RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Institut National du Cancer

LA PRODUCTION DES INDICATEURS D'INCIDENCE NATIONALE ET INFRA NATIONALE, AUJOURD'HUI ET DEMAIN

Partie 2 : Demain

Sophie Houzard– Médecin de santé publique

Registre National des Cancers
inca Institut National du Cancer

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

Institut National du Cancer

Registre National des Cancers
Institut National du Cancer

GT Registres des cancers FRANCIM

CONCORDANCES ENTRE DONNÉES DES REGISTRES DU CANCER ET DE LA COHORTE-CANCER-SNDS : LE CAS DU CANCER DU PANCRÉAS

DÉFINIR ET VALIDER UN ALGORITHME D'IDENTIFICATION ET QUALIFICATION DES CAS

DANS LA COHORTE CANCER /LE SNDS

Pourquoi développer ces algorithmes?

- Aujourd'hui, pour mener des études sur la Cohorte Cancer (CC), extraction SNDS dans le Registre national des cancers), on utilise des algorithmes fondés sur des Règles « Métier ». **Mais** :
  - Quelques imprécisions, erreurs de codage et données manquantes
  - Algorithmes non évalués

- Avec les données gold standard des registres, appariées à la CC ➔ Identification possible des vrais cas de cancer de la CC, des cas « manquants », des faux cas de la CC dans la Zone Registre (ZR)

➔ Un algorithme, construit sur la ZR, avec un double objectif :

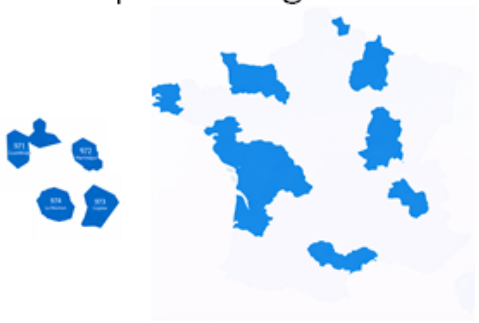
1. Identifier tous les cas SNDS en France pour sélectionner les populations d'intérêt des études descriptives et analytiques de trajectoires, de soins, de coûts… (quand on ne dispose pas des données RG)
2. Identifier les cas de la Zone Hors Registre pour produire les indicateurs épidémiologiques au niveau national : incidence, mortalité

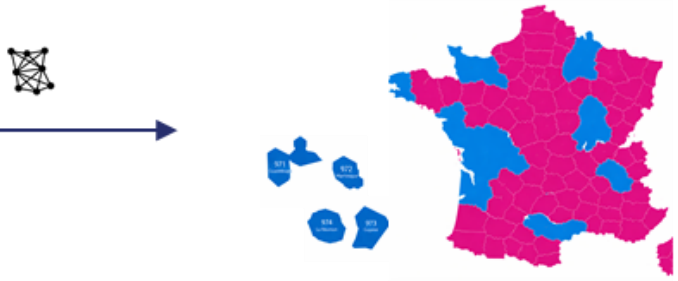
Objectif : Repérage des cas de la zone non couverte par un registre local

Registre National des Cancers

Aujourd'hui

Zones couvertes par les registres = Gold standard

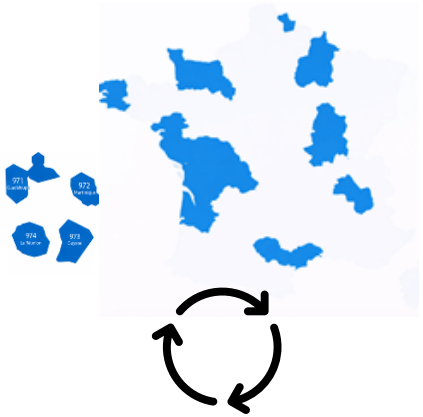




Modélisation statistique pour obtenir le nombre de cas France entière

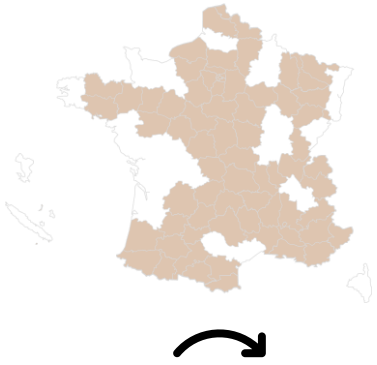
+ A moyen terme

Zones couvertes par les registres = Gold standard

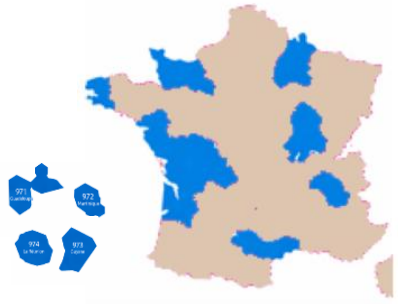


Elaboration de l'algorithme par confrontation des Données RL et CC appariées sur la Zone Registre

