

de traitement lorsque ce confondant est représenté de trois façons à degrés de liberté effectifs appariés, colonnes indépendantes, discret pénalisé, spline continue, la référence étant l'ajustement sur le confondant vrai. Le résultat est le même d'un générateur à l'autre, ce qui est l'essentiel : l'ordre des représentations est conservé, la spline continue récupère l'effet au moins aussi bien que le discret, mieux que les colonnes indépendantes, sur Cox comme sur Weibull comme sur le logistique discret. La robustesse externe du verdict qualitatif est donc établie hors de la famille d'estimation.

Deux nuances que l'honnêteté impose. D'abord, contrairement au résultat sur la représentation du temps, où discret et continu se valaient, la spline garde ici un léger avantage pour un confondant lisse. L'explication est instructive et conforme au principe du papier : l'appariement par la trace n'égalise pas la qualité d'approximation d'une cible lisse par une base en marches. Ensuite, cet avantage rétrécit quand on affine le grain de la représentation discrète, de vingt à quatre-vingt-seize bins l'écart se réduit de moitié à deux tiers sans disparaître, ce qui montre que le grain discret est un levier distinct des degrés de liberté effectifs. La leçon n'est donc pas qu'une représentation gagne universellement, mais qu'une comparaison de représentations doit contrôler à la fois la complexité effective et la classe fonctionnelle, faute de quoi l'écart attribué au format mêle les deux. C'est exactement la contribution que le banc défend. Insistons pour lever une apparence de contradiction : R6 ne contredit pas R5. R5 porte sur la représentation du temps, où le grain était déjà fin et où discret et continu s'égalisaient ; R6 porte sur la représentation d'un confondant lisse, où un grain discret grossier approxime moins bien qu'une spline. Ce sont deux questions expérimentales distinctes, et dans les deux l'écart, quand il existe, s'explique par la complexité effective et le grain, non par une vertu propre de la continuité.

Lecture d'ensemble. Le banc établit un effet de fidélité modeste, de nature régularisatrice, et ne détecte aucune différence robuste d'estimation sous confusion simulée. Le résultat de fidélité, indépendant de la confusion synthétique, porte la charge de preuve, mais le plan factoriel en réduit la lecture à un phénomène de degrés de liberté effectifs plutôt que de continuité. Le résultat nul réduit la crainte d'une construction produisant mécaniquement la supériorité attendue, sans constituer une preuve de neutralité du banc, car un nul est aussi compatible avec deux méthodes également bien spécifiées, une métrique bruitée, une puissance insuffisante ou un DGP peu discriminant. Une montée en effectif, et surtout en contextes cliniques indépendants, pourrait révéler un effet aujourd'hui sous la différence détectable, mais rien dans les données présentes ne permet de l'affirmer.

Augmentation. Piste réservée, non exécutée, présentée en perspective au §7.2.

7. Discussion

7.1 Face aux méthodes statistiques classiques

Les méthodes classiques ne sont pas l'extérieur de la thèse, mais il faut se garder d'une extension excessive du vocabulaire. Un modèle de Cox à risques proportionnels [23] n'est pas une tensorisation ; il impose une contrainte structurelle, la proportionnalité des risques et une forme additive sur les covariables. Kaplan-Meier [24] repose sur une censure

indépendante et une homogénéité du mécanisme de survie dans les strates. Fine-Gray spécifie un modèle pour le hasard de sous-distribution. Ce sont des hypothèses de modèle, et parfois des invariances, non nécessairement des symétries de groupe. Le lien avec notre cadre est plus modeste ainsi, et plus juste : chacune fixe d'avance une structure et une classe de transformations compatibles, là où le banc traite cette structure comme une variable. Le danger, que nous nommons pour l'éviter, est celui d'un concept si extensif qu'il engloberait toute hypothèse statistique et cesserait alors de discriminer quoi que ce soit. Le message opérationnel est méthodologique et non un résultat de cette étude, que le banc n'a pas testée sur un générateur strictement proportionnel : lorsqu'une méthode classique fixe déjà la bonne structure, horloge unique et effet proportionnel pour Cox, le choix de disposition ne devrait guère peser ; c'est quand la structure de l'effet s'écarte de l'hypothèse posée que la disposition redevient un levier potentiel, sous la réserve qu'ajoutent nos résultats, aucun avantage propre n'ayant été détecté au-delà de la régularisation.

