

Appariement des données d'accès précoce au Système National des Données de Santé (SNDS) via la Cohorte Cancer, en France : évaluation des performances et des biais potentiels

I. Charkaoui¹, V.P. Luu², M. Licour³, M. Narbeburu¹, C. Le Bihan-Benjamin¹, N. Ozan⁴
¹ Institut National du Cancer, ² Filière Intelligence Artificielle et Cancers, ³ OncoReal, ⁴ AstraZeneca

Appariement des données d'accès précoce au Système National des Données de Santé (SNDS) via la Cohorte Cancer, en France : évaluation des performances et des biais potentiels

I. Charkaoui¹, V.P. Luu², M. Licour³, M. Narbeburu¹, C. Le Bihan-Benjamin¹, N. Ozan⁴
¹ Institut National du Cancer, ² Filière Intelligence Artificielle et Cancers, ³ OncoReal, ⁴ AstraZeneca

CONTEXTE

- Les programmes d'accès précoce (AP) sont une opportunité pour décrire la première utilisation de traitements innovants en vie réelle.
- La Haute Autorité de Santé (HAS) rend obligatoire la collecte des données cliniques des patients pour les AP et recommande leur appariement au Système National des Données de Santé (SNDS).
- Les données d'AP collectées par les industriels en vertu de l'obligation de traitement médicamenteux sont quasi-exhaustives, mais les données de suivi sont moins complètes.

OBJECTIFS

- Analyser la faisabilité d'un appariement entre des données d'AP et la cohorte cancer (généraliste de données en oncologie) extraits du SNDS de plus de 6 millions de personnes atteintes de cancer et de certaines personnes à haut risque.
- Évaluer les performances et les biais potentiels de cet appariement, ce qui, à notre connaissance, n'a jamais été réalisé auparavant.

MÉTHODE

Un appariement indirect déterministe a été réalisé entre les deux bases de données suivantes :

Données d'AP	Extrait de la Cohorte Cancer (CC)
Patients informés et traités en France par durvalumab pour un cancer du poumon non à petites cellules et non opérable entre le 1 ^{er} octobre 2017 et le 31 décembre 2018 dans 13 centres	Ciblage des patients ayant reçu durvalumab entre le 1 ^{er} octobre 2017 et le 31 décembre 2018
N = 457	N = 666

Les différences clés de l'appariement sont :

- Pré-processing (nettoyage, désignation et harmonisation des données, ...)
- Définition et analyse des variables d'appariement, de contrôle et d'étude
- Application du processus algorithmique d'appariement et évaluation de leurs performances
- Pour l'algorithmie la plus performante, analyse de la concordance et des différences entre les deux sources de données chez les personnes appariées et non appariées

RÉSULTATS

Chaque algorithme va tester l'égalité des variables sélectionnées pour l'appariement entre les deux sources de données, selon le tableau suivant :

Tableau 1 : Variables utilisées pour les différents algorithmes d'appariement

Algorithme	Sexe	Risque de récidive	Année de naissance	Prénoms	Département
A1	x	x	x	x	x
A2	x	x	x	x	x
A3	x	x	x	x	x

• Dans l'Algorithme 2, l'appariement révisé est basé sur l'année de naissance, le département de résidence et le prénom et sur le sexe (si le sexe n'est pas renseigné).

Tableau 2 : Taux d'appariement et répartition des effectifs par les différents algorithmes d'appariement testés

Algorithme	Taux d'appariement	Effectif apparié	Effectif non apparié
A1	85,2%	370	85
A2	84,2%	364	89
A3	85,3%	392	63

Figure 1 : Méthodologie de l'appariement final

Données d'AP (N = 457) et Données de la cohorte cancer (N = 666) sont combinées via un Algorithme d'appariement (A3) pour produire un Extrait de la cohorte cancer (N = 667 (1,0 sur 100)) et un Extrait de l'appariement (N = 392 (1,1)).

CONCLUSION

- Cette étude illustre la faisabilité de l'appariement des données d'AP avec le SNDS par une méthode indirecte déterministe avec un taux d'appariement > 85%.
- D'autres appariements de données d'AP seront nécessaires pour valider cette méthodologie et sa reproductibilité.
- Des travaux sont en cours afin :
 - d'évaluer la performance d'utiliser des données du SNDS pour un appariement avec les données d'AP
 - d'évaluer le suivi des patients et l'efficacité en vie réelle des traitements en AP à partir des données du SNDS pour les patients appariés.