Cohorte cancer = France entière algo CC

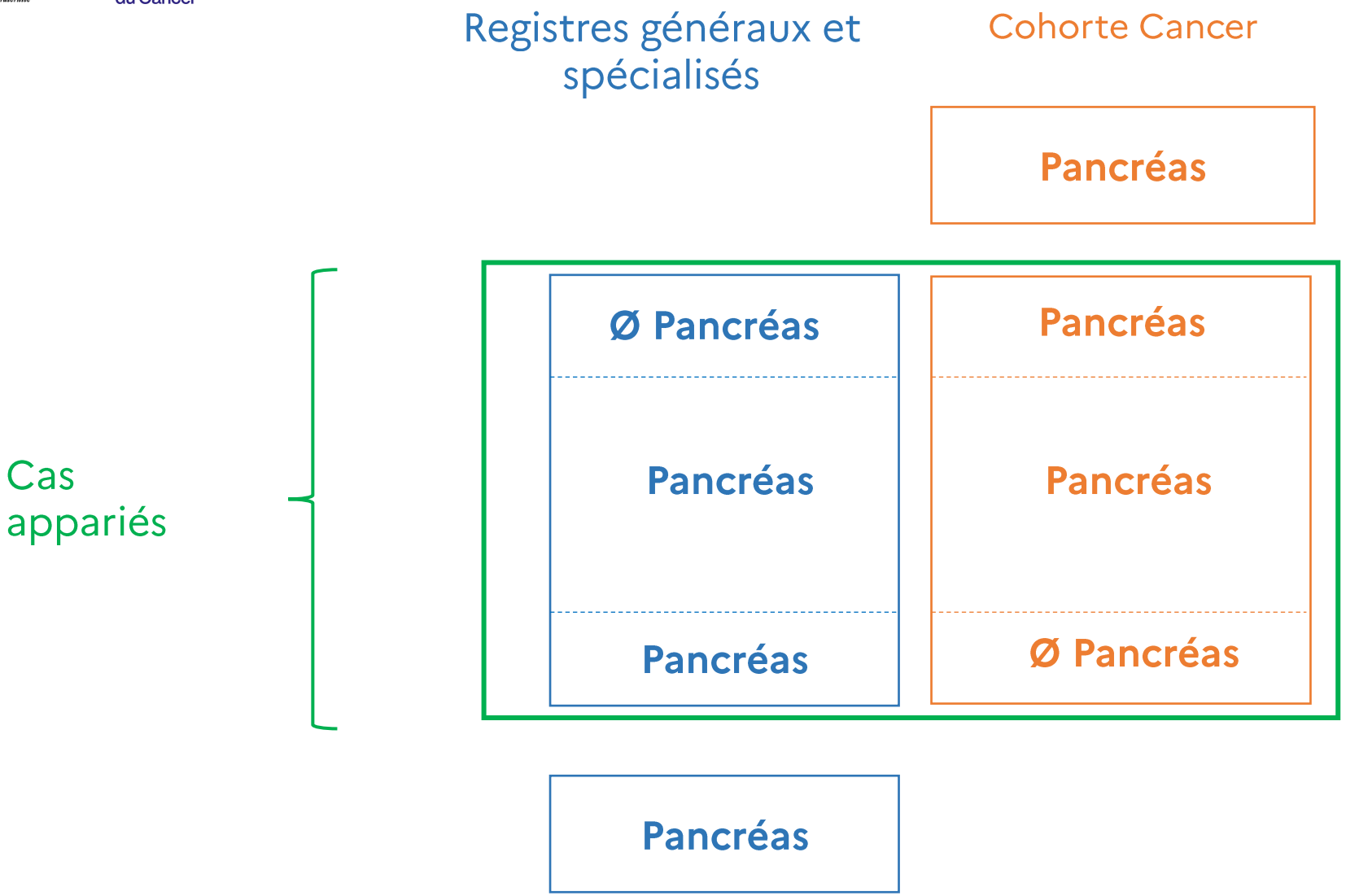


Application de l'algorithme sur la Zone Hors Registre



Méthode complémentaire pour obtenir les cas France entière

Le cancer du Pancréas dans la CC à l'épreuve du gold standard des registres



L'appariement séquentiel repose sur un ensemble de variables sans inclure systématiquement la localisation détaillée à l'organe du cancer

Motifs de non concordance des tumeurs

Registres généraux et spécialisés

Cohorte Cancer

Pancréas ⟷

=**Non appariés**
Tumeurs du registre non appariées (1:n, n:n)
= **Faux Positifs SNDS**
Erreurs de codage SNDS FP
Erreur comportement : tumeurs précancéreuses TIPMP/TMC
Erreur de localisation géographique
Vrais pancréas non appariés (Inaccessibilité source info pour le registre)

Ø Pancréas	**Pancréas** ⟷
Pancréas	**Pancréas** ⟷
Pancréas	**Ø Pancréas** ⟷

=**Faux Positifs SNDS**
Cancer pancréas codé à la place d'un cancer contigu (Foie, Voies biliaires, Estomac, Grêle)
Cancer pancréas + cancer contigu
Erreur appariement

=**Vrais Positifs SNDS**

=**Faux Négatifs SNDS**
Imprécision codage SNDS (ex: App Dig sans prec.)
Cancer autre codé à la place pancréas (Foie Voies biliaires, Estomac, Grêle)
Erreur appariement

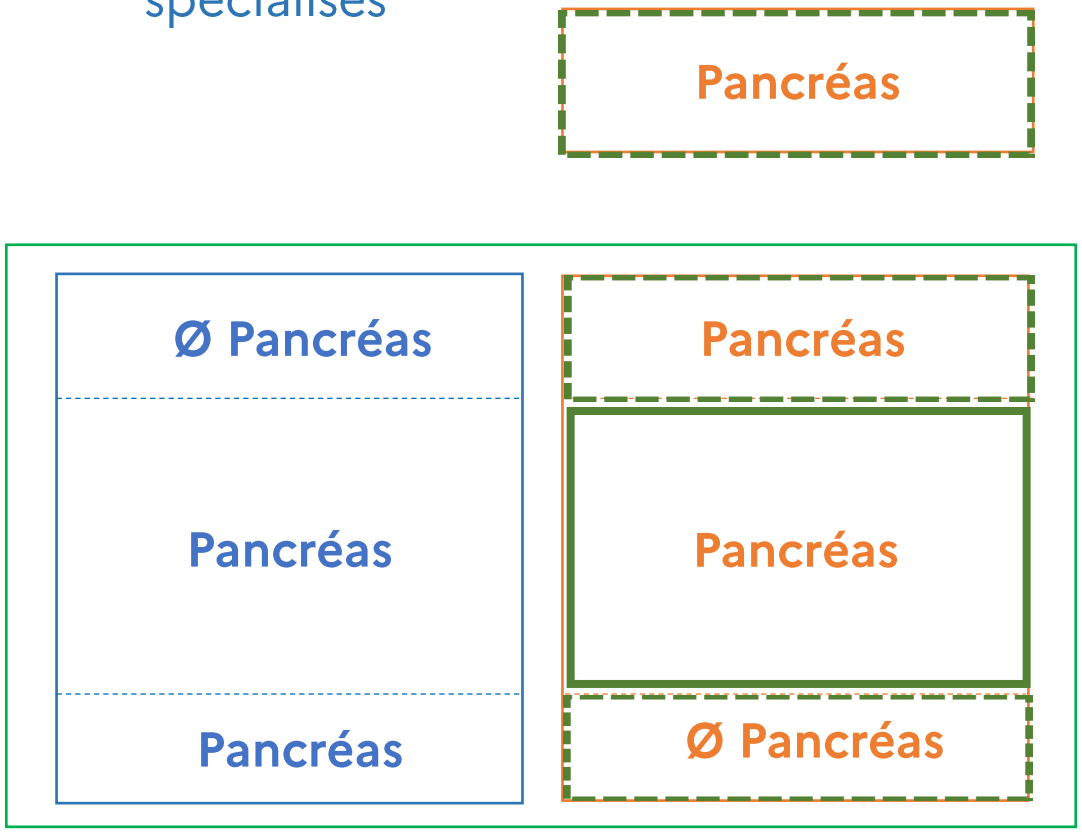
Pancréas ⟷

=**Non appariés**
Tumeurs du registre non appariées
= **Faux Négatifs SNDS**
Patient absent CC (ex: Pancréas sans ALD ni hospitalisation)

Motifs de non concordance des tumeurs

Registres généraux et spécialisés

Cohorte Cancer



=**Non appariés**
Tumeurs du registre non appariées (1:n, n:n)
= **Faux Positifs SNDS**
Erreurs de codage SNDS FP
Erreur comportement : tumeurs précancéreuses TIPMP/TMC
Erreur de localisation géographique
Vrais pancréas non appariés (Inaccessibilité source info pour le registre)

=**Faux Positifs SNDS**
Cancer pancréas codé à la place d'un cancer contigu (Foie, Voies biliaires, Estomac, Grêle)
Cancer pancréas + cancer contigu
Erreur appariement

=**Vrais Positifs SNDS**

=**Faux Négatifs SNDS**
Imprécision codage SNDS (ex: App Dig sans prec.)
Cancer autre codé à la place pancréas (Foie Voies biliaires, Estomac, Grêle)
Erreur appariement

