# CHAÎNAGE DES DONNÉES DES REGISTRES LOCAUX ET DE LA COHORTE CANCER (SNDS) : MÉTHODE ET RÉSULTATS

Marion Narbeburu

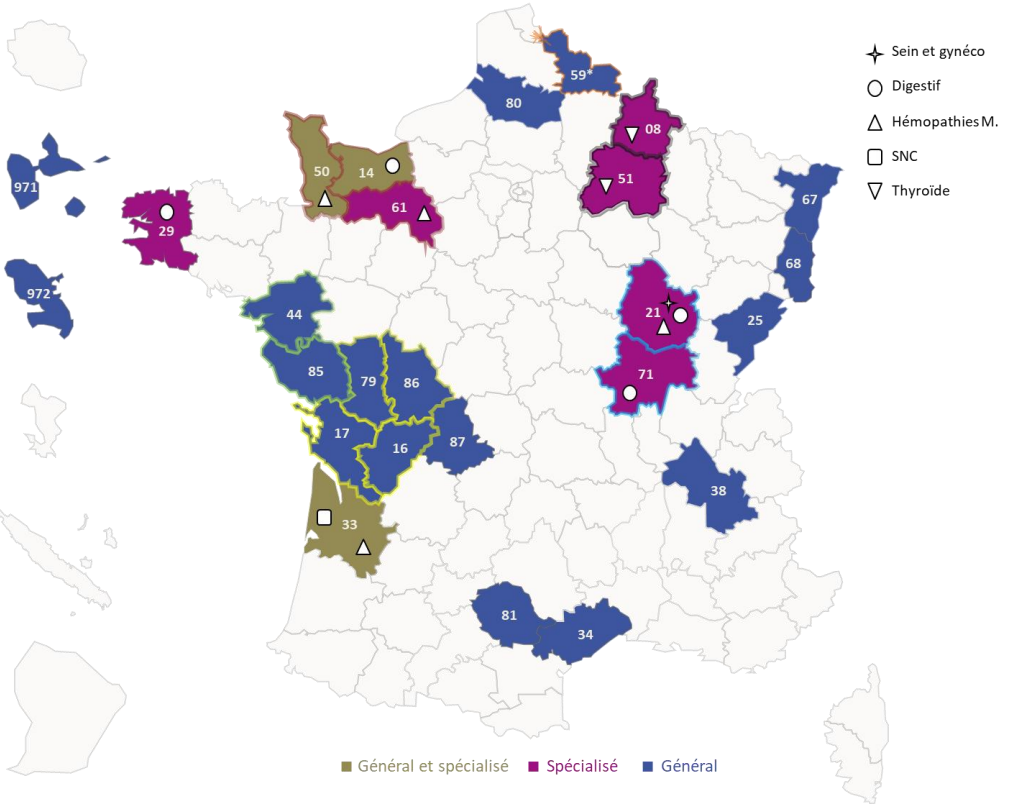
TRAVAUX EN COLLABORATION AVEC

# Les registres des cancers

**Mission** : Recueil **exhaustif** et **continu** des cas de cancers sur le territoire couvert par le registre

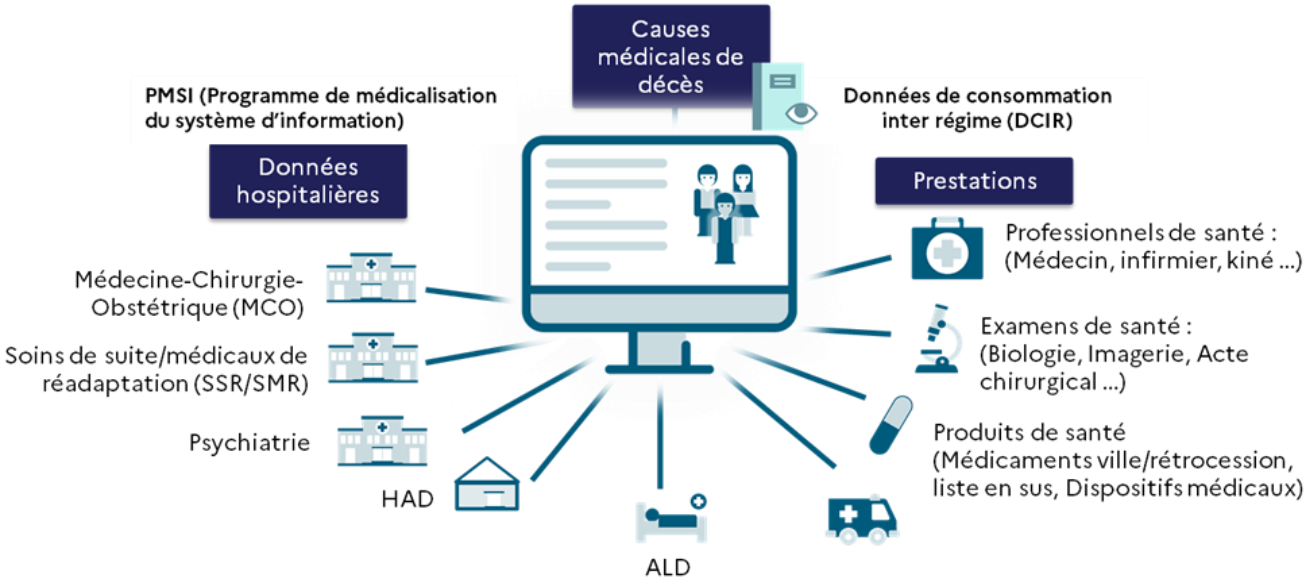
**27** registres de cancers généraux ou spécialisés financés par l'INCa

- 2 registres pédiatriques
- 2 DROM
- Fédérés au sein du réseau
- Couverture **~22-24%** France hexagonale



# La Cohorte Cancer (CC)

= extraction du Système National des Données de Santé (SNDS) des patients atteints de cancer en France



# Objectif

**Chaîner les données des Registres locaux avec la Cohorte Cancer**

**Registres locaux**
Données transmises : tumeurs invasives incidentes 2010-2021 en zone registre

APPARIEMENT

**Cohorte Cancer (CC)**
Extraction du SNDS des patients atteints de cancer en France

Gold standard pour l'enregistrement et la qualification des cas de cancer (topographie, histologie, date de diagnostic, …)

Données longitudinales pour l'étude des trajectoires patients jusqu'au décès à partir des consommations de soins présentées au remboursement à l'AM

# 1. Méthode

# Type d'appariement

3 types d'appariement :

- Appariement direct déterministe :
  - Via un identifiant commun aux 2 sources (ex : NIR)

CADRE DE L'APPARIEMENT REGISTRES – COHORTE CANCER

- Appariement indirect déterministe :
  - Via des variables indirectement identifiantes (ex : sexe, date de naissance, commune de résidence, localisation du cancer, date(s) de soin, date de décès, …)
- Appariement indirect probabiliste :
  - Calcul d'une probabilité que la paire soit la bonne pour chaque paire potentielle

# Stratégie d'appariement

**Tumeurs solides - registres généraux**

Tumeurs solides, ensemble des registres généraux

Exclusion des cas pédiatriques et des hémopathies malignes (dans des appariements dédiés ad hoc)

**Tumeurs solides - registres spécialisés**

Sur la zone géographique couverte par un registre général : SNC Gironde & Digestif Calvados

Sur la zone géographique non couverte par un registre général, couverte par un registre spécialisé, par spécialité :

- Digestif Finistère et Bourgogne
- Sein Gynéco Côte d'Or
- Thyroïde Marne Ardennes

**Hémopathies malignes**

Registres spécialisés HM Gironde, Basse-Normandie et Côte d'Or + cas HM dans les registres généraux

**Pédiatrie**

Registre national de pédiatrie

# Stratégie d'appariement

**OBJET DE LA PRESENTATION = APPARIEMENT TUMEURS SOLIDES DE L'ADULTE DES REGISTRES GENERAUX ET SPECIALISES**

**Tumeurs solides - registres généraux**

Tumeurs solides, ensemble des registres généraux

Exclusion des cas pédiatriques et des hémopathies malignes (dans des appariements dédiés ad hoc)

**Tumeurs solides - registres spécialisés**

Sur la zone géographique couverte par un registre général : SNC Gironde & Digestif Calvados

Sur la zone géographique non couverte par un registre général, couverte par un registre spécialisé, par spécialité :

- Digestif Finistère et Bourgogne
- Sein Gynéco Côte d'Or
- Thyroïde Marne Ardennes

**Hémopathies malignes**

Registres spécialisés HM Gironde, Basse-Normandie et Côte d'Or + cas HM dans les registres généraux

**Pédiatrie**

Registre national de pédiatrie

# Etapes d'appariement

**1 Définition des variables d'appariement**

- Sexe, mois et année de naissance, commune de résidence, département de résidence, date de décès, localisation du cancer (appareil, organe), date de diagnostic

**2 Préparation de la population SOURCE (=Registres)**

- Sélection des patients dans le périmètre de l'appariement
- Reconstitution de la localisation du cancer pour faire converger CIM-O3 (registres) - CIM-10 (CC)

**3 Sélection et préparation de la population CIBLE (= Cohorte Cancer)**

- Blocage de la population Cohorte Cancer sur la zone registre
- Sélection des patients dans le périmètre de l'appariement
- Reconstitution de la localisation du cancer pour faire converger CIM-O3 (registres) - CIM-10 (CC)
- Identificat° des événements d'intérêt pouvant correspondre à la date diag

**4 Construction des clés d'appariement**

- Clés créées à partir de combinaisons des variables d'appariement

**5 Mise en œuvre de l'appariement**

- Appariement séquentiel avec passage successif de toutes les clés
- Ordre de passage des clés d'appariement : de la plus discriminante (générant le moins de doublons dans la population CIBLE) à la moins discriminante

**6 Évaluation de la validité de l'appariement**

- Nombre de matchs selon leur type et de non-matchs (1:1, 1:0, 1:N, N:1, N:N)
- Analyse des clés des 1:1
- Analyse de concordance et invalidation des cas appariés à tort
- Comparaison des appariés aux non appariés

