

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 données de programmes d'accès dérogatoire aux médicaments sont une opportunité pour décrire la première utilisation de traitements innovants en vie réelle.
- Depuis la réforme de 2021 sur les accès dérogatoires aux médicaments, les autorisations temporaires d'utilisation de cohorte (ATUc) ont été remplacées par les accès précoces pré- et post-AMM (AP).
- 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).
- Dans les données d'ATUc et d'AP, les données des fiches de demande d'accès au traitement sont quasi-exhaustives, mais les données de suivi sont moins complètes.

OBJECTIFS

- Evaluer la faisabilité d'un appariement entre des données d'AP et la cohorte cancer (plateforme de données en cancérologie) extraction du SNDS de plus de 8 millions de personnes atteintes de cancer et de certaines personnes à haut risque.
- Etudier 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'ATUc	Extraction 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 113 centres ➤ N = 457	Ciblage des patients ayant reçu durvalumab entre le 1 ^{er} octobre 2017 et le 31 décembre 2018 ➤ N = 666

Les différentes étapes de l'appariement sont :

- Pre-processing (recettage, description et harmonisation des données, ...)
- Définition et analyse des variables d'appariement, de contrôle et d'étude
- Application de plusieurs algorithmes d'appariement et évaluation de leurs performances
- Pour l'algorithme le plus performant, 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

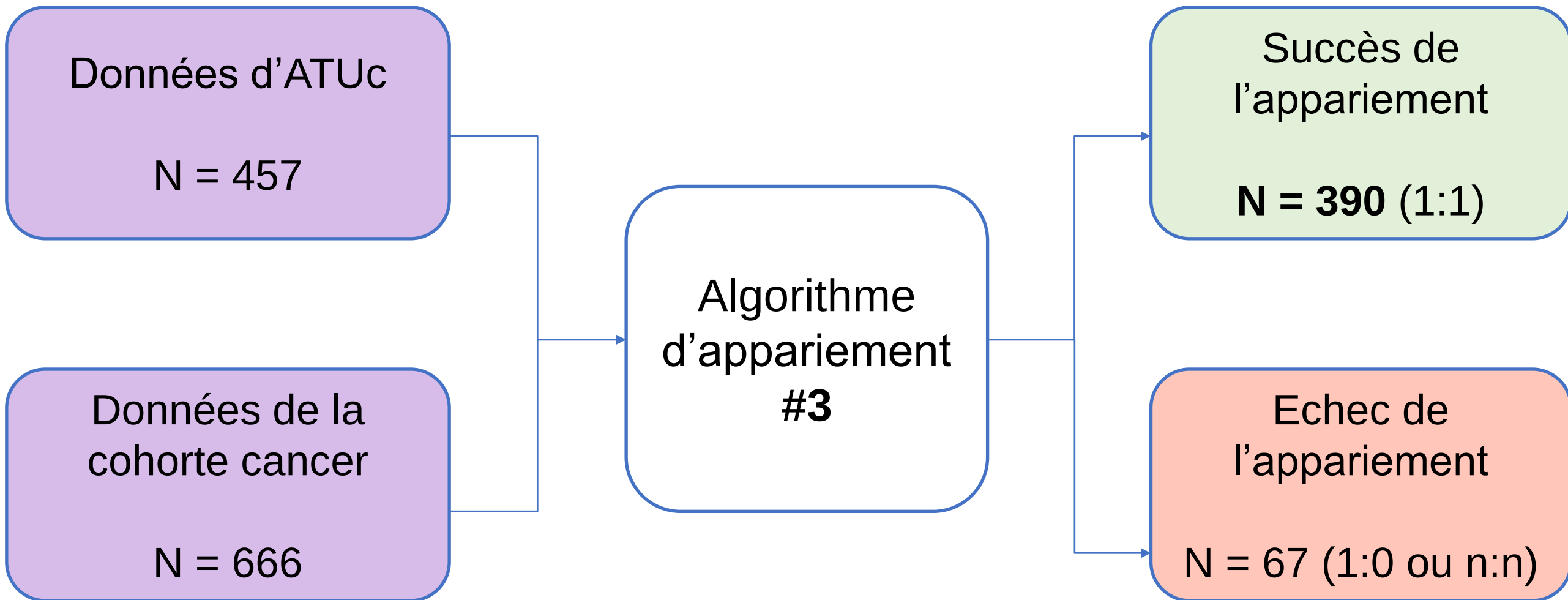
Algorithme	Sexe	Mois de naissance	Année de naissance	Etablissement de santé	
				Finess	Département
#1	x	x	x	x	
#2	x	x	x		x
#3*	x*	x*	x		x

* Dans l'algorithme 3, l'appariement nécessitait une égalité de l'année de naissance, du département de l'établissement de santé ainsi que du (sexe OU mois de naissance).

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

Algorithme	Taux d'appariement	Effectif apparié	Effectif non apparié	
		1:1	1:0	n:n
#1	80,9%	370	83	4
#2	84,0%	384	69	4
#3	85,3%	390	63	4

Figure 1 : Diagramme de flux de l'appariement final



Concordance entre les deux sources de données

La concordance des variables d'appariement est élevée (Cohen's kappa > 0.8) ou excellente (ICC > 0.9).

Tableau 3 : Concordance des variables qualitatives de contrôle chez les appariés

Variables de contrôle	Données d'ATUc	Extraction de la CC
Diagnostic du cancer du poumon	390 (100%)	387 (99%)
Dispensation du traitement d'ATUc	390 (100%)	390 (100%)
Antécédents de chimiothérapie pour le cancer du poumon	389 (100%)	388 (99%)
Antécédents de radiothérapie pour le cancer du poumon	389 (100%)	260 (67%)

Tableau 4 : Concordance des variables quantitatives de contrôle chez les appariés

Variables de contrôle	Différence de médiane (IQR)
1 ^{ère} administration du traitement d'ATUc : date prévue de début (ATUc) vs date effective de dispensation (CC) en jours	-1 (-9 ; 0)
Date de diagnostic du cancer (ATUc) vs date de premier soin pour le cancer (CC) en jours	0 (-24 ; 8)

Différence entre les populations appariées et non appariées

Les patients appariés vs non appariés (ATUc) : aucune différence significative sur les variables d'appariement et de contrôle, sauf concernant les antécédents de chimiothérapie et de radiothérapie.

Les personnes appariées vs non appariées (CC) : aucune différence significative sur les variables d'appariement et de contrôle à l'exception des antécédents de radiothérapie et du département de l'établissement de santé.

Limites de l'étude

- Certaines variables communes aux deux sources de données n'ont pas pu être utilisées pour l'appariement car les définitions sont différentes, comme les comorbidités.
- Les antécédents de radiothérapie n'ont pas pu être utilisés comme variable de contrôle puisque le diagnostic est absent pour les irradiations réalisées en libéral dans le SNDS.

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.