=**Non appariés**
Tumeurs du registre non appariées
= **Faux Négatifs SNDS**
Patient absent CC (ex: Pancréas sans ALD ni hospitalisation)

Préalables aux comparaisons
Trois particularités influencent ces combinaisons

- Côté CC, sont sélectionnés les patients qui ont au moins 1 date de soins alors qu'ils résident dans une commune de la zone registre, même si cette consommation survient à distance du cancer (quelques absences de mise à jour des codes communes dans la CC)

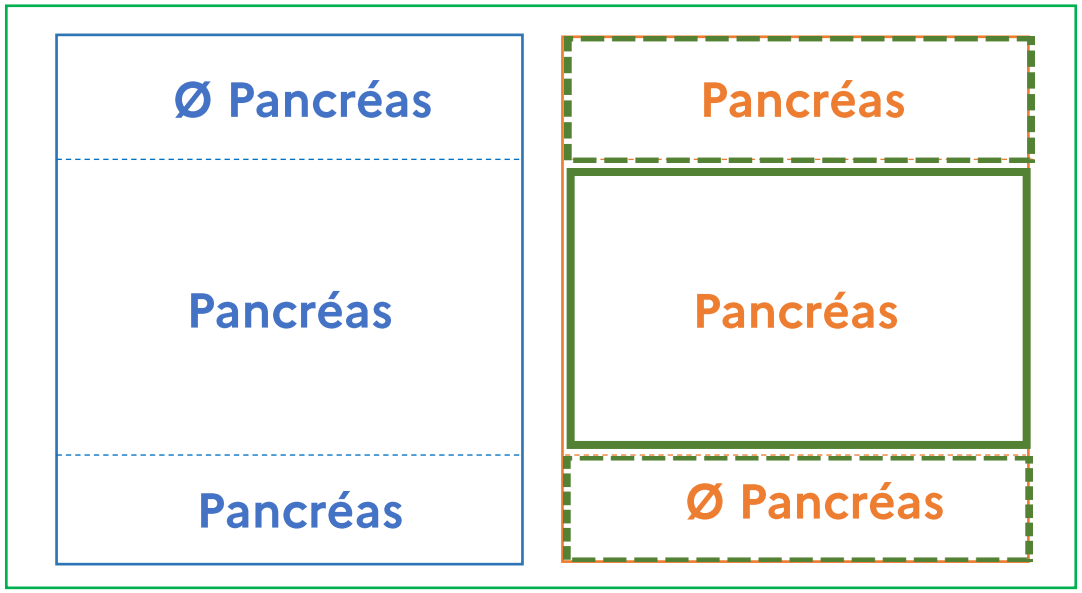
- La CC contient toutes les personnes qui ont une consommation cancer depuis 2010. Elles concernent donc des patients incidents mais également prévalents (distinction non réalisable), notamment dans les premières années observables 2010, 2011…

- Les patients pancréas CC non appariés comprennent des cas pancréas Registres non appariés par insuffisance de précision dans leurs données générant des doublons et inversement

Préalables aux comparaisons

Registres généraux et spécialisés

Cohorte Cancer



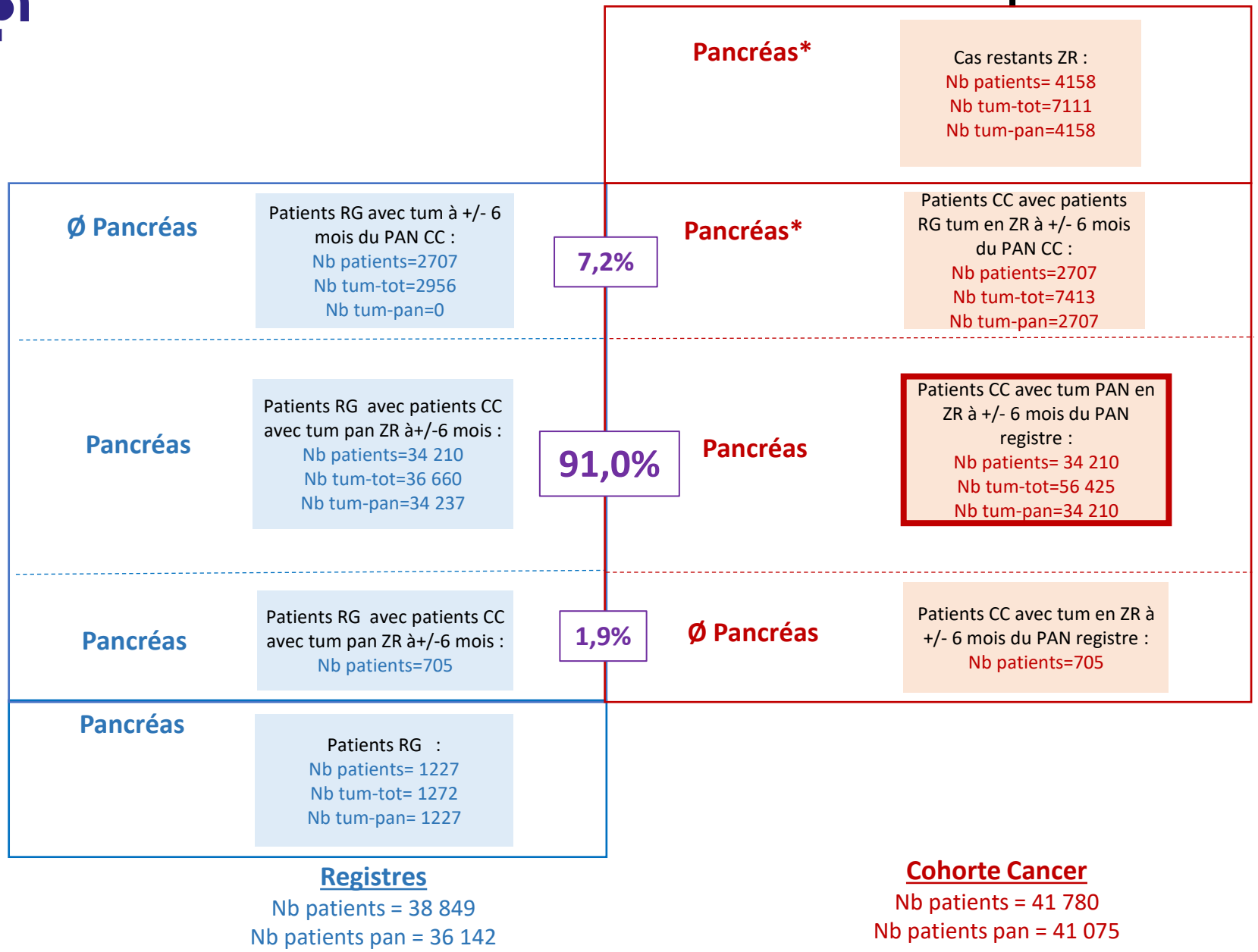
EXCLURE
Rechute FP
Pancréas prévalents pour les années 2010-2011 FP
Tumeurs CC non appariées mais appariables sur un set minimal de variables
Tumeurs Hors Zone Registre (HZR)

EXCLURE
Tumeurs HZR

EXCLURE
Tumeurs HZR

EXCLURE
Tumeurs RG non appariées mais appariables sur un set minimal de variables