CONTEXTE ET OBJECTIF

- Enjeu : enrichir les données recueillies pendant l'Accès Précoce (AP) avec les données du SNDS

→ Objectif : analyse de la **faisabilité** d'un appariement entre des données d'ATUc et la Cohorte Cancer (CC)

MÉTHODE

- Appariement **indirect déterministe** réalisé entre les deux bases de données via des variables socio-démographiques et le numéro FINESS de l'établissement de santé

Données d'ATUc

Patients traités en France par durvalumab pour un cancer du poumon non à petites cellules et non opérable

N = 457

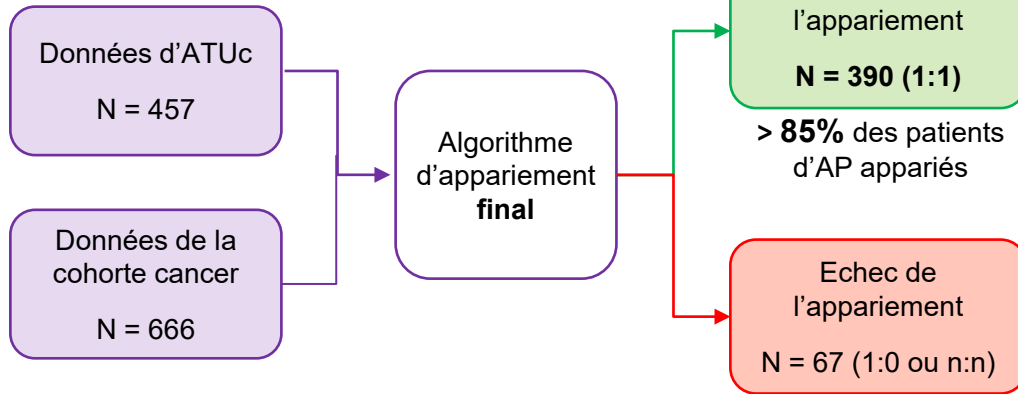
Extraction de la Cohorte Cancer (CC)

Ciblage des patients ayant reçu durvalumab sur la période d'AP

N = 666

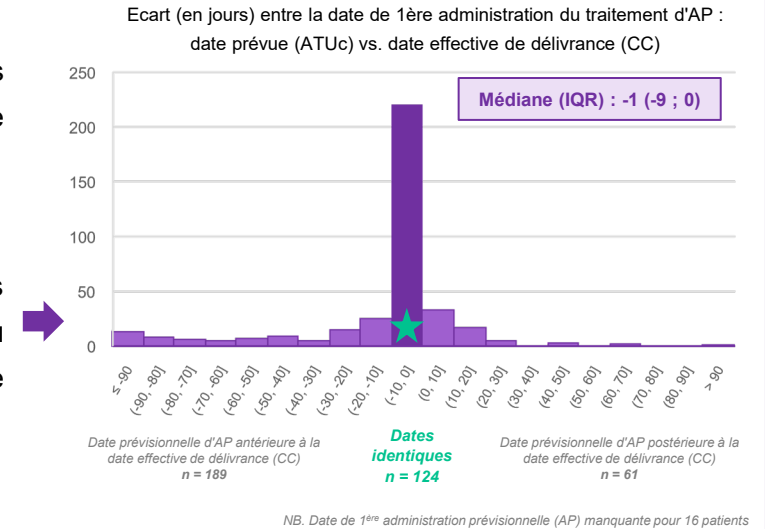
RÉSULTAT DE L'APPARIEMENT

Appariement final



ÉVALUATION DES PERFORMANCES ET BIAIS POTENTIELS

- Concordance élevée sur les variables d'appariement et de contrôle
- Pas de biais de sélection observé
- Certains écarts de dates entre les deux sources (1^{ère} délivrance du traitement d'ATUc et diagnostic de cancer)



CONCLUSION ET PERSPECTIVES

Cette étude illustre la faisabilité de l'appariement des données d'ATUc et d'AP avec le SNDS par une méthode indirecte déterministe avec un taux d'appariement **> 85%**.

D'autres appariements de données d'AP seront nécessaires pour valider cette méthodologie et sa reproductibilité.

- Des travaux d'évaluation du suivi des patients et de l'efficacité en vie réelle du traitement en AP à partir des données du SNDS sont en cours pour les patients appariés. Ceci permettra d'apprécier la pertinence d'apparier les données d'AP avec les données du SNDS dans cet objectif.