7.2 Perspectives

Un prolongement est réservé, sans résultat à ce stade : une augmentation de données respectant la structure déclarée pourrait aider davantage qu'une augmentation aveugle, mais seulement là où cette structure coïncide avec celle de l'effet ; nous ne l'exécutons pas ici.

7.3 Application

La conséquence applicative possible concerne les bras synthétiques et comparaisons indirectes : la disposition pourrait influencer la fidélité du jumeau à la distribution jointe des covariables et la stabilité de la repondération. Elle n'est pas testée ici, aucune mesure de balance, d'overlap, de taille d'échantillon effective ou de biais de MAIC n'étant produite, et une expérience dédiée serait nécessaire avant toute affirmation.

7.4 Biais de simulation et comparaison neutre

Une simulation comparative mal bridée permet de revendiquer la supériorité de presque n'importe quelle méthode [25], et c'est l'objection de fond adressée à tout banc dont l'auteur fabrique le monde. Nous y répondons par le design plutôt que par la protestation. Le plan suit le standard ADEMP des études de simulation [26] et se veut une comparaison neutre au sens de Boulesteix et collègues [27]. Les garde-fous sont le protocole gelé et haché avant tirage, la grille de régularité préspecifiée de sorte qu'aucune disposition ne l'exprime par construction, la grille de récupérabilité oracle découplée de l'estimation et aveugle à la disposition gagnante. Nous ne prétendons pas que ces garde-fous prouvent la neutralité : un générateur peut favoriser une méthode de façon dépendante du régime si les régimes eux-mêmes sont choisis pour coïncider avec la théorie, et la dépendance au régime n'est donc pas une preuve interne de neutralité. La neutralité ne s'établit que par le protocole gelé, les comparateurs neutres, les ablations, un ancrage plasmode sur covariables réelles [28] et, à terme, une réplication externe. Que le résultat causal soit ici nul, aucune disposition n'étant favorisée, est du reste cohérent avec un banc qui ne penche pas.

8. Limites

Le domaine de validité est explicite et il borne les conclusions, plus étroitement que les versions antérieures ne le laissaient croire. Une seule famille de modèle est employée, et l'unité réellement testée est le pipeline représentationnel complet, non la disposition isolée ; l'effet observé mêle disposition, paramétrisation et pénalisation, et leur séparation exigerait plusieurs architectures témoins, Cox, modèle discret plat, spline pénalisée, gradient boosting, perceptron minimal. L'appariement des budgets est paramétrique et non fonctionnel, deux représentations de même dimension n'ayant ni la même classe fonctionnelle ni le même biais de lissage. Le mécanisme générateur causal est écrit dans la langue de la famille d'estimation, ce qui favorise les représentations compatibles avec cette paramétrisation ; une validation robuste emploierait plusieurs générateurs, Cox, temps de vie accéléré, hasards non proportionnels ou non monotones, un simulateur indépendant de la famille estimée, voire un ancrage semi-réel sur covariables réelles. L'estimand est un coefficient conditionnel non colligeable, et une différence de survie moyenne restreinte serait plus robuste. L'expérience de fidélité de premier ordre, R1, et la reconstruction Guyot ne disposent pas encore de la séparation apprentissage-test franche dont bénéficie déjà le plan factoriel R5, ni d'une analyse de sensibilité à la numérisation des courbes et aux tables d'effectifs à risque. La validation de la reconstruction par proximité du rapport de risque valide un scalaire, non la structure complète, temps de censure, non-proportionnalité, survie moyenne restreinte. Le support empirique se réduit à deux essais, une pathologie en immunothérapie cellulaire. Enfin le résultat causal est nul à ce niveau d'effectif : nous ne pouvons exclure un effet fin, mais nous ne pouvons pas davantage l'affirmer. Ces limites ne fragilisent pas la thèse théorique sur la représentation, elles bornent sévèrement ce que la vérification empirique permet de conclure, et cette borne est déclarée plutôt que tue.