# Etape 1 – Définition des variables d'appariement

| | Registres | Cohorte Cancer |
|---|---|---|
| **Sexe** | Variable sexe | Variable sexe |
| **Mois et année de naissance** | Variables mois et année de naissance | Variables mois et année de naissance |
| **Commune de résidence** | Commune de résidence au moment du diagnostic | Extraction de l'ensemble des communes disponibles dans la base par identifiant (DCIR, PMSI, référentiel bénéficiaire), avec la date correspondante, sur la période d'intérêt 2010-2021 |
| **Département de résidence** | Reconstitué à partir de la commune de résidence au diagnostic | Reconstitué à partir des communes de résidence extraites ci-dessus |
| **Date de décès** | Reconstituée à partir de la date de dernières nouvelles pour les patients avec un état aux dernières nouvelles = décédé<br>Sources : Rnipp, dossier médical, … | Sources : référentiel bénéficiaires, mode de sortie d'une hospitalisation, causes médicales de décès, … avec vérif. de cohérence entre les différentes sources<br>*Point d'attention : non exhaustif hors régime RG* |
| **Localisation du cancer** (appareil et organe) | Reconstituée à partir du codage **CIM-O3** (codes topo/morpho) | Reconstituée à partir du codage **CIM-10** (PMSI / ALD) et/ou la délivrance de médicaments anti-cancéreux spécifiques d'une localisation de cancer |
| **Date diagnostique du cancer** | Variable date de diagnostic du cancer<br>(à partir du jour / mois / année de diagnostic) | Approximation de la date diagnostique par des dates de consommations de soins ou d'événements d'intérêt (actes d'intérêt : biopsie, anapath, endoscopie, exérèse, …, séjour PMSI, mise en ALD, …) |

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Institut National du Cancer

2 Préparation de la population SOURCE (=Registres)

# Etape 2 – Préparation des données des registres pour l'appariement tumeurs solides

**Transformation CIM-O3 → organe / appareil** ← CIM-10 pour les besoins de l'appariement :

- Le codage de la localisation du cancer diffère dans les bases :
  - Côté registre : utilisation de la CIM-O3 3.2
  - Côté Cohorte Cancer : utilisation de la CIM-10, moins précise

→ **Nous avons fait converger CIM-O3 (et CIM-10) vers une liste de 46 organes, eux-mêmes rattachés à 15 appareils.**

**Exclusions et exceptions :**

- Si morpho = Sarcome, organe = Tissus Mous. + organe de la topo
- Exclusion des tumeurs de la peau hors mélanomes et hors sarcomes de Kaposi de la peau

| Liste TOPO CIM-O | Organe | Liste CIM-10 (simplifié sur 3 caractères) |
|---|---|---|
| C00 Lèvre | Lèvre | C00 |
| C01 Base de la langue, C02 Autres localisations et localisations (…) | Langue | C01, C02 |
| C03 Gencive, C04 Plancher de la bouche, C05 Palais, (…) | Cav. Orale | C03, C04, C05, C06, C46 |
| C07 Glande parotide, C08 Autres glandes salivaires principales (…) | Gland. Saliv. | C07, C08 |
| C09 Amygdale, C10 Oropharynx, C11 Nasopharynx (…) | Pharynx | C09, C10, C11, C12, C13, C14 |
| C14 Autres localisations et localisations mal définies de la lèvre, (…) | VADS sans précis. | C00-C14, C46, D00, D37, Z85 |
| C15 Oesophage | Oesophage | C15, D00 |
| C16 Estomac | Estomac | C16, D00, D37 |
| C17 Intestin grêle | Intestin grêle | C17, D01, D37 |
| C18 Côlon, C19 Jonction recto-sigmoidienne, C20 Rectum, C21 (…) | Colon-Rectum-Anus | C18, C19, C20, C21, D01, D37 |
| C22 Foie et voies biliaires intrahépatiques, C23 Vésicule biliaire, (…) | Foie et voies bil. | C22, C23, C24, D01, D37 |
| C25 Pancréas | Pancréas | C25 |
| C26 Autres localisations et localisations mal définies des (…) | App. Dig. sans précis. | C26, C46, D01, D37, Z85 |
| C30 Fosse nasale et oreille moyenne, C31 Sinus annexes de la face | Fosses nas., sinus, oreille | C30, C31 |
| C32 Larynx | Larynx | C32, D02, D38 |
| C33 Trachée, C34 Bronche et poumon | Trach., Bronch., Poum. | C33, C34, D02, D38, Z85 |
| C37 Thymus, C38 Coeur, médiastin et plèvre, C39 Autres loc. (…) | App. Respi. sans précis. | C37, C38, C39, C45, C46, D02, D38, Z85 |
| C38 Coeur, médiastin et plèvre | Médiastin | C38, D38 |
| C38 Coeur, médiastin et plèvre | Plèvre | C38, C45, D38 |
| C40 Os, articulations et cartilage articulaire des membres (…) | Os, artic., Cartil. Artic. | C40, C41, D48 |
| C44 Peau | Peau | C43, ~~C44~~, C46, D03, D04, D48 |
| C48 Rétropéritoine et péritoine | Peri.-Retroperi. | C45, C48, D48 |
| C47 Nerfs périphériques et système nerveux autonome | Nerfs Périph. | C47, D48 |
| C49 Tissu conjonctif, tissu sous- cutané et autres tissus mous | Tiss. Mous | B21, C46, C49, D48 |
| C50 Sein | Sein | C50, D05, D48, Z85 |
| C51 Vulve, C52 Vagin | Vulve et vagin | C51, C52, D07 |
| C53 Col utérin | Col utérus | C53, D06 |
| C54 Corps utérin, C55 Utérus SAI | Utérus | C54, C55, D07 |
| C56 Ovaire | Ovaire | C56, D39 |
| C57 Organes génitaux féminins, autres et (…) | Org. Gen. Fem. sans précis. | C57, C58, D07, D39 |
| C60 Verge | Pénis | C60, D07 |
| C61 Prostate | Prostate | C61, D07, D40 |
| C62 Testicule | Testicule | C62, D40 |
| C63 Organes génitaux masculins, autres et non spécifiés | Org. Gen. Masc. sans précis. | C63, D07, D40 |
| C64 Rein | Rein | C64, D41 |
| C65 Bassinet (du rein), C66 Uretère | Voies Uri. Hautes | C65, C66, D41 |
| C67 Vessie | Vessie | C67, D09, D41 |
| C68 Organes urinaires, autres et non spécifiés | App. Uri. sans précis. | C68, D09, D41, Z85 |
| C69 OEil et annexes | Oeil | C69, D09, Z85 |
| C70 Méninges, C71 Encéphale, C72 Moelle épinière, nerfs (…) | Syst. Nerv. Centr. | C70, C71, C72, D42, D43, Z85 |
| C73 Glande thyroide | Thyroïde | C73, D44 |
| C74 Glande surrénale | Surrénale | C74, D44 |
| C75 Autres glandes endocrines et structures apparentées | Gland. Endo. sans précis. | C75, D09, D44, Z85 |
| C76 autres localisations et localisations mal définies, C80 Site primaire inconnu | Org. non Attrib. | B21, C45, C76, C80, C97, D00, D02, D07, D09, D37, D38, D48, Z85 |
| C77 Ganglions | Ganglions | C77, D48 |

# Etape 3 – Sélection et préparation de la population CIBLE (= Cohorte Cancer)

## Approximation de la date de diagnostic du cancer

- Extraction des actes et événements d'intérêt disponibles dans les données de la Cohorte Cancer
- Par ordre de priorité, afin de coller au mieux aux règles de codage des registres des cancers (ENCR) :




| Priorité 1 | Priorité 2 | Priorité 3 |
|---|---|---|
| • Biopsie<br>• Acte d'anapath.<br>• Endoscopie<br>• Exérèse<br>• NABM | • Séjour PMSI niveau organe | • Séjour PMSI niveau appareil<br>• Mise en ALD<br>• Acte d'imagerie spécifiques<br>• 1ère délivrance d'un médicament anti-cancéreux spécifique de l'organe |

# Etape 4 – Construction des clés d'appariement



Les clés sont créées à partir de combinaisons des variables d'appariement, certaines niveau Patient et certaines niveau Tumeur :

| Clé | Sexe | Mois-année nais. | Dép. résidence | Com. résidence | Date de décès | | Organe | Appareil | Date de diagnostic prio 1 | | | Date de diagnostic prio 2 | | | Date de diagnostic prio 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Exacte | Mois-année | | | Exacte | [-7;+21j] | Année | Exacte | [-7;+21j] | Année | Exacte | [-7;+21j] | Année |
| 1 | x | x | x | x | x | | x | | x | | | | | | | | |
| 2 | x | x | x | x | x | | x | | | | | x | | | | | |
| 3 | x | x | x | x | x | | x | | | | | | | | x | | |
| 4 | x | x | x | x | | x | x | | x | | | | | | | | |
| 5 | x | x | x | x | | x | x | | | | | x | | | | | |
| 6 | x | x | x | x | | x | x | | | | | | | | x | | |
| 7 | x | x | x | x | x | | | x | x | | | | | | | | |
| 8 | x | x | x | x | x | | | x | | | | x | | | | | |
| 9 | x | x | x | x | x | | | x | | | | | | | x | | |
| 10 | x | x | x | x | | x | | x | | | | | | | | | |
| 11 | x | x | x | x | | x | | x | | | | | | | | | |
| 12 | x | x | x | x | | x | | x | | | | | | | | | |
| … | | | | | | | | | | | | | | | | | |
| 144 | x | x | x | | | | | | | | | | | | | | |

Variables obligatoires

→ 144 combinaisons possibles de ces variables, en imposant Sexe, Mois-Année de naissance et Département comme variables obligatoires

# Etape 4 – Construction des clés d'appariement

Les clés sont créées à partir de combinaisons des variables d'appariement, certaines niveau Patient et certaines niveau Tumeur :

| Clé | Sexe | Mois-année nais. | Dép. résidence | Com. résidence | Date de décès | | Organe | Appareil | Date de diagnostic prio 1 | | | Date de diagnostic prio 2 | | | Date de diagnostic prio 3 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Exacte | Mois-année | | | Exacte | [-7;+21j] | Année | Exacte | [-7;+21j] | Année | Exacte | [-7;+21j] | Année |
| 1 | x | x | x | x | x | | x | | x | | | | | | | | |
| 2 | x | x | x | x | x | | x | | | | | x | | | | | |
| 3 | | | | | | | | | | | | | | | | | |
| 4 | | | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | | |
| 6 | | | | | | | | | | | | | | | | | |
| 7 | | | | | | | | | | | | | | | | | |
| 8 | | | | | | | | | | | | | | | | | |
| 9 | | | | | | | | | | | | | | | | | |
| 10 | | | | | | | | | | | | | | | | | |
| 11 | x | x | x | x | | x | | x | | | | | | | | | |
| 12 | x | x | x | x | | x | | x | | | | | | | | | |
| ... | | | | | | | | | | | | | | | | | |
| 144 | x | x | x | | | | | | | | | | | | | | |