Concordance pancréas en Zone Registre

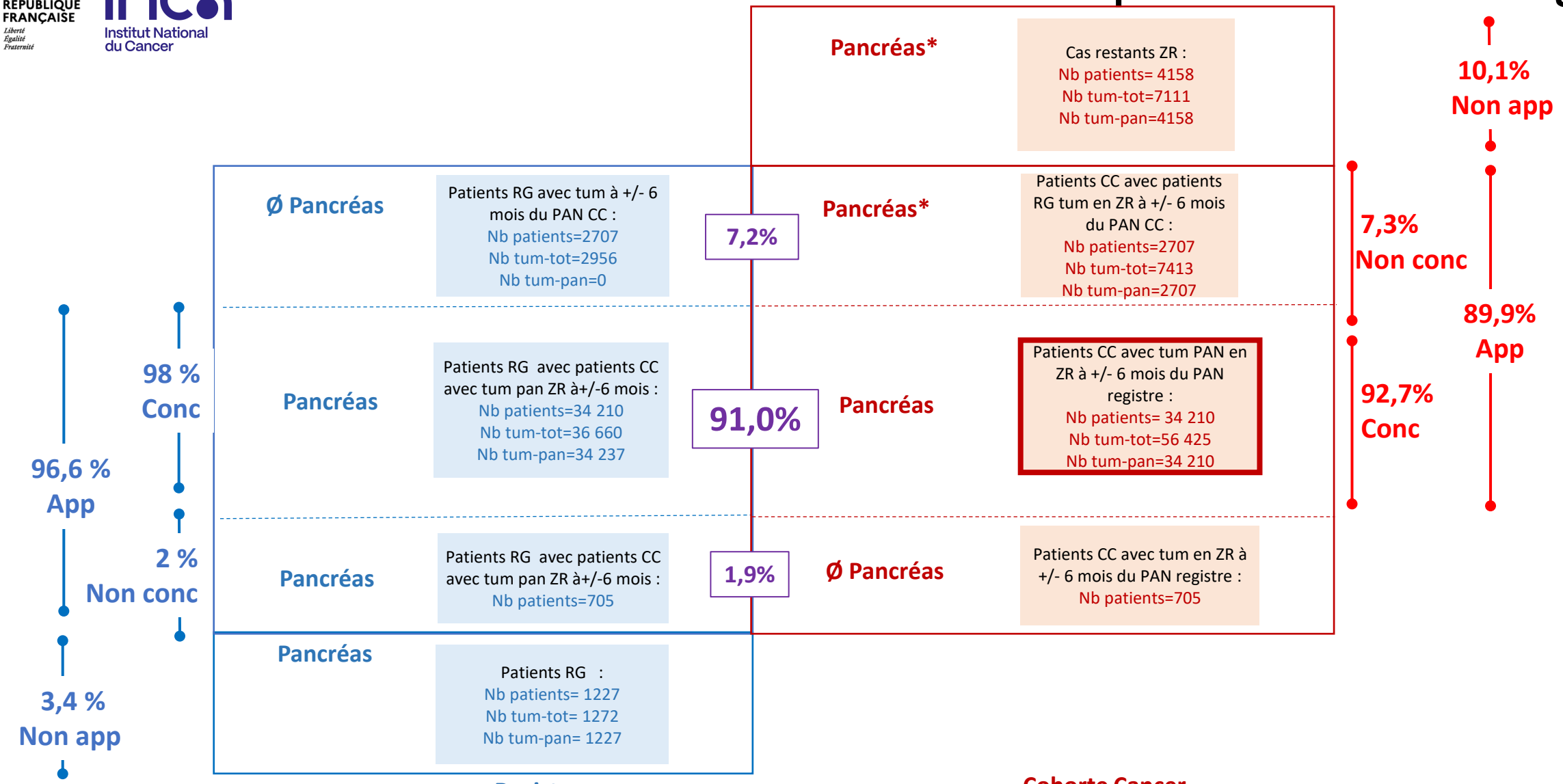


Registres
Nb patients = 38 849
Nb patients pan = 36 142

Cohorte Cancer
Nb patients = 41 780
Nb patients pan = 41 075

*ZR population filtrée aux identifiants résidant strictement en zone registre à +/- 1 an de la 1ère date CC du pancréas

Concordance pancréas en Zone Registre



Registres
Nb patients = 38 849
Nb patients pan = 36 142

Cohorte Cancer
Nb patients = 41 780
Nb patients pan = 41 075

*ZR population filtrée aux identifiants résidant strictement en zone registre à +/- 1 an de la 1ère date CC du pancréas

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

- Focus sur 7 caractéristiques :
  - Variables sociodémographiques : sexe, âge, survie
  - Caractéristiques de la tumeur : comportement, localisations associées, sources d'information de la localisation pancréas, traitements

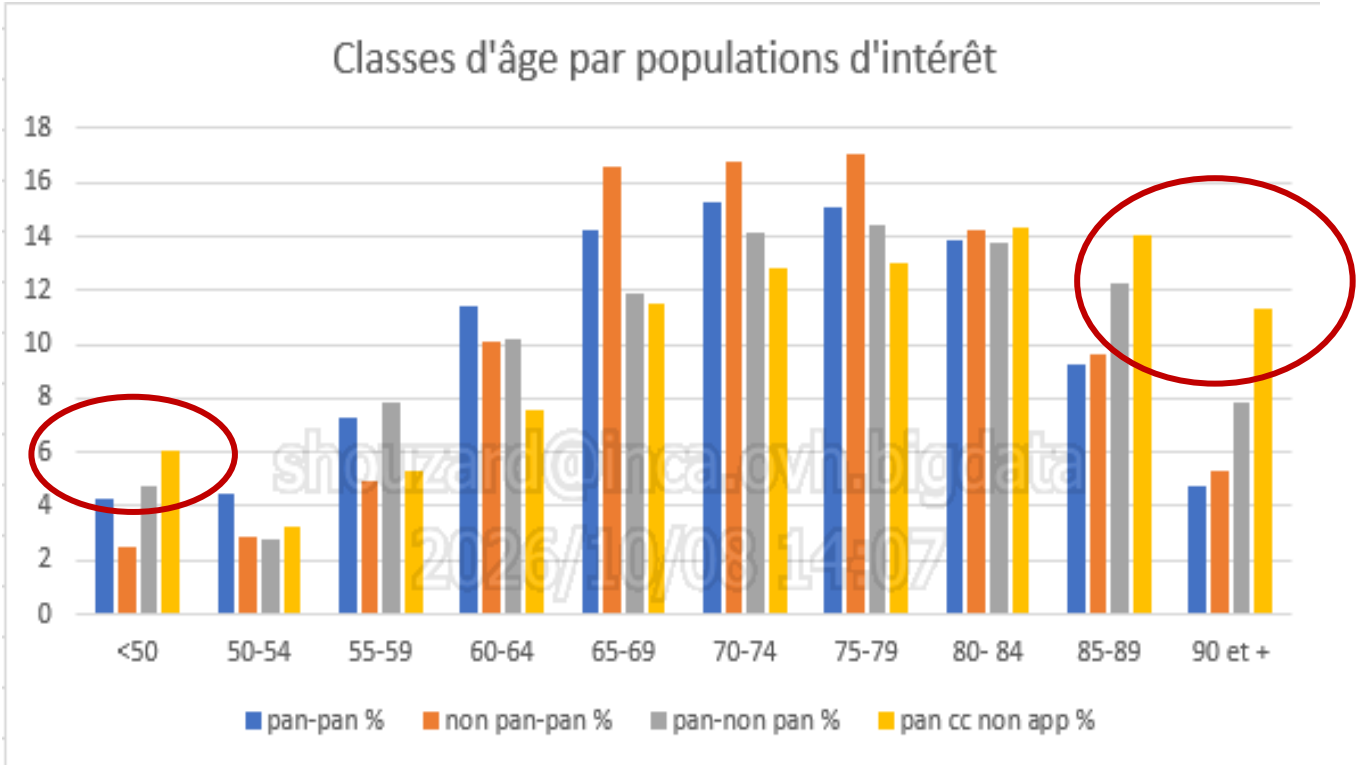
Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