Note : Disponibilité du code et reproductibilité

Le banc est reproductible de bout en bout à partir de sources publiques. La reconstruction opère par `guyot.py` sur les courbes de Kaplan-Meier publiées de TRANSFORM et ZUMA-7. La conception gelée avant le run confirmatoire, budgets d'expressivité appariés, famille de DGP paramétrée en régularité, filtre oracle, est fixée dans un protocole haché avant tout tirage, de sorte que ni les budgets ni les formes ne peuvent être ajustés après avoir vu les données ; ce hachage local ne vaut cependant pas dépôt horodaté par un tiers. Le run confirmatoire applique l'amendement du filtre décrit en §5.6, grille de récupérabilité découplée de l'estimation et régime à plus d'événements, et rapporte l'incertitude par différences appariées, probabilités et intervalles bootstrap. Aucune donnée individuelle n'est mobilisée, seules les courbes publiées et leurs tables d'effectifs à risque sont en entrée, ce qui rend le banc répliquable et auditable par un tiers.

Figures

De l'effet apparent au principe : la logique du banc

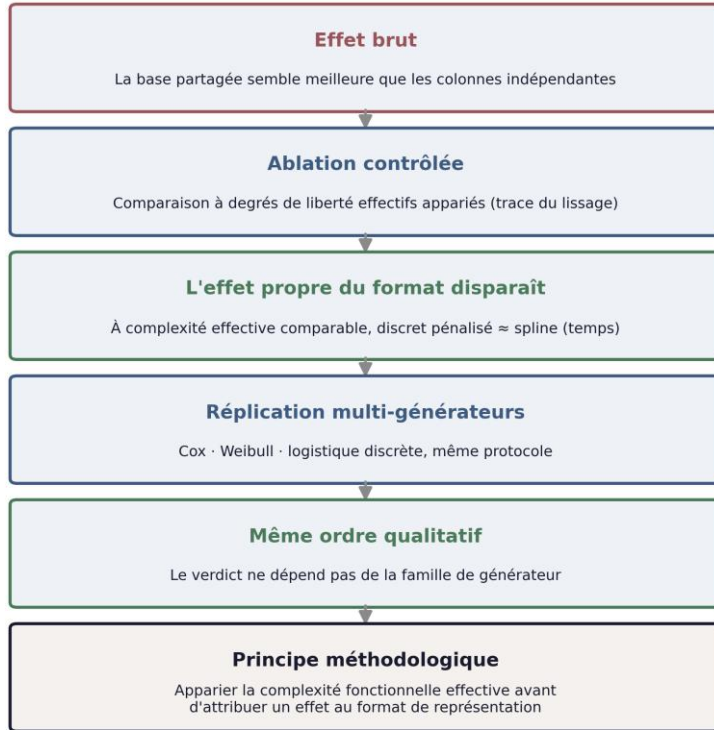


Figure R0. Synthèse de la logique du banc : de l'effet brut au principe méthodologique, en passant par l'ablation, l'appariement de complexité, la disparition de l'effet et la validation multi-générateurs.



Figure 1. Les structures rendues disponibles par quatre dispositions d'une même information ; les libellés décrivent la structure de l'espace d'entrée, non une invariance imposée.

Graphe causal du confondant : U agit sur le traitement et sur le hasard

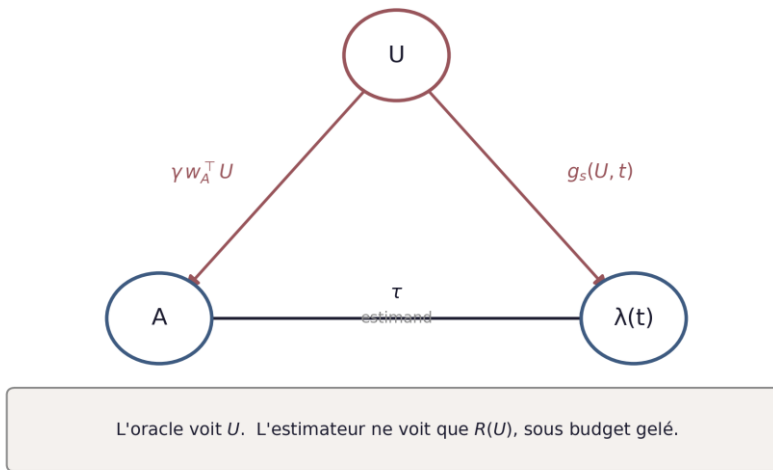


Figure 2. Graphe causal du confondant, U agissant sur le traitement et sur le hasard, l'oracle voyant U et l'estimateur $R(U)$.