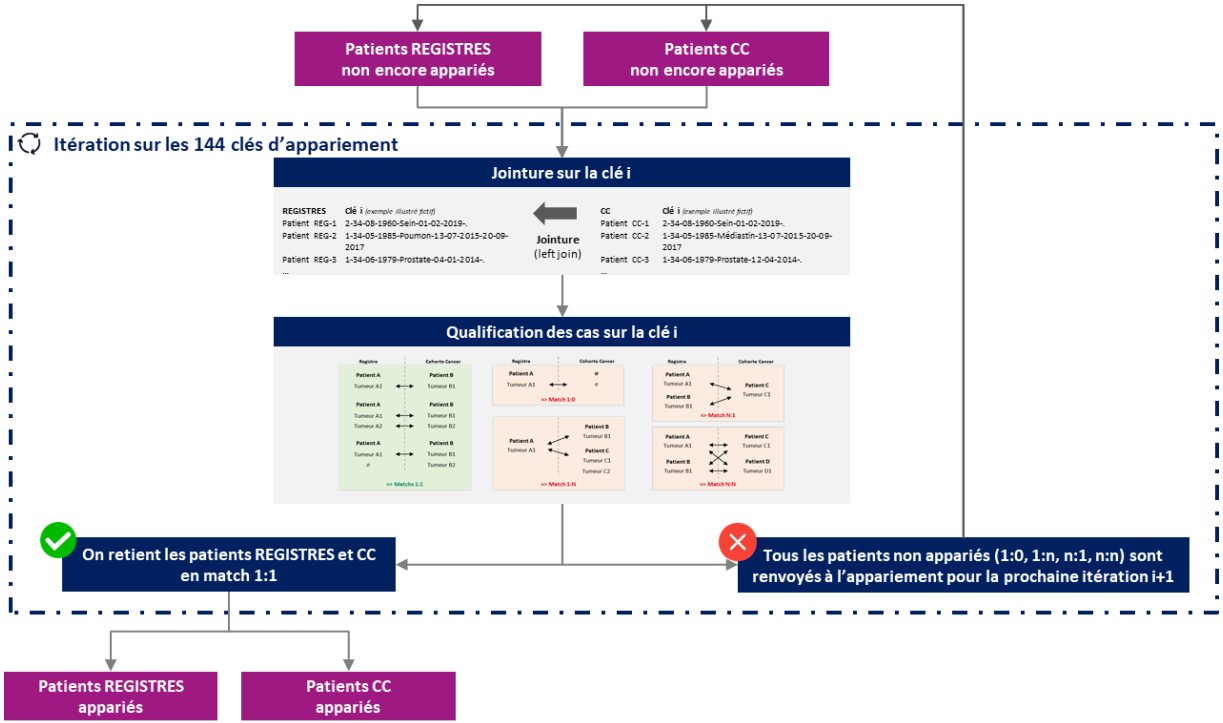
Variables obligatoires

→ 144 combinaisons possibles de ces variables, en imposant Sexe, Mois-Année de naissance et Département comme variables obligatoires

**Illustration d'une clé = concaténation des variables d'appariement contenues dans la clé :**

| Patient | Sexe | Mois-année nais. | Dép. résidence | Com. résidence | Date de décès | Organe | Date de diagnostic prio 1 | Clé 1 |
|---|---|---|---|---|---|---|---|---|
| **PAT-1** | 2 | 02/1956 | 38 | 38421 | 10/06/2015 | Sein | 25/08/2014 | 2-02/1956-38-38421-10/06/2015-Sein-25/08/2014 |

# Etape 5 – Mise en œuvre de l'appariement



# Etape 6 – Evaluation de la validité de l'appariement

1. Nombre de matchs selon leur type et de non-matchs (1:1, 1:0, 1:N, N:1, N:N)

2. Analyse des clés des 1:1

3. Analyse de concordance

   → En fonction de l'analyse de concordance, définition de règles de décision pour arbitrer sur des matchs 1:1 à retirer de l'appariement

4. Comparaison des appariés aux non appariés

# 2. Résultats

## Rappel du périmètre : Tumeurs solides chez l'adulte

# Descriptif des données registres

## AVANT APPARIEMENT

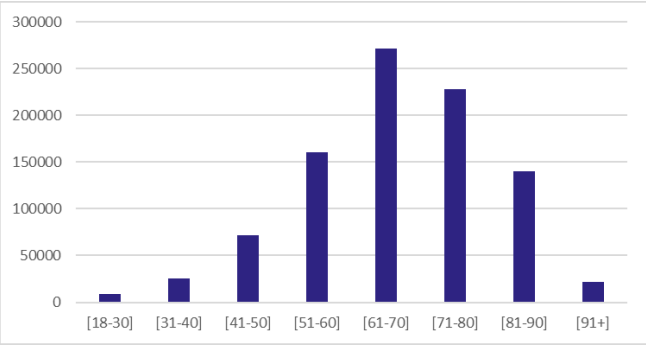
**874 659** patients

**926 887** tumeurs invasives solides, incidentes 2010-2021

**46 %** de femmes
**54** % d'hommes

**53 %** de patients vivants (aux dernières nouvelles)

**Classe d'âge au diagnostic**




| Site de la tumeur solide *(invasif)* | Nb de tumeurs (%) |
|---|---|
| Col de l'utérus | 8 380 (0,9%) |
| Corps de l'utérus | 19 416 (2,1%) |
| **Côlon et rectum** | **130 146 (14,0%)** |
| Estomac | 20 711 (2,2%) |
| Foie | 29 711 (3,2%) |
| Intestin grêle | 4 420 (0,5%) |
| Larynx | 8 417 (0,9%) |
| Lèvre-bouche-pharynx | 33 822 (3,6%) |
| Mélanome de la peau | 35 280 (3,8%) |
| Mésothéliome de la plèvre | 2 486 (0,3%) |
| Œsophage | 16 059 (1,7%) |
| Ovaire | 13 302 (1,4%) |
| Pancréas | 36 988 (4,0%) |
| **Poumon** | **108 423 (11,7%)** |
| **Prostate** | **150 431 (16,2%)** |
| Rein | 34 964 (3,8%) |
| Sarcome | 12 612 (1,4%) |
| **Sein** | **152 036 (16,4%)** |
| Système nerveux central (SNC) | 12 665 (1,4%) |
| Testicule | 5 689 (0,6%) |
| Thyroïde | 22 952 (2,5%) |
| Vessie | 31 774 (3,4%) |
| Vésicule et voies biliaires | 8 624 (0,9%) |
| Autre site | 27 579 (3,0%) |

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Institut National du Cancer

# Descriptif des données CC

## AVANT APPARIEMENT

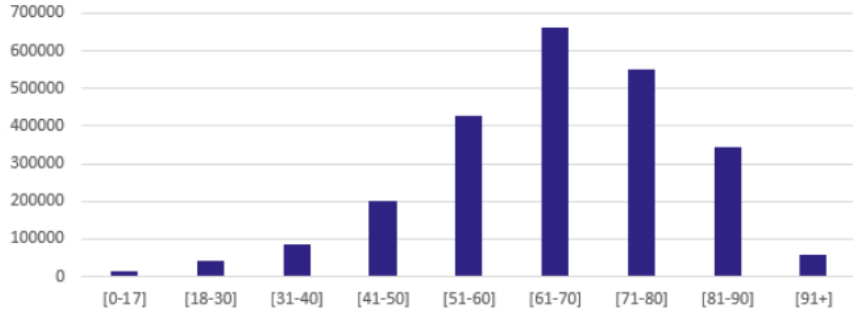
**1 686 546** patients ayant résidé à un moment dans la zone registre entre 2010 et 2021

