COMBICANCER-SEIN

Analyse des interactions entre comédications/comorbidités et le pronostic sur la cohorte quasi-exhaustive des patientes traitées pour un cancer du sein en France, issue du Système National des Données de Santé (SNDS)

Dossier de présentation du protocole au
Comité d'expertise pour les recherches, études et évaluations en santé

SOMMAIRE

1. INTRODUCTION

1.1. Présentation de l'équipe projet

Responsable de traitement	**Pr Fabien REYAL** Service de Chirurgie Sénologique Gynécologique et Reconstructrice UMR932 Immunité et Cancer. Laboratoire de recherche Résidu Tumoral et Résistance au Traitement (RT2Lab) Institut Curie 26 rue d'Ulm 75005 Paris fabien.reyal@curie.fr
Responsable de mise en oeuvre	**Dr Anne-sophie HAMY-PETIT** UMR932 Immunité et Cancer. Laboratoire de recherche Résidu Tumoral et Résistance au Traitement (RT2Lab) Institut Curie 26 rue d'Ulm 75005 Paris anne-sophie.hamy-petit@curie.fr
Responsable informatique et Data	**Julien GUERIN** Responsable de l'équipe Health Data Factory - Direction des Données Institut Curie 25 rue d'Ulm 75005 PARIS
Délégué à la protection des données	**Astrid LANG** Responsable Sécurité Système d'Information Déléguée à la Protection des données Institut Curie 25 rue d'Ulm 75005 PARIS astrid.lang@curie.fr
Doctorante	**Elise DUMAS** Ecole Doctorale de Cancérologie, Biologie, Médecine et Santé (CBMS 582) Université Paris Sud Laboratoire d'accueil : RT2Lab, Résidu tumoral et réponse au traitement, U932 Immunité et Cancer
Direction du doctorat	**Pr Fabien REYAL** Service de Chirurgie Sénologique Gynécologique et Reconstructrice UMR932 Immunité et Cancer. Laboratoire de recherche Résidu Tumoral et Résistance au Traitement (RT2Lab) fabien.reyal@curie.fr **Dr Chloé Agathe AZENCOTT** Centre de Biologie Computationnelle (CBIO), Mines ParisTech chloe-agathe.azencott@mines-paristech.fr

1.2. Contexte et justification de l'étude

Contexte scientifique

En 2018, 18,1 millions de nouveaux cas de cancers et 9,6 millions de décès par cancer ont été dénombrés dans le monde, mélanomes exclus [1]. Entre 40% [2] et 70% [3] des patients atteints de cancer ont de plus des pathologies associées au moment du diagnostic (comorbidités). Ces comorbidités peuvent être –ou non - associées à la prise de comédications chroniques pendant la prise en charge du cancer (désignées comme comédications dans l'ensemble du projet). Des preuves épidémiologiques ont rapporté une association entre certains médicaments tels que l'aspirine ou les anti-inflammatoires non-stéroïdiens [4] et une diminution du risque de développer un cancer. Les statines [5], les bêta-bloquants [3] et la metformine [6] ont aussi été associés à une baisse du nombre de cas de récidives et à une amélioration notable de la survie dans certains types de cancer. L'approche de chimioprévention est un sujet d'intérêt grandissant dans le champ de la cancérologie. Le **repositionnement de médicaments** consiste en particulier en l'utilisation de certaines molécules développées et prescrites pour certaines indications thérapeutiques dans le traitement d'autres maladies. Il a été considéré dans plusieurs conférences récentes comme une approche potentielle de prévention du cancer [7].

Le cancer du sein est le premier cancer chez les femmes, avec 2.1 millions de nouveaux cas estimés en 2018 [8]. Plusieurs études ont illustré l'influence des comorbidités/comédications sur 1) le risque de développer un cancer du sein [9], 2) sa présentation clinique et pathologique initiale [10], 3) les traitement reçus [11], 4) l'évolution à long terme [2], [12].

=> **Nous faisons l'hypothèse que de nombreux médicaments peuvent avoir un effet sur l'histoire naturelle du cancer du sein et peuvent modifier l'efficacité des traitements.**

Études préliminaires

En ré-analysant les données de l'essai randomisé REMAGUS02, notre équipe a récemment démontré que la prise quotidienne de celecoxib (inhibiteur de Cox2, classe des anti-inflammatoires non stéroïdiens) pendant la chimiothérapie néo-adjuvante pour un cancer du sein était associée à un sur-risque de rechute et de décès majeur dans certains sous-groupes de patients (tumeurs avec faible expression de PTGS2 et/ou n'exprimant pas les récepteurs aux estrogènes) (Hamy et al, fev 2019, Journal of Clinical Oncology) [13]. A la suite de ces travaux, nous avons effectué les **démarches de signalement d'effets indésirables** auprès du laboratoire Pfizer* commercialisant la molécule, ainsi que de l'ANSM (Agence Nationale de Sécurité du médicament). Nous avons été auditionnés par les responsables de la branche oncologie à l'ANSM le 11 septembre 2019. Les experts de l'ANSM ont jugé les données et le travail présentés comme suffisamment convaincants pour entamer une procédure auprès de l'EMA (European Medecines Agency), visant à émettre des ***recommandations contre-indiquant l'utilisation de ce médicament pendant les traitements de tous les cancers, quelle qu'en soit la localisation***. Une étude parallèle que notre équipe a mené en collaboration avec l'Institut National du Cancer (INCa) sur les données de la Plateforme de Données en Cancérologie démontre qu'environ 5% des patients atteints de cancer traité par chimiothérapie -tous cancers confondus, hors mélanomes- sont concernés par la prise de celecoxib au cours du traitement (**n=153 821)** entre 2011 et 2016). Une différence significative de **mortalité** (16% contre 10% dans la population de contrôle) confirme l'effet délétère du celecoxib en population générale.