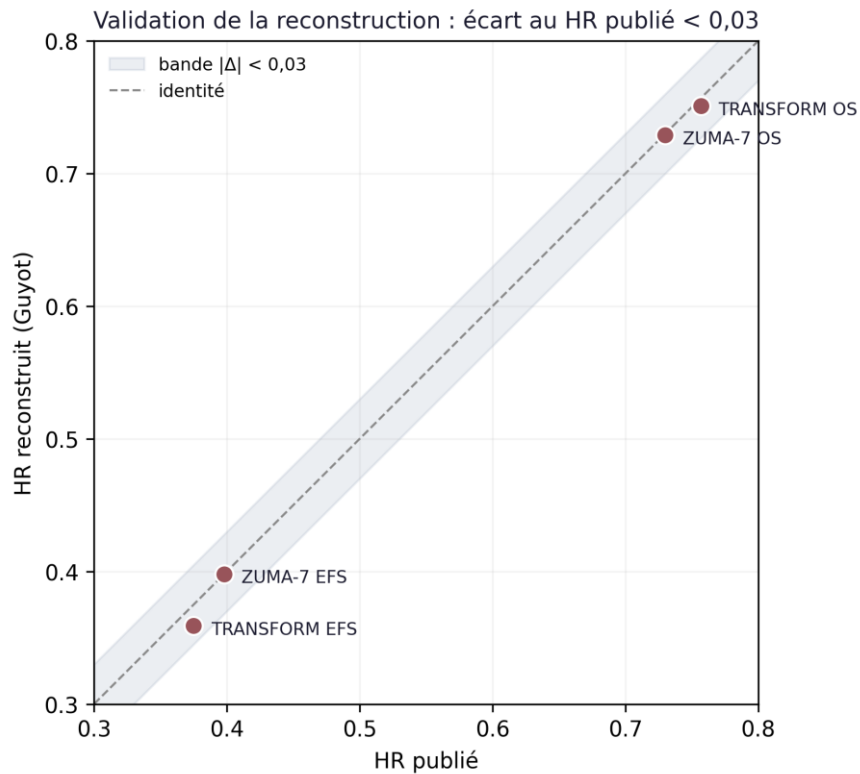


Figure 3. Validation du HR sur les données reconstruites, écart au rapport de risque publié inférieur à trois centièmes pour les quatre courbes ; validation d'un scalaire par courbe, non de la structure complète.

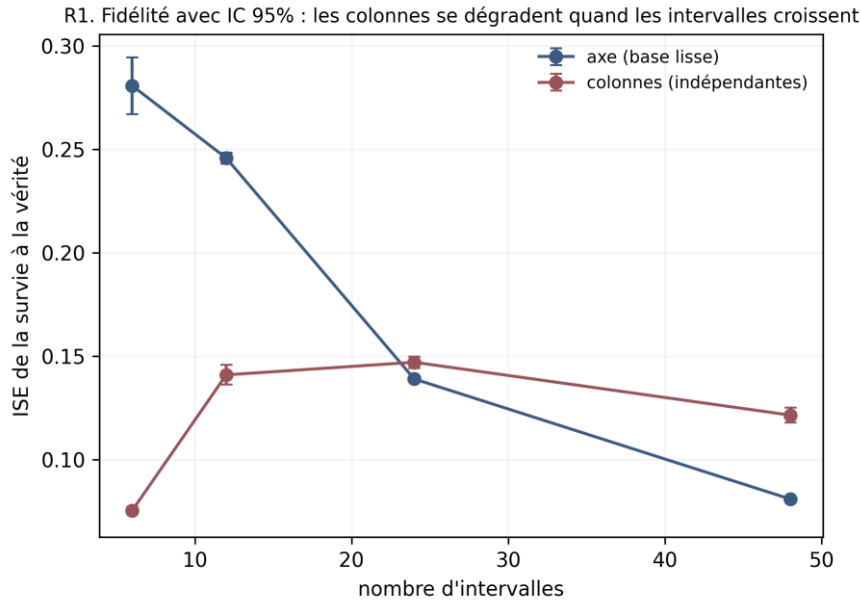


Figure R1. Fidélité : à pénalisation fixée, l'erreur des colonnes indépendantes augmente relativement à la base partagée lorsque la granularité croît, intervalles de confiance à 95%. R1 mesure l'effet de la granularité à paramétrisation donnée, R5 celui de la flexibilité à granularité fixée.