**2 392 104** tumeurs invasives solides, détectées entre le 01/10/2009 et le 31/03/2022 chez les patients sélectionnés

**48 %** de femmes
**52** % d'hommes

**61 %** de patients vivants (aux dernières nouvelles)

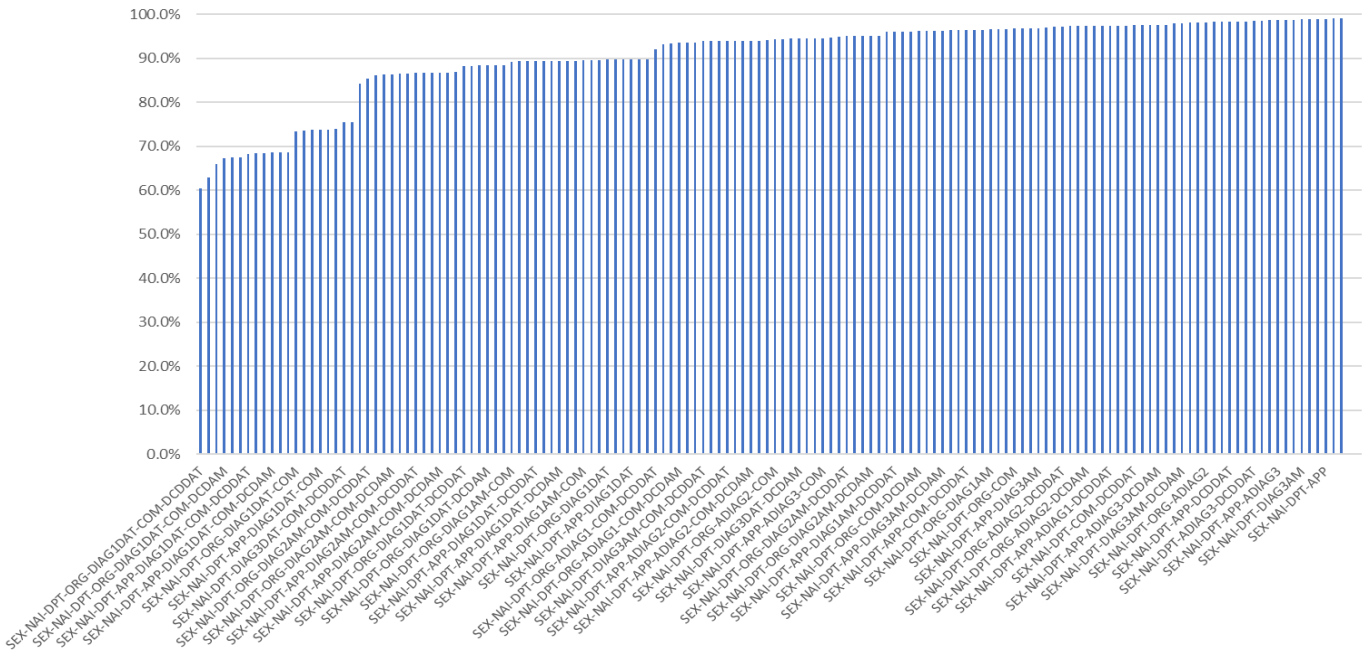
**Classe d'âge au diagnostic**




| Site de la tumeur solide *(tout comportement)* | Nb de tumeurs (% de patients) |
|---|---|
| App. Dig. sans précis. | 37767 (2.2%) |
| App. Respi. sans précis. | 11355 (0.7%) |
| App. Uri. sans précis. | 6831 (0.4%) |
| Cav. Orale | 17012 (1%) |
| Col utérus | 50606 (3%) |
| **Colon-Rectum-Anus** | **415725 (24.7%)** |
| Estomac | 42918 (2.5%) |
| Foie et voies bil. | 77458 (4.6%) |
| Fosses nas., sinus, oreille | 6337 (0.4%) |
| Gland. Endo. sans précis. | 10241 (0.6%) |
| Gland. Saliv. | 6178 (0.4%) |
| Intestin grêle | 17665 (1%) |
| Langue | 16640 (1%) |
| Larynx | 22961 (1.4%) |
| Lèvre | 3670 (0.2%) |
| Meta. seul | 6675 (0.4%) |
| Médiastin | 5352 (0.3%) |
| Nerfs Périph. | 2517 (0.1%) |
| Oeil | 5656 (0.3%) |
| Oesophage | 30606 (1.8%) |
| Org. Gen. Fem. sans précis. | 6366 (0.4%) |
| Org. Gen. Masc. sans précis. | 2021 (0.1%) |
| Org. non Attrib. | 51420 (3%) |
| Os, artic., Cartil. Artic. | 19692 (1.2%) |
| Ovaire | 34237 (2%) |
| Pancréas | 62378 (3.7%) |
| Peau | 69770 (4.1%) |
| Peri.-Retroperi. | 17023 (1%) |
| Pharynx | 37087 (2.2%) |
| Plèvre | 7742 (0.5%) |
| **Prostate** | **229817 (13.6%)** |
| Pénis | 2436 (0.1%) |
| Rein | 59291 (3.5%) |
| **Sein** | **262827 (15.6%)** |
| Surrénale | 8433 (0.5%) |
| Syst. Nerv. Centr. | 51119 (3%) |
| Syst. Nerv. sans précis. | 93 (0%) |
| Testicule | 10773 (0.6%) |
| Thyroïde | 38945 (2.3%) |
| Tiss. Mous | 21703 (1.3%) |
| **Trach., Bronch., Poum.** | **178224 (10.6%)** |
| Utérus | 42067 (2.5%) |
| VADS sans précis. | 11462 (0.7%) |
| Vessie | 106185 (6.3%) |
| Voies Uri. Hautes | 12023 (0.7%) |
| Vulve et vagin | 7086 (0.4%) |

# Taux de matchs 1:1 par clé

**Taux cumulé de matchs 1:1 par ordre de passage des 144 clés de l'appariement séquentiel**

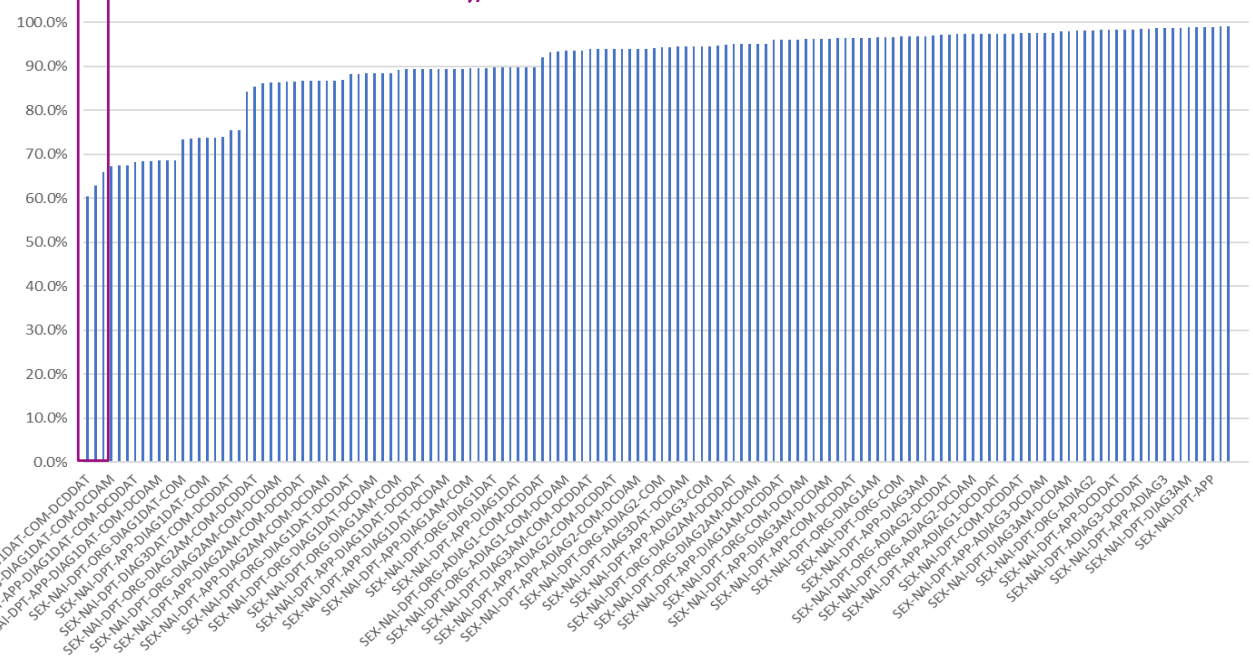



**866 388 patients registres en 1:1 (99.1%) à l'issue du séquentiel**

# Taux de matchs 1:1 par clé

**Taux cumulé de matchs 1:1 par ordre de passage des 144 clés de l'appariement séquentiel**

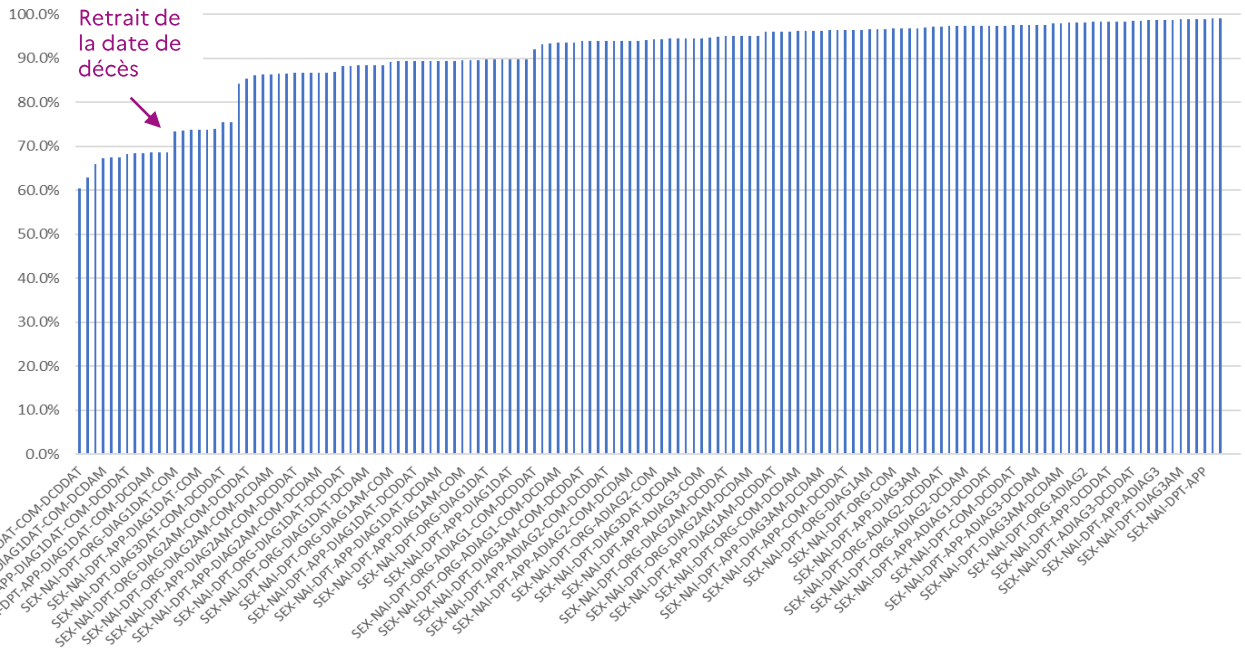
→ **Les 3 premières clés, les plus discriminantes (SEX-NAI-DPT-COM-ORG-DIAGDAT-DCDDAT), ramènent 66.3% des matchs 1:1**