Registres généraux et spécialisés

Cohorte Cancer

Registres généraux et spécialisés	Cohorte Cancer	N	Femmes	Age >90 ans	Survie à 1 an	Survie à 5 ans
	Pancréas	4 158	54% ➔ TMC?	11%	58% ➔TIPMP?	42%
Ø Pancréas	Pancréas	2 707	44%	5%	42%	19%
Pancréas	Pancréas	34 210	49%	5%	37%	9%
Pancréas	Ø Pancréas	705	46%	8%	35%	22%
Pancréas						

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?



Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

Registres généraux et spécialisés

Cohorte Cancer

		N	Organe concordant	Organe non concordant
	Pancréas	4 158	NA	NA
Ø Pancréas	Pancréas	2 707	82% (2 209)	18% (498)
Pancréas	Pancréas	34 210	100 %	-
Pancréas	Ø Pancréas	705	16% (110)	84% (595)
Pancréas				

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

Registres généraux et spécialisés

Cohorte Cancer

		N	Organe concordant	Organe non concordant
	Pancréas	4 158	NA	NA
Ø Pancréas	Pancréas	2 707	82% (2 209)	18% (498)
Pancréas	Pancréas	34 210	100 %	-
Pancréas	Ø Pancréas	705	16% (110)	84% (595)
Pancréas				

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

Organes Matchés chez les NON Pancréas RG - Pancréas CC N=2707		N(%)
Foie Voies Biliaires RG	- Foie Voies Biliaires CC Pancréas CC	1184 (44%)
Poumon RG	- Poumon CC Pancréas CC	234 (9%)
Colon Rectum RG	- Colon Rectum CC Pancréas CC	174 (6%)
Intestin grêle RG	- Intestin grêle CC Pancréas CC	153 (6%)
Estomac RG	- Estomac CC Pancréas CC	129 (5%)
Organe non Attrib RG	- Organe non Attrib CC Pancréas CC	107 (4%)

Pancréas CC codé à tort mais en association avec une tumeur retrouvée dans le RG

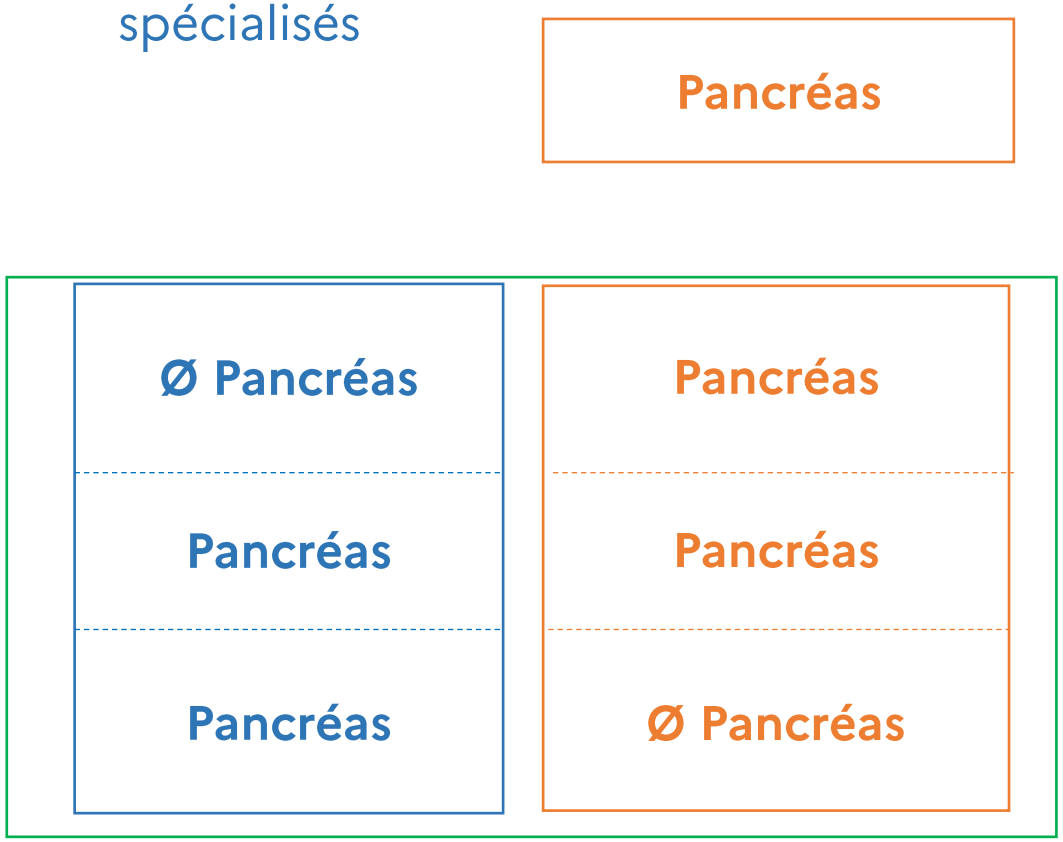
Organes Matchés chez les Pancréas RG - NON Pancréas CC N=705	N(%)
Pancréas RG – App. Dig. sans précision CC	228 (32%)
Pancréas RG – Foie Voies Biliaires CC	181 (26%)
Pancréas RG – Méta Foie Voies Biliaires CC	122 (26%)
Pancréas RG – Organe non Attrib. CC	90 (13%)
Pancréas RG – Colon Rectum CC	52 (3%)
Pancréas RG – Estomac CC	21 (3%)

Autre organe codé à tort à la place de pancréas

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

Registres généraux et spécialisés

Cohorte Cancer



N	Comportement	Tumeur bénigne D136-137
4 158	99% invasif	8%
2 707	100% invasif	1,6%
34 210	100% invasif	2,2%
705	59 % invasif 32% TEI* D37 Autres org. Dig. 6% métastatique FVB +/-perit.	4,7%

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

Registres généraux et spécialisés

Cohorte Cancer

	N	ALD seule	MCO	Exérèse	Chimio	Soins palliatifs
Pancréas	4 158	18,2%	74,5%	14%	20%	33%
Ø Pancréas	2 707	17,5%	75,7%	36%	52%	50%
Pancréas	34 210	0%	98,4%	23%	60%	51%
Pancréas	705	NA	NA	NA	NA	NA
Pancréas						
Pancréas						
Ø Pancréas						