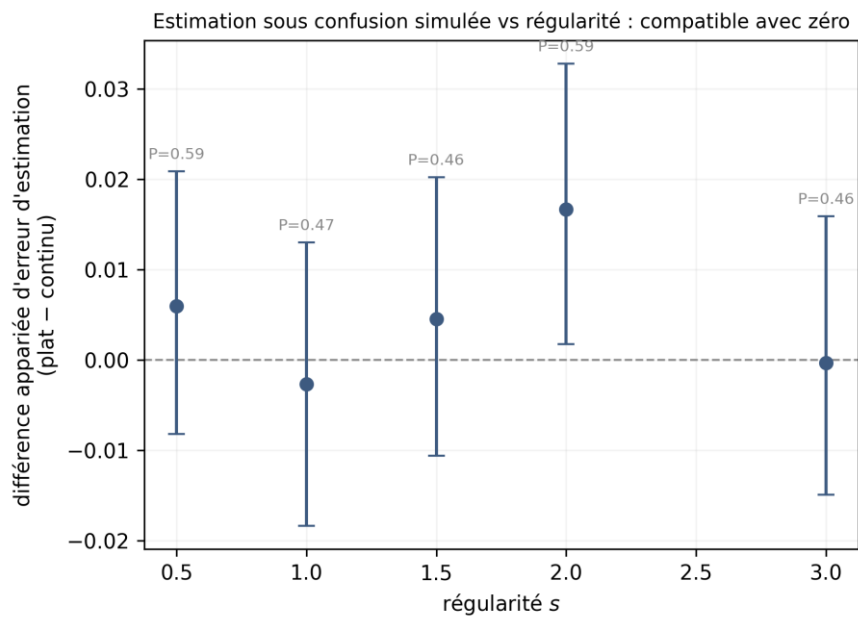


Figure R2. Différence appariée d'erreur d'estimation du coefficient structurel, plat moins continu, en fonction de la régularité s, IC bootstrap à 95%, compatible avec zéro.

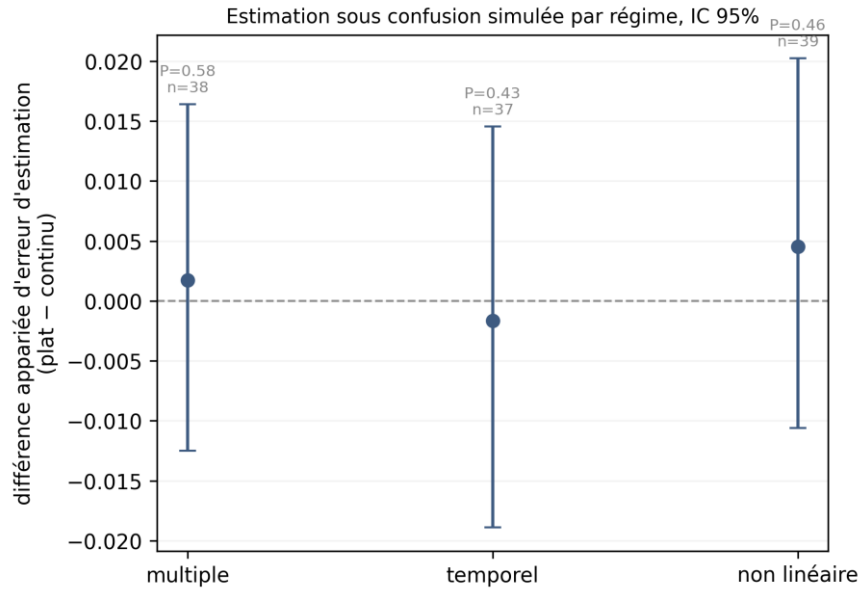


Figure R3. Différence appariée d'erreur d'estimation du coefficient structurel par régime, IC à 95%, effectif et probabilité annotés.