**866 388 patients registres en 1:1 (99.1%) à l'issue du séquentiel**

# Taux de matchs 1:1 par clé

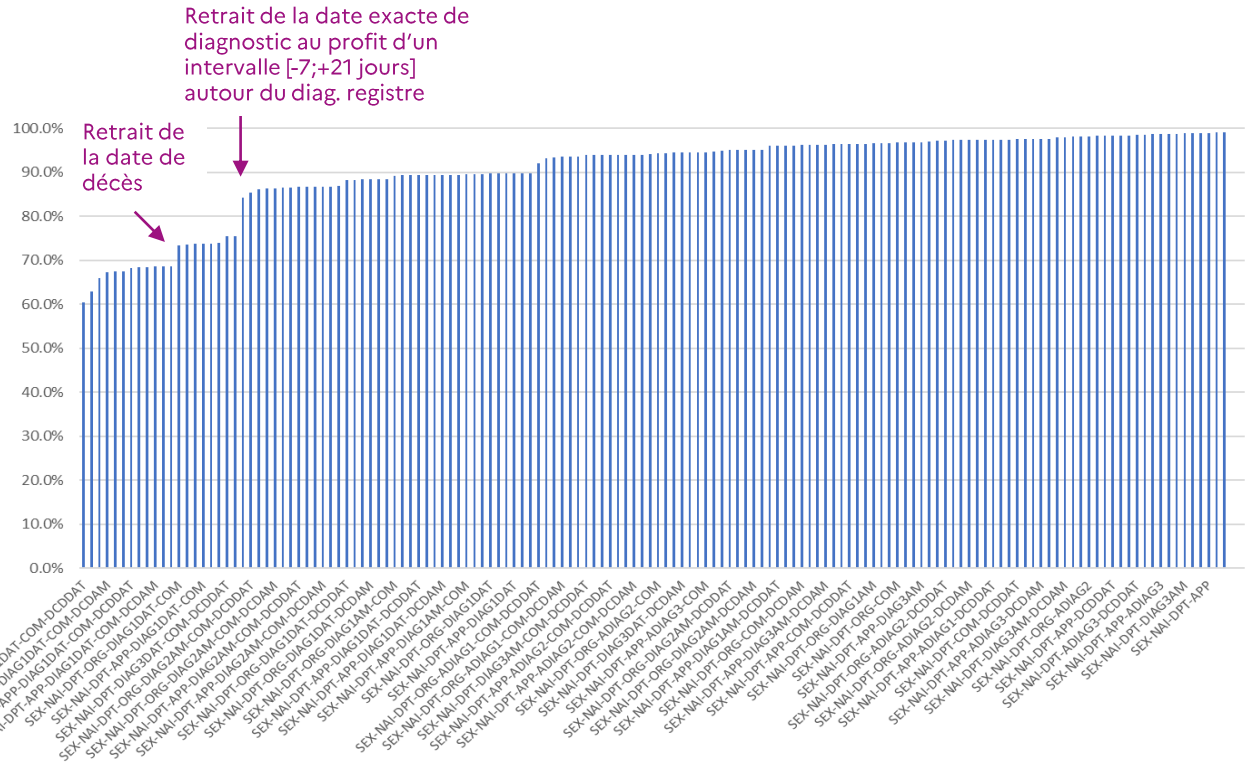
**Taux cumulé de matchs 1:1 par ordre de passage des 144 clés de l'appariement séquentiel**




**866 388 patients registres en 1:1 (99.1%) à l'issue du séquentiel**

# Taux de matchs 1:1 par clé

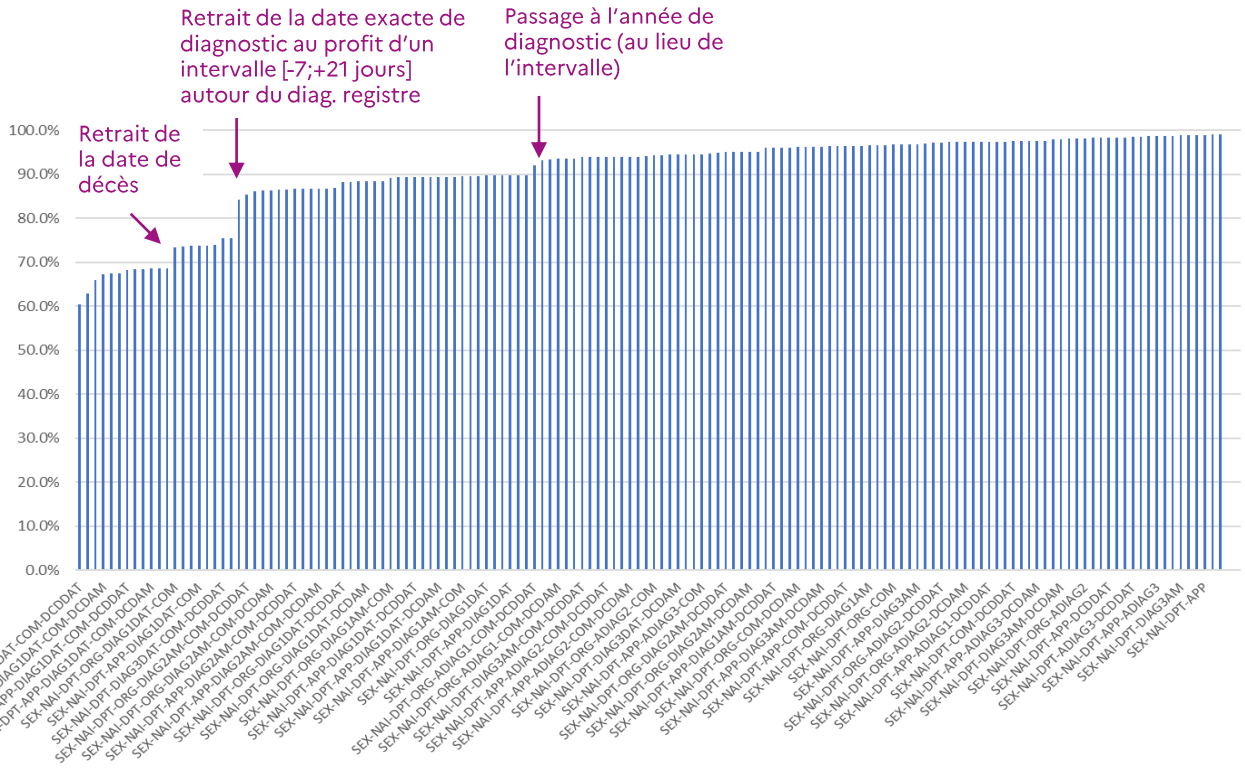
**Taux cumulé de matchs 1:1 par ordre de passage des 144 clés de l'appariement séquentiel**




**866 388 patients registres en 1:1 (99.1%) à l'issue du séquentiel**

# Taux de matchs 1:1 par clé

**Taux cumulé de matchs 1:1 par ordre de passage des 144 clés de l'appariement séquentiel**




**866 388 patients registres en 1:1 (99.1%) à l'issue du séquentiel**

# Concordance : analyse des variables une à une

| | Avant RD |
|---|---|
| **Sexe, mois année de naissance et département de résidence** (= variables obligatoires) | 100,0 % |
| **Commune de résidence** | 96,2 % |
| **Statut vital et date de décès** | 98,6 % |
| **Appareil** | 97,7 % |
| **Organe** | 96,6 % |
| **Présence d'un événement d'intérêt CC à +/- 1 mois du diagnostic registre** | 96,5 % |
| • dont Biopsie de l'organe à +/- 1 mois | 44.7 % |
| • dont Biopsie non attribuable à un organe à +/- 1 mois | 74.7 % |
| • dont Endoscopie de l'organe à +/- 1 mois | 23.6 % |
| • dont Exérèse de l'organe à +/- 1 mois | 32.8 % |
| • dont Séjour PMSI de l'organe à +/- 1 mois | 67.5 % |
| • dont Mise en ALD de l'organe à +/- 1 mois | 53.6 % |

**Commune de résidence** : Concordance si la commune de résidence au diagnostique côté registre, correspond à l'une des communes enregistrées dans la CC entre 2010 et 2021.

**Date de décès** :
Si date de décès renseignée dans les 2 sources : concordance si date de décès registre à [-7;+7] jours de la date de décès CC
Si aucune date de décès côté CC ni côté registre : considérés concordants
Si date de décès côté CC et aucune date de décès côté registre : considérés concordants
Si date de décès côté registre et aucune date de décès côté CC : considérés discordants, sauf si le patient CC est rattaché à un registre hors rég. général

**Appareil / organe :** Concordance si l'appareil / l'organe côté registre (déduit de la CIM-O3) correspond à l'un des appareils / organes du patient côté CC (déduit de la CIM-10) entre 2010 et 2021

# Etude des discordances sur les matchs 1:1

**1. Sur la localisation** :

Il existe des approximations sur le codage CIM-10 dans les données médico-administratives, notamment sur des localisations contigües (ex : Estomac-Œsophage / Pancréas-Foie et voies biliaires / Colon-Rectum / …) et des localisations non précisées.

*Exemple de discordance : patient avec une tumeur de l'intestin grêle côté registre ayant matché avec un patient avec un code C286 « Appareil digestif sans précision » dans le PMSI et une exérèse de l'intestin grêle retrouvée dans la CC à quelques jours de la date de diagnostic registre*

Etude des discordances chez les patients décédés à la même date (= référence) :

**Lorsque l'organe registre n'est pas précisé**
*(5 plus fréquents)*

| Organe CC | Organe RG |
|---|---|
| Vessie | App. Uri. sans précis. |
| Ovaire | Org. Gen. Fem. sans précis. |
| Voies Uri. Hautes | App. Uri. sans précis. |
| Foie et voies bil. | App. Dig. sans précis. |
| Rein | Voies Uri. Hautes |

**Lorsque l'organe CC n'est pas précisé**
*(5 plus fréquents)*

| Organe CC | Organe RG |
|---|---|
| Org. non Attrib. | App. Dig. sans précis. |
| Org. non Attrib. | Org. Gen. Fem. sans précis. |
| App. Respi. sans précis. | Médiastin |
| Org. non Attrib. | Peri.-Retroperi. |
| App. Dig. sans précis. | Pancréas |

**Pour les organes contigus**

| Organe CC | Organe RG |
|---|---|
| Peri.-Retroperi. | Org. Gen. Fem. sans précis. |
| Trach., Bronch., Poum. | Médiastin |
| Colon-Rectum-Anus | Intestin grêle |
| Estomac | Intestin grêle |
| App. Dig. sans précis. | Peri.-Retroperi. |
| Oesophage | Estomac |
| Pancréas | Foie et voies bil. |
| Foie et voies bil. | Pancréas |
| Plèvre | Trach., Bronch., Poum. |
| Médiastin | Trach., Bronch., Poum. |

Il existe également des **approximations sur le codage des métastases** (métastase codée en cancer primitif ou cancer primitif codé en métastase).

→ Ajout d'une tolérance sur les discordances sur la localisation pour ces cas particuliers

# Etude des discordances sur les matchs 1:1

**2. Sur la date de diagnostic :**

Pour certaines localisations diagnostiquées et traitées principalement en ville, le délai entre la date de diagnostic du registre et la détection dans la CC (via un séjour PMSI à posteriori) peut être important.