Algorithmes à tester

Méthode 1 : Algorithme fondé sur des règles déterministes

Variables candidates

- Age
- Sexe
- ALD
- localisations associées d'autres cancers
- source PMSI (DP, DR, DAS, MCO, HAD, SMR)
- consommations de soins (Exérèse, CT, soins palliatifs)

Méthode 2 : Algorithme induit par des méthodes statistiques ou Machine Learning

Métriques
Sensibilité, Spécificité, VPP, VPN

En conclusion

- Du point de vue des Registres :
  - 98 % de Concordance « Organe » des tumeurs pancréas
  - 3,4% de cas non appariés

- Du point de vue de la CC :
  - 92,7 % de Concordance « Organe » des tumeurs pancréas
  - 10,1% de cas non appariés

- Des discordances qui cachent des réalités différentes :
  - Les non appariés :
    - Des erreurs de comportement dans la CC
    - Des ALD seules chez les patients âgés
  - Des erreurs dans le PMSI :
    - Cancer pancréas codé à la place d'un cancer contigu (Foie, Voies biliaires, Estomac, Grêle)/ appareil digestif sans précision et inversement
    - Cancer pancréas codé en plus d'un cancer contigu

Perspectives

- Des différences entre la population concordante des Pancréas RG-CC et les autres populations
  ➔ Un algorithme possible

- Le projet LOKAN avec les Registres Poitou- Charentes et Gironde sur les localisations Sein, Prostate, Colon-rectum, Poumon

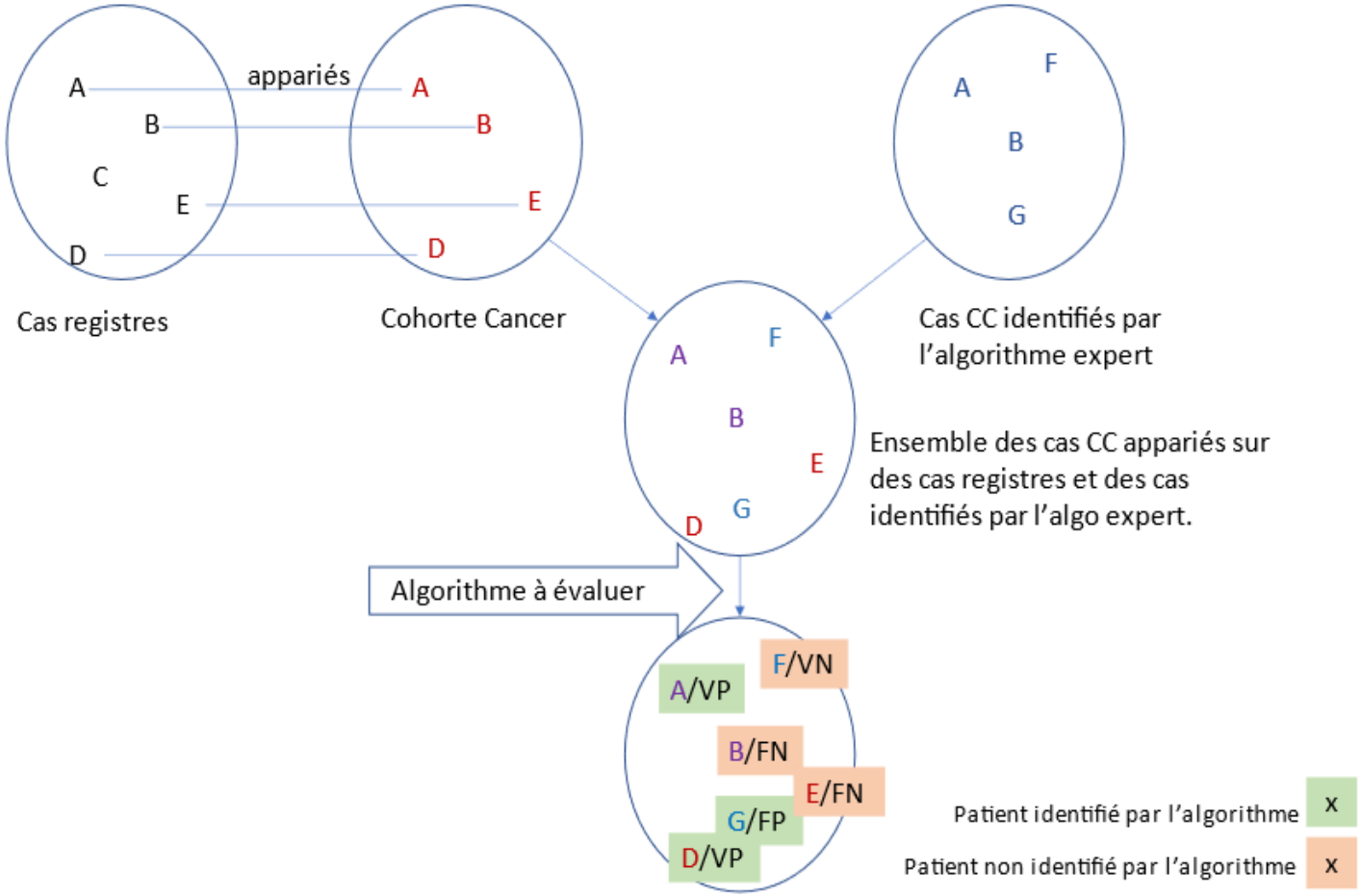
- L'appariement avec le NIR

- Intérêt des sources complémentaires ++ sur la zone non couverte par un registre

Nous remercions tous les acteurs participant à ce projet

Des questions?

Evaluation de la performance
Détermination des VP, FP, FN, VN



A/VP (Vrai Positif) Patient du registre, identifié correctement par l'algo (il était identifié antérieurement par l'algo expert)

D/VP (Vrai Positif) Patient du registre, identifié correctement par l'algo (il n'était pas identifié antérieurement par l'algo expert ; Exdans le cadre d'un algo qui recherche les cancers invasifs du sein *: cancer in situ dans le SNDS qui est en réalité un invasif)

B/FN (Faux Négatif) Patient du registre, non identifié par l'algo (il était identifié antérieurement par l'algo expert ; Ex* : L'algo à évaluer ne le retient pas car il ne retient que les cas MCO et non SSR par exemple)

E/FN (Faux Négatif) Patient du registre, non identifié par l'algo (il n'était pas identifié antérieurement par l'algo expert. Ex * : cancer situ dans le SNDS qui est en réalité un invasif)

F/VN (Vrai Négatif) Patient n'étant pas dans le registre et non identifié par l'algo (il était identifié antérieurement par l'algo expert ; Ex* : L'algo à évaluer ne le retient pas car i ne retient que les cas MCO et non SSR par exemple)

G/FP (Faux Positif) Patient n'étant pas dans le registre et identifié à tort par l'algoi(était identifié antérieurement par l'algo expert ; Ex* : un cancer *in situ* qui aurait été codé à tort dans le SNDS avec un code d'invasif)