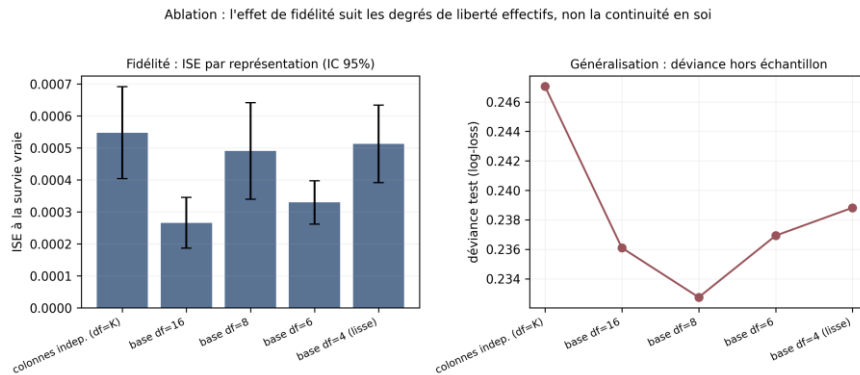


Figure R4. Ablation de fidélité, ISE et déviance hors échantillon selon les degrés de liberté de la représentation temporelle, montrant que l'effet suit la flexibilité effective et non la continuité en soi.

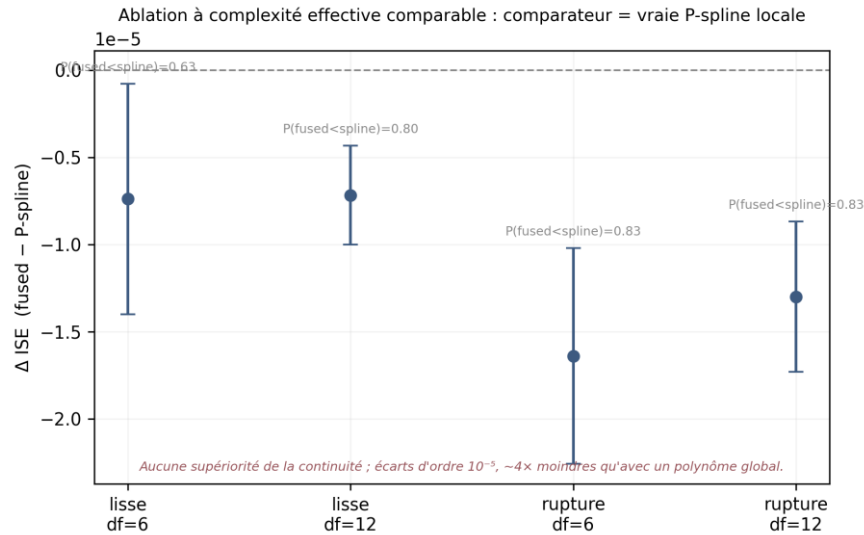


Figure R5. Ablation contrôlée à complexité effective globale comparable, contraste apparié discret contre spline pénalisée sur vérité lisse et à rupture : aucune supériorité propre de la représentation continue, les écarts résiduels étant d'un ordre de grandeur inférieur à ceux d'un polynôme global.