→ Ajout d'une tolérance sur les écarts sur la date de diagnostic pour les tumeurs de la peau, du sein et de la prostate

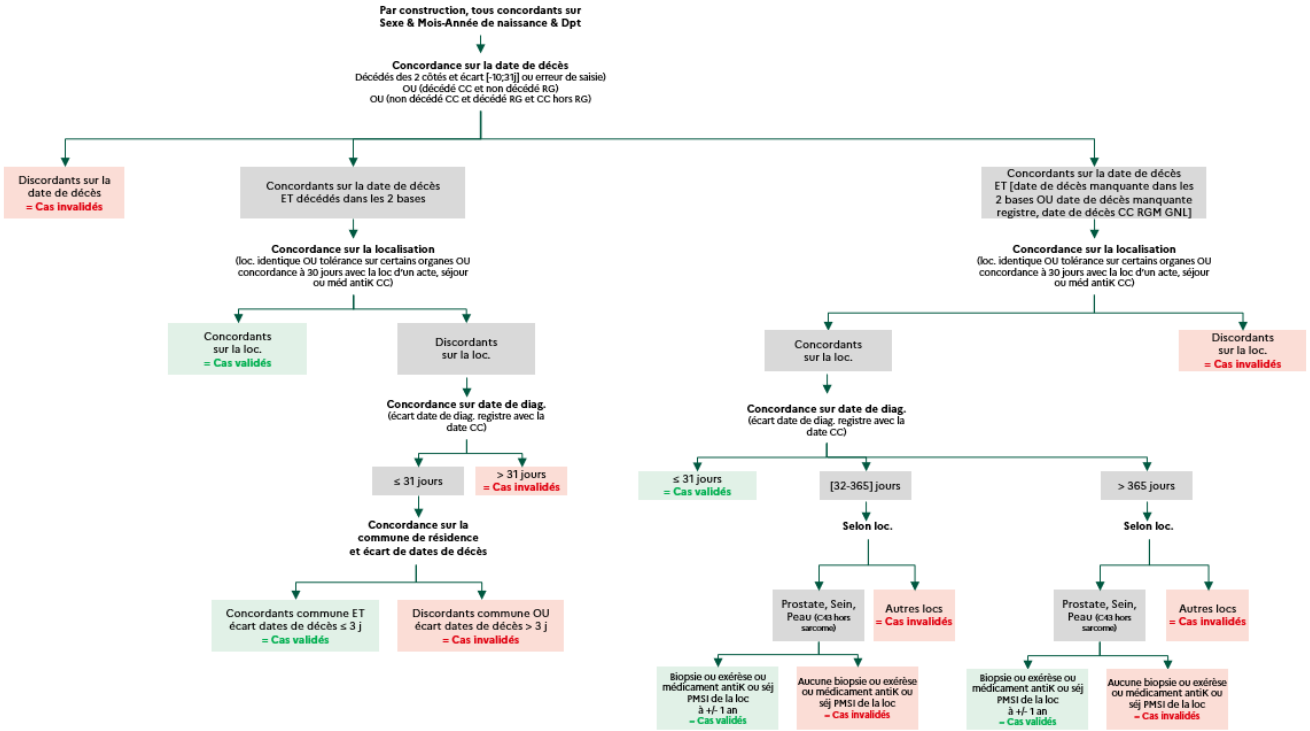
**3. Sur le statut vital et la date de décès :**

- L'information peut ne pas être exhaustive dans la CC pour les patients affiliés à un régime hors RG ou SLM
- Le jour de décès n'est pas toujours disponible dans la CC
- La date de décès côté registre peut être parfois manquante (chaînage Rnipp)

→ Ajout de tolérances sur les discordances sur le statut vital pour ces cas particuliers

# Règles de décision

## Construction d'un arbre de décision pour valider / invalider les cas de l'appariement



# Concordance : analyse des variables une à une

| | Avant RD | Après RD |
|---|---|---|
| **Sexe, mois année de naissance et département de résidence** (= variables obligatoires) | 100,0 % | 100,0 % |
| **Commune de résidence** | 96,2 % | 97,7 % |
| **Statut vital et date de décès** | 98,6 % | 100,0 % |
| **Appareil** | 97,7 % | 99,1 % |
| **Organe** | 96,6 % | 98,3 % |
| **Présence d'un événement d'intérêt CC à +/- 1 mois du diagnostic registre** | 96,5 % | 98,9 % |
| • dont Biopsie de l'organe à +/- 1 mois | 44.7 % | 46.1 % |
| • dont Biopsie non attribuable à un organe à +/- 1 mois | 74.7 % | 76.7 % |
| • dont Endoscopie de l'organe à +/- 1 mois | 23.6 % | 24.3 % |
| • dont Exérèse de l'organe à +/- 1 mois | 32.8 % | 33.8 % |
| • dont Séjour PMSI de l'organe à +/- 1 mois | 67.5 % | 69.5 % |
| • dont Mise en ALD de l'organe à +/- 1 mois | 53.6 % | 55.3 % |

**Commune de résidence** : Concordance si la commune de résidence au diagnostique côté registre, correspond à l'une des communes enregistrées dans la CC entre 2010 et 2021.

**Date de décès** :
Si date de décès renseignée dans les 2 sources : concordance si date de décès registre à [-7;+7] jours de la date de décès CC
Si aucune date de décès côté CC ni côté registre : considérés concordants
Si date de décès côté CC et aucune date de décès côté registre : considérés concordants
Si date de décès côté registre et aucune date de décès côté CC : considérés discordants, sauf si le patient CC est rattaché à un registre hors rég. général

**Appareil / organe :** Concordance si l'appareil / l'organe côté registre (déduit de la CIM-O3) correspond à l'un des appareils / organes du patient côté CC (déduit de la CIM-10) entre 2010 et 2021

# Concordance : analyses croisées

| Concordance Sexe | Concordance Mois-Année de naissance | Concordance Département de résidence | Concordance Commune de résidence | Concordance Statut vital et date de décès | Concordance Appareil | Concordance Organe | Evénement d'intérêt CC à +/- 1 mois | Avant RD (% de tum. des patients appariés) | Après RD (% de tum. des patients appariés) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 92,0 | 95,2 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1,4 | 1,0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0,7 | 0,8 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0,8 | 0,7 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0,4 | 0,0 |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0,4 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 2,0 | 2,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0,5 | 0,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0,1 | 0,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0,2 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0,4 | 0,0 |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0,1 | |

1 = concordants / 0 = non concordants

# Concordance : analyses croisées

| Concordance Sexe | Concordance Mois-Année de naissance | Concordance Département de résidence | Concordance Commune de résidence | Concordance Statut vital et date de décès | Concordance Appareil | Concordance Organe | Evénement d'intérêt CC à +/- 1 mois | Avant RD (% de tum. des patients appariés) | Après RD (% de tum. des patients appariés) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 92,0 | 95,2 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1,4 | 1,0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0,7 | 0,8 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 1 | 1 | 0 | | | | |
| 1 | 1 | 1 | 1 | 1 | 0 | | | | |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0,4 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 2,0 | 2,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0,5 | 0,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0,1 | 0,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0,2 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0,4 | 0,0 |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0,1 | |

= Cas discordants date de décès, invalidés par les règles de décision

1 = concordants / 0 = non concordants

# Concordance : analyses croisées

| Concordance Sexe | Concordance Mois-Année de naissance | Concordance Département de résidence | Concordance Commune de résidence | Concordance Statut vital et date de décès | Concordance Appareil | Concordance Organe | Evénement d'intérêt CC à +/- 1 mois | Avant RD (% de tum. des patients appariés) | Après RD (% de tum. des patients appariés) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 92,0 | 95,2 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | | |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0,8 | 0,7 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0,4 | 0,0 |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0,4 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 2,0 | 2,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0,5 | 0,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0,1 | 0,1 |
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0,2 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0,4 | 0,0 |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0,1 | |

= Tumeurs 100% concordantes

1 = concordants / 0 = non concordants

# Concordance : analyses croisées

| Concordance Sexe | Concordance Mois-Année de naissance | Concordance Département de résidence | Concordance Commune de résidence | Concordance Statut vital et date de décès | Concordance Appareil | Concordance Organe | Evénement d'intérêt CC à +/- 1 mois | Avant RD (% de tum. des patients appariés) | Après RD (% de tum. des patients appariés) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 92,0 | 95,2 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1,4 | 1,0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0,7 | 0,8 |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0,8 | 0,7 |
| 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0,4 | 0,0 |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0,4 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0,0 | |
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 2,0 | 2,1 |
| 1 | 1 | 1 | 0 | | | | | | |
| 1 | 1 | 1 | 0 | | | | | | |
| 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0,1 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0,2 | 0,0 |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0,4 | 0,0 |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0,2 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0,1 | |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0,1 | |

= Tumeurs concordantes, sauf pour la commune de résidence (mais au sein du même département)

1 = concordants / 0 = non concordants

# Comparaison appariés / non appariés côté registres

| | Total | Appariés | Non appariés |
|---|---|---|---|
| **Sexe** | | | |
| Hommes | 476361 (54.5%) | 453773 (54.4%) | 22588 (55.7%) |
| Femmes | 398298 (45.5%) | 380333 (45.6%) | 17965 (44.3%) |
| **Statut vital** (aux dernières nouvelles) | | | |
| Vivant | 467065 (53.4%) | 443101 (53.1%) | **23964 (59.1%)** |
| Décédé | 407594 (46.6%) | 391005 (46.9%) | 16589 (40.9%) |
| **Classe d'âge au diagnostic** | | | |
| [18-30] | 8991 (1%) | 8123 (0.9%) | **868 (2.1%)** |
| [31-40] | 25195 (2.7%) | 23350 (2.6%) | **1845 (4.5%)** |
| [41-50] | 71532 (7.7%) | 67860 (7.7%) | 3672 (8.9%) |
| [51-60] | 160521 (17.3%) | 153827 (17.4%) | 6694 (16.2%) |
| [61-70] | 271141 (29.3%) | 261159 (29.5%) | 9982 (24.1%) |
| [71-80] | 228083 (24.6%) | 218958 (24.7%) | 9125 (22.1%) |
| [81-90] | 140000 (15.1%) | 132624 (15%) | **7376 (17.8%)** |
| [91+] | 21424 (2.3%) | 19623 (2.2%) | **1801 (4.4%)** |

# Taux d'appariement après règles de décision

| | Patient* Registre | Tumeur* Registre | Taux de matchs 1:1 avant règles de décision | Taux app <u>après</u> application des règles de décision |
|---|---|---|---|---|
| **TOTAL Registres généraux et spécialisés** | **874 659** | **926 887** | **99.1%** | **95.4%** |

* Rappel du cadre de l'appariement présenté : exclusion des cas pédiatriques (traités ad hoc), des hémopathies malignes (traitées ad hoc) et des tumeurs de la peau hors mélanome et hors sarcome de Kaposi de la peau