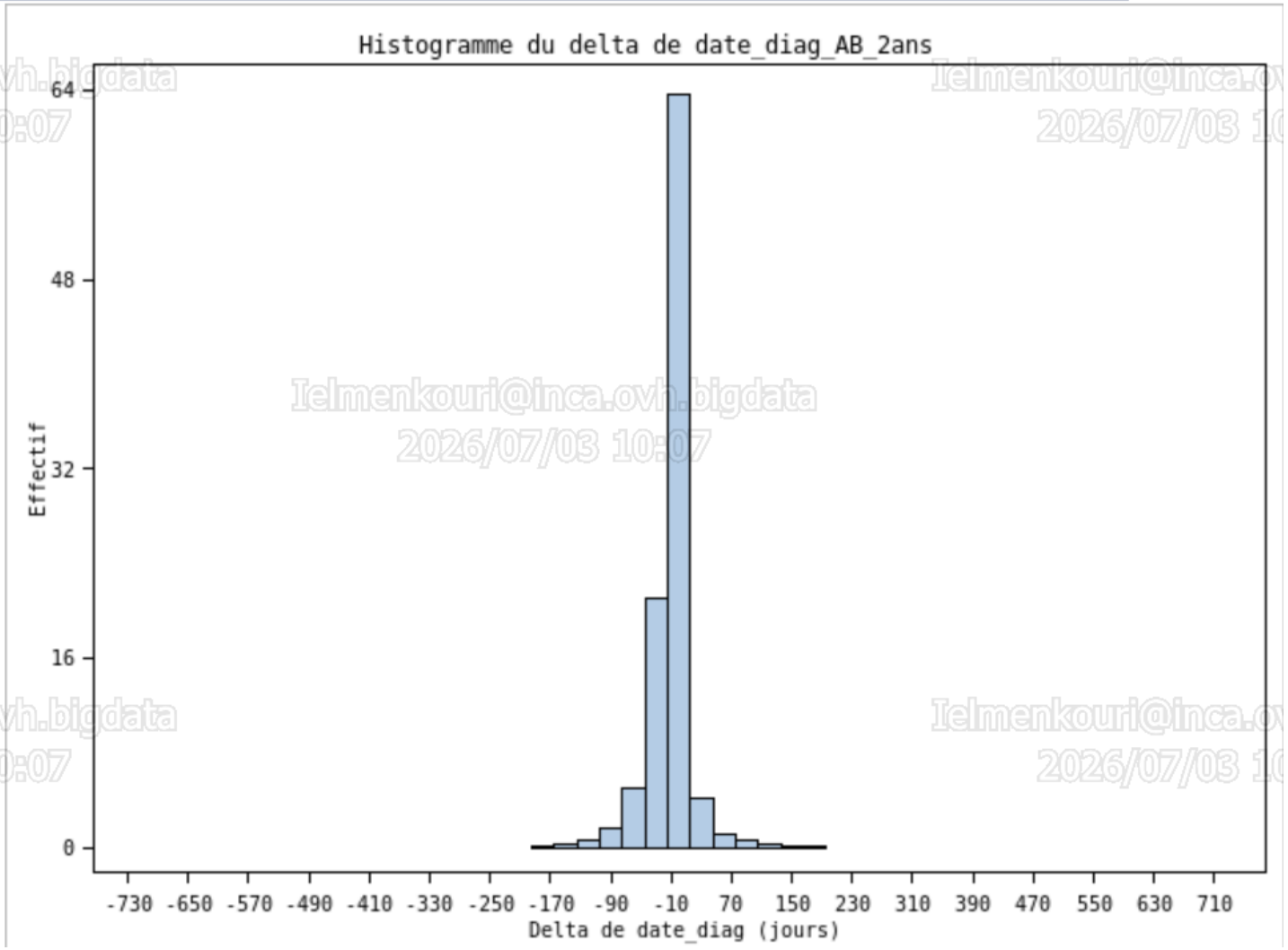
NB: exclure des cas de cancers incidents CC non appariés :

- ceux qui ne résident pas strictement dans la zone registre au moment de leur cancer
- ceux qui ressemblent aux quelques cas des registres qui n'ont pas trouvé leur match 1 :1

		Registre Cancer du pancréas incident en année N	
		Oui	Non
CC Personne retenue par l'algorithme	Oui	VP	FP
	Non	FN	VN

- **Sensibilité = VP / (VP + FN)** : proportion de malades correctement identifiés
- **Spécificité = VN / (VN + FP)** : proportion de non-malades correctement identifiés
- **Valeur prédictive positive (VPP) = VP / (VP + FP)** : probabilité qu'un patient soit réellement malade (ou porteur de la condition) **sachant que son test est positif**
- **Valeur prédictive négative (VPN) = VN / (VN + FN)** : probabilité qu'un patient soit réellement sain (ou exempt de la condition) **sachant que son test est négatif**. [1]