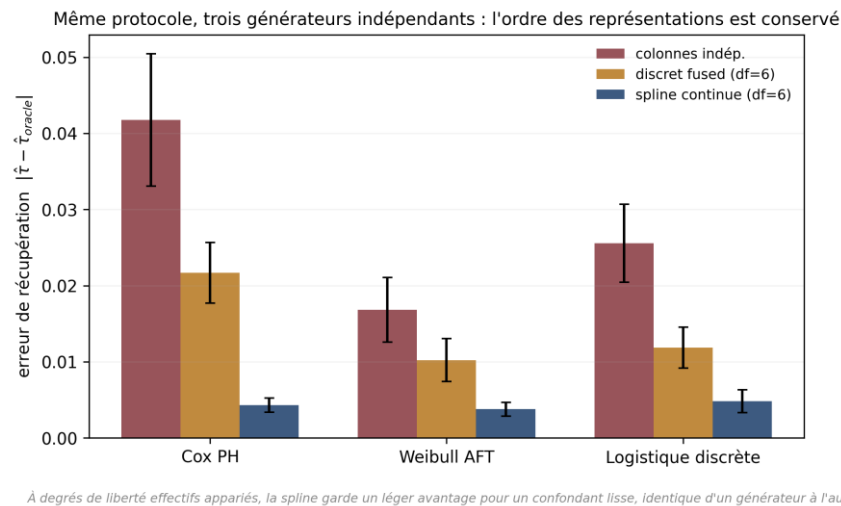


Figure R6. Même protocole sur trois générateurs indépendants, Cox à risques proportionnels, Weibull à temps accéléré, logistique discret : erreur de récupération du coefficient de traitement par représentation du confondant à degrés de liberté effectifs appariés. L'ordre des représentations est conservé d'un générateur à l'autre.

9. Portée de l'inférence

Plutôt qu'une liste défensive, nous bornons explicitement chaque affirmation par sa preuve et sa limite. Ce tableau tient lieu de conclusion : il fixe ce que le banc autorise à conclure, et rien de plus.

Affirmation	Preuve dans ce banc	Limite
Le choix de représentation et de pénalisation modifie la classe fonctionnelle accessible	ablation factorielle, fidélité	attribution conditionnelle au pipeline
Un partage de paramètres réduit modestement l'erreur de reconstruction à grille fine	R1, R4	effet petit, mécanisme de df effectifs
Aucun effet propre de la continuité détecté	R5, le discret pénalisé égale ou dépasse le continu à df apparié	configurations lisse et à rupture seulement
Le verdict qualitatif est robuste à la famille de générateur	R6, trois générateurs indépendants, même ordre	avantage spline modeste pour confondant lisse, dépendant du grain discret
Aucune différence d'estimation sous confusion	R2, R3, Holm, signes	nul dans une marge de 0,033, non équivalence
La disposition est une hypothèse d'invariance au sens fort	aucune	non établi, formule doctrinale
Supériorité de l'axe continu, réduction du biais causal	aucune	non soutenu par les résultats, non détecté dans la marge
Généralisation à d'autres pathologies, DGP, architectures	aucune	hors champ
Amélioration d'une MAIC ou d'un bras synthétique	aucune	hypothèse de transfert non testée

La nouveauté n'est pas la disposition comme hypothèse d'invariance, thèse que ce banc ne soutient pas, mais une déconfusion méthodologique : montrer que des effets attribués à la disposition peuvent être confondus avec la régularisation fonctionnelle, et proposer une décomposition entre représentation, classe fonctionnelle, architecture, contrainte et complexité effective. La formulation défendable est donc celle-ci : à information nominale identique, le pipeline de représentation modifie la classe fonctionnelle et la régularisation accessibles à l'estimateur ; dans ce banc de survie, l'avantage apparent d'une représentation continue est expliqué par les degrés de liberté effectifs plutôt que par la continuité elle-même, et aucune différence robuste d'estimation n'est détectée sous confusion simulée. De là un principe de design, plus transportable que le résultat local sans prétendre que celui-ci se généralise : toute comparaison de représentations doit apparié ou caractériser explicitement la complexité fonctionnelle effective et la pénalisation induite, faute de quoi l'effet attribué au format peut n'être qu'un effet de régularisation déguisé. Ce principe déborde la survie. Il vaut partout où l'on oppose des représentations à capacité inégale, données tabulaires, texte, image, graphes, et c'est probablement la portée la plus large du travail : non pas une propriété de la continuité, mais une exigence de contrôle expérimental pour toute comparaison de formats. Nous l'énonçons ici sans prétendre le démontrer au-delà du banc de survie, mais nous ne le sous-vendons pas non plus. Deux réserves de portée que

l'honnêteté impose. D'abord, quarante ou mille graines autour de quatre courbes issues de deux essais ne font pas quarante ou mille contextes cliniques ; l'incertitude rapportée est simulationnelle, non un transport empirique, et la pseudo-réplication conceptuelle demeure. Ensuite, la neutralité du banc n'est pas prouvée par le protocole gelé seul : les représentations, régimes et métriques ont été définis selon la théorie, et la validation forte, générateurs construits par une équipe indépendante, jeu de challenge externe, plasmods sur covariables réelles, réplication par une seconde implémentation, analyse en aveugle des libellés, n'est pas une extension optionnelle mais une condition de validité externe. Le passage à une revue méthodologique exigeante suppose ce programme, plus l'ajout de plusieurs générateurs indépendants de la famille estimée, une grille de taux d'événements et de censure, un estimand marginal de type survie moyenne restreinte, et l'évaluation des dispositions grille et hiérarchique.