# Taux d'appariement par localisation

| Site de la tumeur solide | Taux d'appariement | Nombre de tumeurs total |
|---|---|---|
| Testicule | **97.9%** | 5 689 |
| Mésothéliome de la plèvre | **97.8%** | 2 486 |
| Larynx | **97.7%** | 8 417 |
| Ovaire | **97.6%** | 13 302 |
| **Sein** | **97.6%** | **152 036** |
| Corps de l'utérus | **97.1%** | 19 416 |
| **Côlon et rectum** | **97.1%** | **130 146** |
| Œsophage | **97.0%** | 16 059 |
| **Poumon** | **96.8%** | **108 423** |
| Pancréas | **96.7%** | 36 988 |
| Vésicule et voies biliaires | **96.4%** | 8 624 |
| Vessie | **96.3%** | 31 774 |
| Lèvre-bouche-pharynx | **96.2%** | 33 822 |
| **Prostate** | **96.1%** | **150 431** |
| Rein | **96.0%** | 34 964 |
| Estomac | **95.9%** | 20 711 |
| Foie | **95.7%** | 29 711 |
| Col de l'utérus | **95.6%** | 8 380 |
| Intestin grêle | **95.5%** | 4 420 |
| Système nerveux central (SNC) | **95.0%** | 12 665 |
| Thyroïde | **93.9%** | 22 952 |
| Autre site | **92.8%** | 27 579 |
| Sarcome | **91.2%** | 12 612 |
| Mélanome de la peau | **73.0%** | 35 280 |

# Limites

- **1 patient est apparié sur les données d'une tumeur**
  Dès qu'une des tumeurs d'un patient est matchée, le patient est considéré comme apparié. En pratique, la concordance a été évaluée distinctement sur les tumeurs matchées et les tumeurs non matchées des patients appariés.

- **Variabilité de la probabilité d'appariement d'un patient à l'autre**
  Par exemple, les patients décédés ont plus de chance d'être appariés que les non décédés car la date de décès est une information précise et très discriminante.

- **Ajout de tolérances sur les erreurs connues dans la CC**
  En retirant certaines variables dans les clés, pour permettre le match malgré les erreurs connues dans la CC, on autorise des discordances sur les patients appariés. Chacune de ces discordances a été étudiée avant de décider si elle invalidait ou non le match. Ces discordances sont justifiées par des règles métier.

- **Classifications CIMO et CIM10 non superposables**
  L'appariement a nécessité au préalable la construction d'une liste d'entités (organes / appareils) communes, introduisant des approximations pour permettre l'appariement.

# CONCLUSION

- Production d'une **base permanente** :
  - avec **~ 885 000 tumeurs confirmées et qualifiées** par les registres (CIM-O3 gold standard) appariées **avec leurs données longitudinales de consommations de soins** (CC-SNDS)
  - exhaustive sur la zone registre
  - **accessible sur l'EDS-RNC sur projet d'étude**

- Mise à jour annuelle avec envoi des cas N-3 par les registres locaux

**Perspectives / prochaines étapes :**

- Appariement des hémopathies malignes en cours
- Appariement pédiatrie à venir

- Cible : appariement direct avec le NIR

# DES QUESTIONS ?

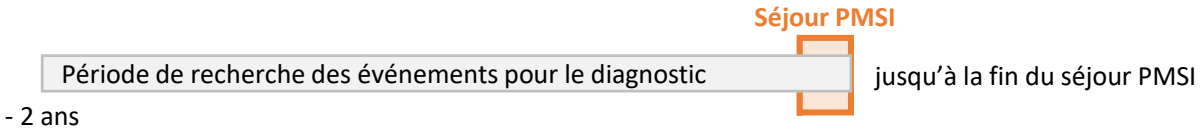
[APPARIEMENT]

# Annexes

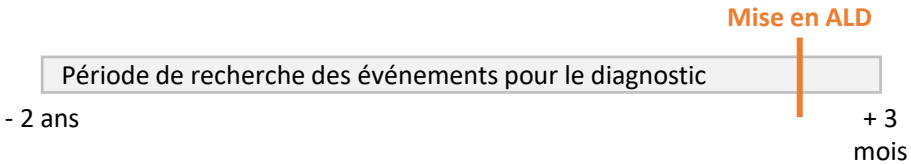
# Etape 3 – Sélection et préparation de la population CIBLE (= Cohorte Cancer)

Période de recherche des actes et événements d'intérêt dans la CC pour approcher la date de diagnostic :

**Cas 1** : si la localisation du cancer est repérée par un séjour PMSI en premier :




**Cas 2** : si la localisation du cancer est repérée par une mise en ALD ou invalidité en premier* :




**Cas 3** : si la localisation du cancer est repérée par la délivrance d'un médicament anti-cancéreux spécifique de la loc. :




**Ajout post-réunion :**

Cas 1 = **65,4 %** des localisations sont repérées par un séjour en PMSI en premier dans la CC

Cas 2 = **32,7 %** des localisations sont repérées par une mise en ALD / invalidité en premier dans la CC (avant tout séjour PMSI)

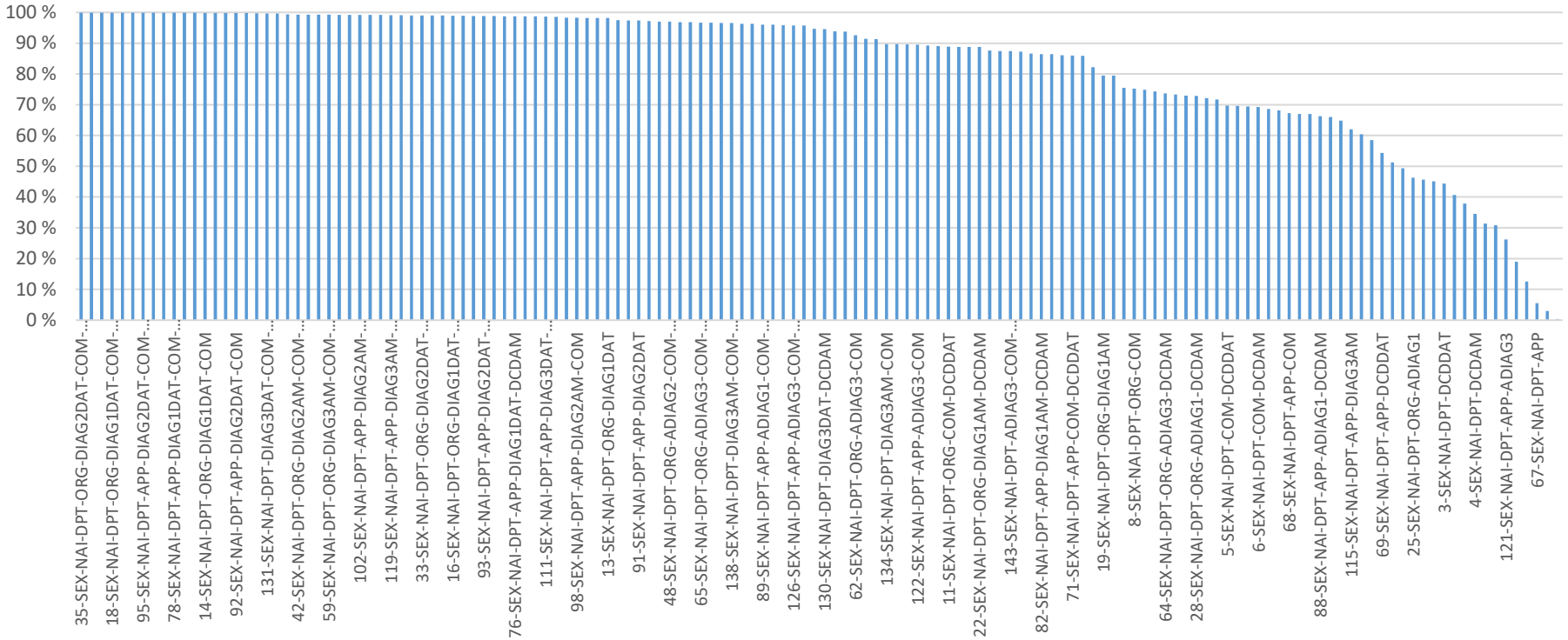
Cas 3 = **1,8 %** des localisations sont repérées par une délivrance d'un médicament anti-cancéreux spécifique en premier dans la CC (avant tout séjour PMSI ou mise en ALD / invalidité)

**→ L'ensemble des dates extraites sur la période sont proposées à l'appariement, par ordre de priorité, comme dates possibles pour correspondre à la date de diagnostic côté registre.**

* Existence de cas particuliers traités de manière spécifique

## Etape 5 – Mise en œuvre de l'appariement

# Pourcentage de valeurs uniques, par clé, dans les données CC



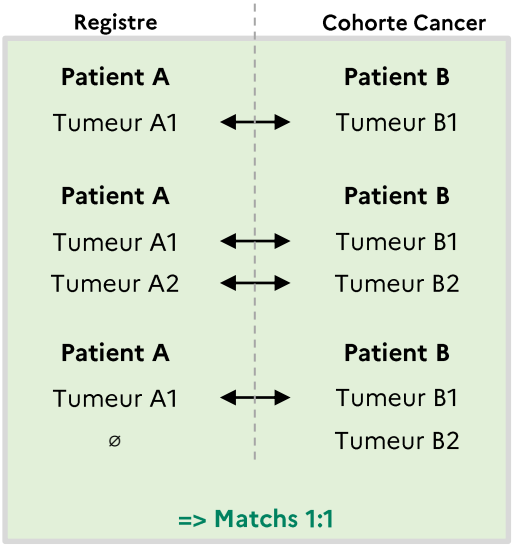
## Etape 5 – Mise en œuvre de l'appariement

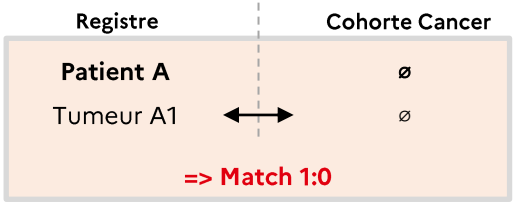
# Qualification du type de match

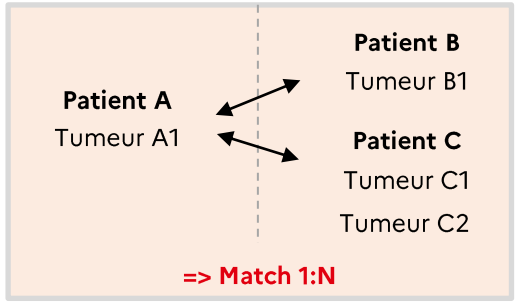
Principe : on apparie des patients via leurs tumeurs.