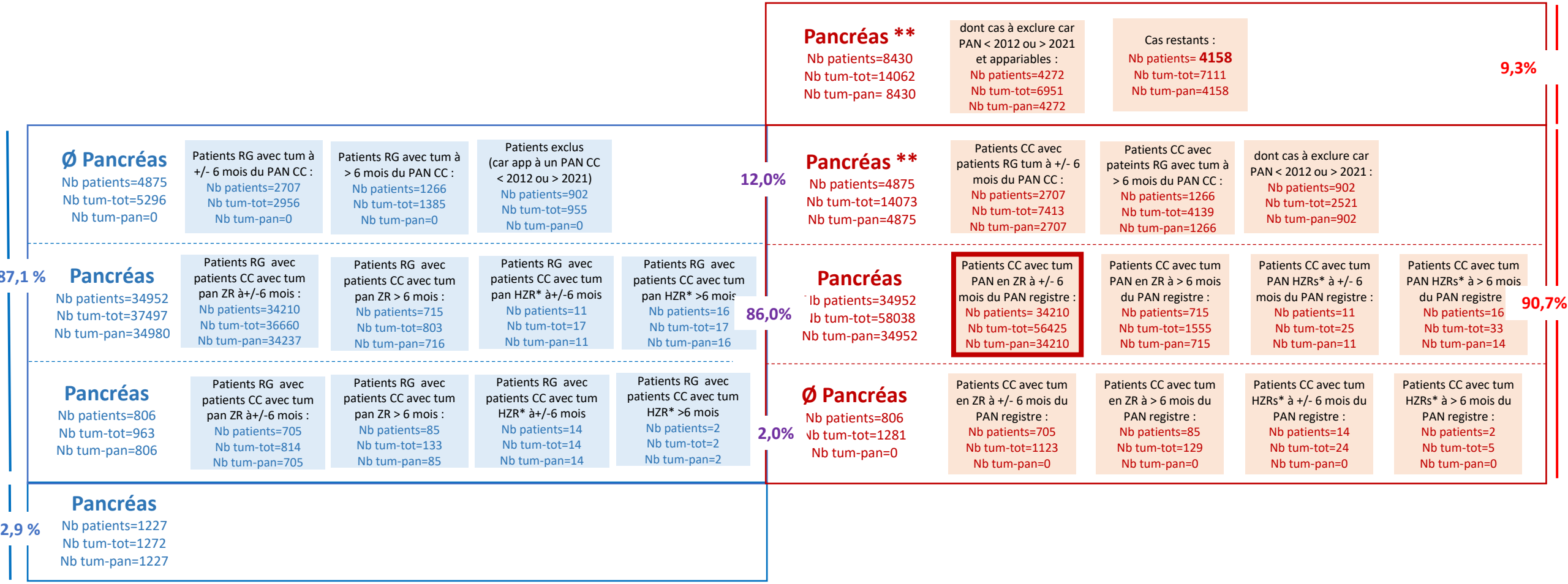
Pancréas_Pancréas



date_diag>date_min

pour faire le filtre période

Localisation pancréas



Registres
Nb patients = 41 860
Nb patients pan = 37 013

Cohorte Cancer
Nb patients = 49 063
Nb patients pan = 48 257

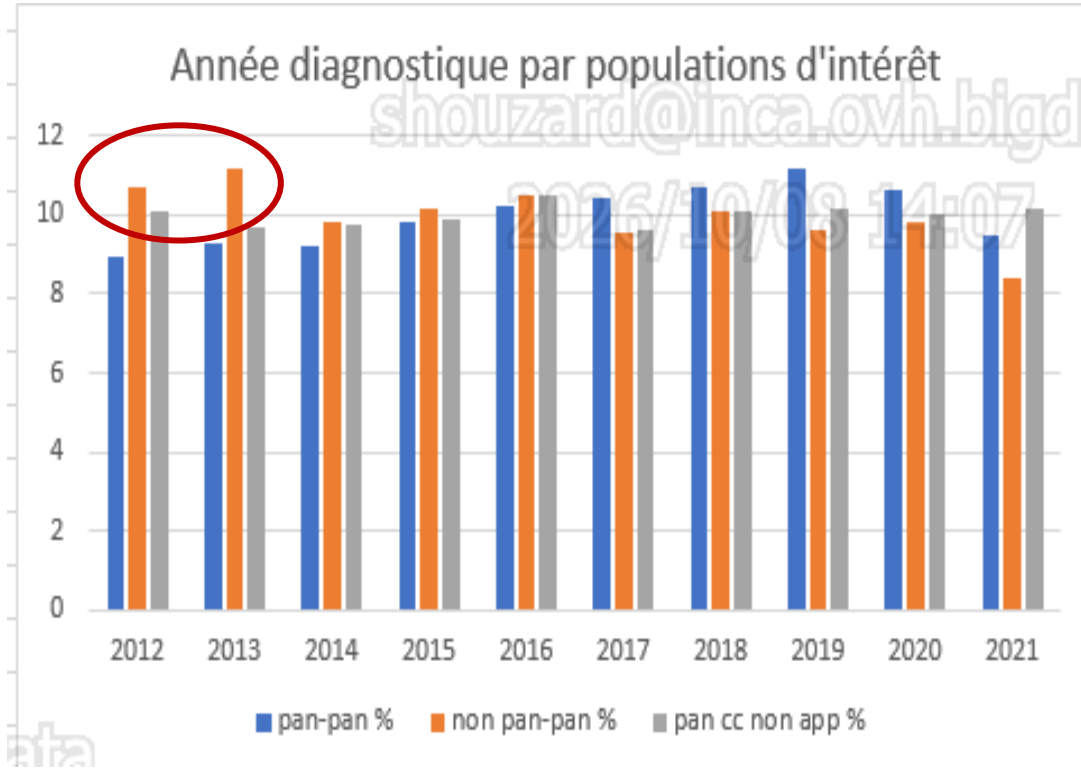
* : HZRs = patients résidant strictement hors zone registre (seulement codes communes hors zone registre lors de leurs consommations de soins à +/- 1 an de la min_dat CC
** : population filtrée aux identifiants résidant strictement en zone registre à +/- 1 an de la min_dat CC du pancréas

C25	Tumeur maligne du pancréas
C250	Tumeur maligne de la tête du pancréas
C251	Tumeur maligne du corps du pancréas
C252	Tumeur maligne de la queue du pancréas
C253	Tumeur maligne du canal pancréatique
C254	Tumeur maligne du pancréas endocrine
C254+0	Tumeur maligne du pancréas endocrine avec néoplasie endocrine multiple de type I [NEM 1]
C254+8	Tumeur maligne du pancréas endocrine, autre et non précisée
C257	Tumeur maligne d'autres parties du pancréas
C258	Tumeur maligne à localisations contiguës du pancréas
C259	Tumeur maligne du pancréas, sans précision
C259+0	Tumeur maligne du pancréas, familiale
C259+8	Tumeur maligne du pancréas, autre et non précisée
D136	Tumeur bénigne du pancréas
D137	Tumeur bénigne du pancréas endocrine

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

	PAN RG PANCC (LL1) N=34 210	PAN RG NON PAN CC (LOL1) N=705	NON PAN RG PAN CC (OLL1) N= 4 875	LCNAPP N=4 158
Dpt 971	0.9%	1.6%	3.5%	373 (9.0%)
ALD seule	0%		**18.6%**	**18.6%**
MCO+ALD	43.4%		12.4%	9.3%
MCO	31%		55.8%	52.7%
Exérèse	23%		36%	**14%**
Chimiothérapie	60%		52%	**20%**
Soins Palliatifs	51%		50%	**33%**

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?



Caractéristiques	TIPMP	TMC/MCN	Adénocarcinome
Nature	Précancéreuse	Précancéreuse	Cancer invasif
Age moyen	65-75ans	40-60ans	Env 70ans
Sexe	Plutôt masculin	Féminin	Plutôt masculin
Découverte fortuite	Très fréquente	Fréquente	Rare
Survie à 5 ans	> 95% non invasive	>90%	10-13%

[1] https://www.sciencedirect.com/science/article/abs/pii/S142439031631208X

Les populations discordantes diffèrent-elles des Pancréas RG-Pancréas CC ?

Organes Matchés chez les Pancréas RG - NON Pancréas CC N=705	N(%)
Pancréas RG – App. Dig. sans précision CC	228 (32%)
Pancréas RG – Foie Voies Biliaires CC	181 (26%)
Pancréas RG – Méta Foie Voies Biliaires CC	122 (26%)
Pancréas RG – Organe non Attrib. CC	90 (13%)
Pancréas RG – Colon Rectum CC	52 (3%)
Pancréas RG – Estomac CC	21 (3%)

Autre organe codé à tort à la place de pancréas

Organes Matchés chez les NON Pancréas RG - Pancréas CC N=2707		N(%)
Foie Voies Biliaires RG	- Foie Voies Biliaires CC Pancréas CC	1184 (44%)
Poumon RG	- Poumon CC Pancréas CC	234 (9%)
Colon Rectum RG	- Colon Rectum CC Pancréas CC	174 (6%)
Intestin grêle RG	- Intestin grêle CC Pancréas CC	153 (6%)
Estomac RG	- Estomac CC Pancréas CC	129 (5%)
Organe non Attrib RG	- Organe non Attrib CC Pancréas CC	107 (4%)

Pancréas CC codé à tort mais en association avec une tumeur retrouvée dans le RG