Références

1. Mitchell TM. The need for biases in learning generalizations. Rutgers University Technical Report CBM-TR-117; 1980.
2. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*. 2013;35(8):1798-1828.
3. Cohen T, Welling M. Group equivariant convolutional networks. In: *Proceedings of the International Conference on Machine Learning (ICML)*; 2016.
4. Battaglia PW, Hamrick JB, Bapst V, et al. Relational inductive biases, deep learning, and graph networks. *arXiv:1806.01261*; 2018.
5. Bronstein MM, Bruna J, Cohen T, Velicković P. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478*; 2021.
6. Katzman JL, Shaham U, Cloninger A, Bates J, Jiang T, Kluger Y. DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC Med Res Methodol*. 2018;18:24.
7. Lee C, Zame WR, Yoon J, van der Schaar M. DeepHit: a deep learning approach to survival analysis with competing risks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*; 2018.
8. Lee C, Yoon J, van der Schaar M. Dynamic-DeepHit: a deep learning approach for dynamic survival analysis with competing risks based on longitudinal data. *IEEE Trans Biomed Eng*. 2020;67(1):122-133.
9. Kvamme H, Borgan Ø, Scheel I. Time-to-event prediction with neural networks and Cox regression. *J Mach Learn Res*. 2019;20(129):1-30.
10. Bender A, Rügamer D, Scheipl F, Bischl B. A general machine learning framework for survival analysis. In: *Proceedings of ECML PKDD*. Springer; 2021.
11. Huang X, Khetan A, Cvitkovic M, Karnin Z. TabTransformer: tabular data modeling using contextual embeddings. *arXiv:2012.06678*; 2020.
12. Gorishniy Y, Rubachev I, Khrulkov V, Babenko A. Revisiting deep learning models for tabular data. *Adv Neural Inf Process Syst*. 2021;34:18932-18943.

13. Somepalli G, Goldblum M, Schwarzschild A, Bruss CB, Goldstein T. SAINT: improved neural networks for tabular data via row attention and contrastive pre-training. arXiv:2106.01342; 2021.
14. Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? Adv Neural Inf Process Syst; 2022.
15. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol. 2012;12:9.
16. Signorovitch JE, Wu EQ, Yu AP, et al. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons. Pharmacoeconomics. 2010;28(10):935-945.
17. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. NICE DSU Technical Support Document 18: methods for population-adjusted indirect comparisons in submissions to NICE. Sheffield: NICE Decision Support Unit; 2016.
18. Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. Med Decis Making. 2018;38(2):200-211.
19. Wood SN. Generalized additive models: an introduction with R. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2017.
20. Hastie T, Tibshirani R. Varying-coefficient models. J R Stat Soc Series B. 1993;55(4):757-796.
21. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. J R Stat Soc Series B. 2005;67(1):91-108.
22. Kim SJ, Koh K, Boyd S, Gorinevsky D. l1 trend filtering. SIAM Rev. 2009;51(2):339-360.
23. Cox DR. Regression models and life-tables. J R Stat Soc Series B. 1972;34(2):187-220.
24. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53(282):457-481.
25. Pawel S, Kook L, Reeve K. Pitfalls and potentials in simulation studies: questionable research practices in comparative simulation studies allow for spurious claims of superiority of any method. Biom J. 2024;66(1):e2200091.
26. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019;38(11):2074-2102.
27. Boulesteix AL, Lauer S, Eugster MJ. A plea for neutral comparison studies in computational sciences. PLoS One. 2013;8(4):e61562.
28. Schreck N, Slynko A, Saadati M, Benner A. Statistical plasmode simulations: potentials, challenges and recommendations. Stat Med. 2024;43(9):1804-1825.