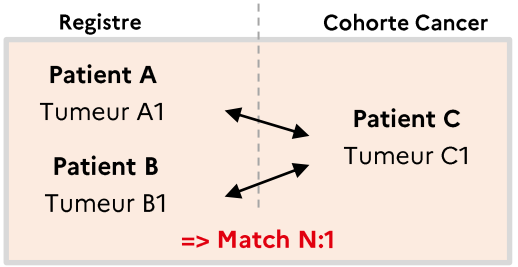
**Qualification du type de match :**

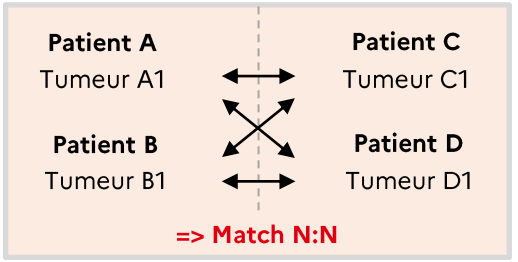
La qualification du type de match s'évalue clé par clé : un doublon sur une clé d'appariement, peut être un 1:1 sur une clé plus précise.



# Evènements d'intérêt dans les +/- 1 mois du diagnostic registre

| Appareil | Biopsie N (%) | Exérèse N (%) | Endoscopie N (%) | Séjour N (%) | ALD N (%) |
|---|---|---|---|---|---|
| APP. DIG. | 186617 (78,6%) | 93871 (39,6%) | 155608 (65,6%) | 209735 (88,4%) | 121801 (51,3%) |
| APP. NON ATTRIB. | 7972 (62%) | 47 (0,4%) | 22 (12,6%) | 4415 (34,3%) | 583 (4,5%) |
| APP. RESPI | 92071 (87,8%) | 14728 (14,1%) | 66003 (63%) | 90460 (86,3%) | 47570 (45,4%) |
| GANGL. SEC. | 16 (76,2%) | 10 (47,6%) | 0 (0%) | 18 (85,7%) | 3 (30%) |
| GLAND. ENDO. | 19055 (87,8%) | 18500 (85,3%) | 599 (2,8%) | 18178 (83,8%) | 9280 (42,8%) |
| OEIL | 364 (25,7%) | 537 (37,9%) | 0 (0%) | 1145 (80,9%) | 683 (48,2%) |
| ORG. GEN. FEM. | 35317 (81,3%) | 18969 (43,7%) | 12477 (28,7%) | 31959 (73,6%) | 24512 (56,4%) |
| ORG. GEN. MASC. | 136633 (92,6%) | 24116 (16,3%) | 111 (0,1%) | 65364 (44,3%) | 88624 (60,1%) |
| OS | 1026 (86,9%) | 257 (21,8%) | 1 (0,1%) | 828 (70,2%) | 468 (39,7%) |
| PEAU | 22401 (90,9%) | 19219 (78%) | 1 (0%) | 7690 (31,2%) | 15070 (61,1%) |
| SEIN | 142269 (96,6%) | 46723 (31,7%) | 0 (0%) | 65578 (44,5%) | 118864 (80,7%) |
| SYST. NERV. | 9044 (75,4%) | 5639 (47%) | 15 (0,1%) | 11278 (94%) | 6565 (54,7%) |
| TISS. MOUS | 8497 (85,6%) | 4002 (40,3%) | 1703 (18,3%) | 6522 (65,7%) | 3442 (34,7%) |
| VADS | 33482 (83%) | 12957 (32,1%) | 30630 (75,9%) | 33625 (83,4%) | 22296 (55,3%) |
| VOIES URI. | 48422 (79,2%) | 47388 (77,5%) | 12606 (20,6%) | 57807 (94,6%) | 24988 (40,9%) |

# Comparaison appariés / non appariés côté registres

**Année de diagnostic**

| | Total | Appariés | Non appariés |
|---|---|---|---|
| 2010 | 72927 (7.9%) | 68476 (7.7%) | **4451 (10.8%)** |
| 2011 | 74806 (8.1%) | 71291 (8.1%) | 3515 (8.5%) |
| 2012 | 74692 (8.1%) | 71358 (8.1%) | 3334 (8.1%) |
| 2013 | 75423 (8.1%) | 72121 (8.1%) | 3302 (8.0%) |
| 2014 | 77316 (8.3%) | 74055 (8.4%) | 3261 (7.9%) |
| 2015 | 79291 (8.6%) | 75988 (8.6%) | 3303 (8.0%) |
| 2016 | 81837 (8.8%) | 78493 (8.9%) | 3344 (8.1%) |
| 2017 | 81661 (8.8%) | 77945 (8.8%) | 3716 (9.0%) |
| 2018 | 83158 (9.0%) | 79436 (9%) | 3722 (9.0%) |
| 2019 | 83411 (9.0%) | 80023 (9%) | 3388 (8.2%) |
| 2020 | 76206 (8.2%) | 73105 (8.3%) | 3101 (7.5%) |
| 2021 | 66159 (7.1%) | 63233 (7.1%) | 2926 (7.1%) |

**Base de diagnostic**

| | Total | Appariés | Non appariés |
|---|---|---|---|
| clinique et para-clinique *(rayon X, ultrason, exploration chir.)* | 51200 (5.5%) | 47090 (5.3%) | **4110 (9.9%)** |
| clinique seule | 374 (0%) | 322 (0%) | 52 (0.1%) |
| cytologie | 13101 (1.4%) | 12532 (1.4%) | 569 (1.4%) |
| histologie sur métastase | 44297 (4.8%) | 42776 (4.8%) | 1521 (3.7%) |
| histologie sur primitif | 812189 (87.6%) | 777443 (87.8%) | 34746 (84%) |
| inconnu | 185 (0%) | 161 (0%) | 24 (0.1%) |
| marqueurs tumoraux spécifiques | 5541 (0.6%) | 5200 (0.6%) | 341 (0.8%) |

* Rappel du cadre de l'appariement présenté : exclusion des cas pédiatriques (traités ad hoc), des hémopathies malignes (traitées ad hoc) et des tumeurs de la peau hors mélanome et hors sarcome de Kaposi de la peau

# Description des tumeurs matchées côté CC

| Comportement | % des tumeurs matchées |
|---|---|
| Invasif | **98,7 %** |
| Tumeur à évolution imprévisible (TEI) | **0,7 %** |
| In situ | **0,3 %** |
| Méta seul | **0,2 %** |
| In situ + TEI | **0,04 %** |
| Ganglion seul | **0,04 %** |
| Antécédent personnel seul | **0,02 %** |

| Source de la 1ère occurrence de la localisation | % des tumeurs matchées |
|---|---|
| ALD | **47,4 %** |
| PMSI | **46,3 %** |
| ALD + PMSI | **5,2 %** |
| Médicament anti-cancéreux spécifique de la localisation | **1,0 %** |
| ALD + méd. | **0,1 %** |
| PMSI + méd. | **0,0 %** |
| ALD + PMSI + méd. | **0,0 %** |

# Etape 2 –
# Préparation des données des registres pour l'appariement tumeurs solides

**Transformation CIM-O3 → organe / appareil** ← CIM-10 :

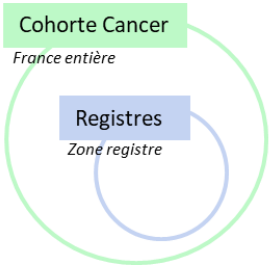
**→ Nous avons fait converger CIM-O3 (et CIM-10) vers une liste de 15 appareils.**

| Organes | APPAREIL |
|---|---|
| Cav. Orale, Fosses nas., sinus, oreille, Gland. Saliv., Langue, Larynx, Lèvre, Pharynx, VADS sans précis. | VADS |
| App. Dig. sans précis., Colon-Rectum-Anus, Estomac, Foie et voies bil., Intestin grêle, Oesophage, Pancréas, Peri.-Retroperi. | APP. DIG. |
| App. Respi. sans précis., Médiastin, Plèvre, Trach., Bronch., Poum. | APP. RESPI |
| Os, artic., Cartil. Artic. | OS |
| Peau | PEAU |
| Nerfs Périph., Syst. Nerv. Centr., Syst. Nerv. sans précis. | SYST. NERV. |
| Tiss. Mous | TISS. MOUS |
| Sein | SEIN |
| Col utérus, Corps utérus, Org. Gen. Fem. sans précis., Ovaire, Vulve et vagin | ORG. GEN. FEM. |
| Org. Gen. Masc. sans précis., Prostate, Pénis, Testicule | ORG. GEN. MASC. |
| App. Uri. sans précis., Rein, Vessie, Voies Uri. Hautes | VOIES URI. |
| Oeil | OEIL |
| Gland. Endo. sans précis., Surrénale, Thyroïde | GLAND. ENDO. |
| Org. non Attrib. | APP. NON ATTRIB. |
| Ganglions | GANGLIONS |

# Etape 3 – Sélection et préparation de la population CIBLE (= Cohorte Cancer)

**3.1/ Blocage de la population côté Cohorte Cancer**

- Sélection des patients résidant dans la zone registre entre 2010 et 2021 (= à au moins 1 date entre 2010 et 2021)
- Prise en compte des tumeurs invasives, in situ et TEI en raison des approximations de codage du comportement dans le PMSI




**3.2/ Reconstitution de la localisation du cancer** (organe & appareil)

- Via codage CIM-10 dans le PMSI
- Via codage CIM-10 en ALD ou invalidité
- Via délivrance de médicament anti-cancéreux spécifique d'une localisation

## Etape 5 – Mise en œuvre de l'appariement

# Principe de l'appariement séquentiel

**Appariement séquentiel** = passage successif de plusieurs clés d'appariement

- Ordre de passage = de la clé la plus discriminante à la moins discriminante (% de valeurs uniques)
- Les clés d'appariement sont créées à partir des combinaisons de variables d'appariement.
- A chaque itération (= 1 clé d'appariement), les patients appariés de manière unique (1:1) sont retenus à l'appariement, et ne passent plus dans les prochaines itérations. Le reste des patients non encore appariés sont candidats pour la prochaine itération (= prochaine clé d'appariement).