Nous avons également ré-analysé les données de vie réelle de 1023 patientes traitées par chimiothérapie néoadjuvante pour un cancer du sein afin d'étudier les relations entre les comédications et la réponse au traitement évaluée par la réponse histologique complète. Près de la moitié des patientes ont déclaré utiliser une ou plusieurs comédications. Nous avons observé que la prise de certaines comédications (médicaments psychotropes, anti-acides, anti-hypertenseurs) permettait de doubler le taux de réponse à la chimiothérapie dans certains sous-groupes de malades. Des expérimentations animales ont permis de reproduire ces associations [12].

Un projet d'envergure (COMBIMMUNO) visant à valider la robustesse de ces associations sur des bases de données structurées d'essais cliniques nationaux, internationaux, ou des bases des données de vie réelle est actuellement en cours (EORTC 10994/BIG 1-00 n=1856, GeparSixto, GeparSepto n≈1794, PACS-08, PACS-09, n≈1000, données Institut Curie (n=15 000)).

> **=> Ces travaux démontrent le potentiel impact (bénéfique ou délétère) des comédications sur la réponse au traitement et l'évolution à long terme dans le cancer du sein. Ils justifient le besoin d'études complémentaires à ce sujet (validation sur des cohortes indépendantes, test sur d'autres molécules…).**
>
> **=> *L'objectif du projet COMBICANCER-SEIN est d'étudier les associations comédications/comorbidités/ cancer à une très large échelle, sur l'ensemble de la population française des patients atteints de cancer du sein entre 2007 et 2017, à partir de données de vie réelle issues du Système National des Données de Santé (SNDS).***

<u>Utilisation du SNDS</u>

En matière de santé, les données de vie réelle représentent une **source d'information encore sous-exploitée** à l'heure actuelle. En particulier, en France, la Sécurité Sociale génère à des fins administratives un volume d'informations gigantesque, agrégées dans une base de données structurée : le **Système National des Données de Santé (SNDS)**. Le SNDS réunit des données individualisées ***administratives*** exhaustives et actualisées de 98% de la population française (âge, sexe, régime d'affiliation, département de résidence, affections de longue durée (ALDs)), les données de remboursement de ***dépenses de santé en ville*** (actes, consultations, médicaments, ***dispositifs médicaux*** (DCIR)) et les données des ***hôpitaux*** (séjours, diagnostics, actes, activité externe du PMSI). Les ***dates de décès*** y sont notamment actualisées depuis 2007 pour le Régime Général. Les causes de décès sont disponibles via l'agrégation des données du Centre d'épidémiologie sur les causes médicales de Décès (CépiDC) de 2013 à 2015.

> **=> *En raison du nombre de patients qu'elle contient (dépassant la taille de toutes les cohortes françaises), en raison de l'exhaustivité de la population représentée, et en raison de la mise à jour* quasi *en temps réel de l'information, l'exploitation du SNDS à des fins de recherche constitue une opportunité exceptionnelle d'élargir le champ de recherche sur les comédications/comorbidités dans le cancer du sein.***

Cependant, les données du SNDS ont initialement une vocation administrative et de remboursement. De nombreux algorithmes d'extraction et d'identification de variables d'intérêt ont été testés en comparaison à des bases de données structurées [14]. Ils permettent de valider l'utilisation de certaines variables de cette base administrative à visée de recherche médicale et scientifique : date et type de traitements anti-cancéreux [15], extraction et regroupement des comorbidités [16]... Concernant le cancer du sein, les données du SNDS contiennent peu d'indicateurs pronostiques (absence de résultats d'analyse, absence de renseignements sur la biologie tumorale et l'histologie…) mais certains éléments peuvent être des **indicateurs indirects** (ex : les cancers du sein présentant une positivité pour les récepteurs hormonaux (RH) sont traités par hormonothérapie ; les cancers du sein HER2+ sont traités par trastuzumab). Par ailleurs, même s'il n'existe pas d'**indicateur structuré permettant d'identifier la survenue d'une rechute,** celle-ci s'accompagne d'un faisceau d'arguments suffisamment forts, détectables au niveau du système de remboursement, pour pouvoir être identifiée.

Notre équipe a récemment analysé la **concordance de certains indicateurs de cancer du sein** sur la base institutionnelle de l'Institut Curie et les données SNDS correspondantes appariées (étude en cours de publication) :

- Nous avons tout d'abord développé un **algorithme permettant l'identification de la survenue de récidives** par la reprise de traitements anticancéreux à la suite d'une période d'interruption de plusieurs mois survenue après le traitement du cancer du sein incident. Une période d'interruption de 5 mois a été retenue après analyses (voir Figure 1).

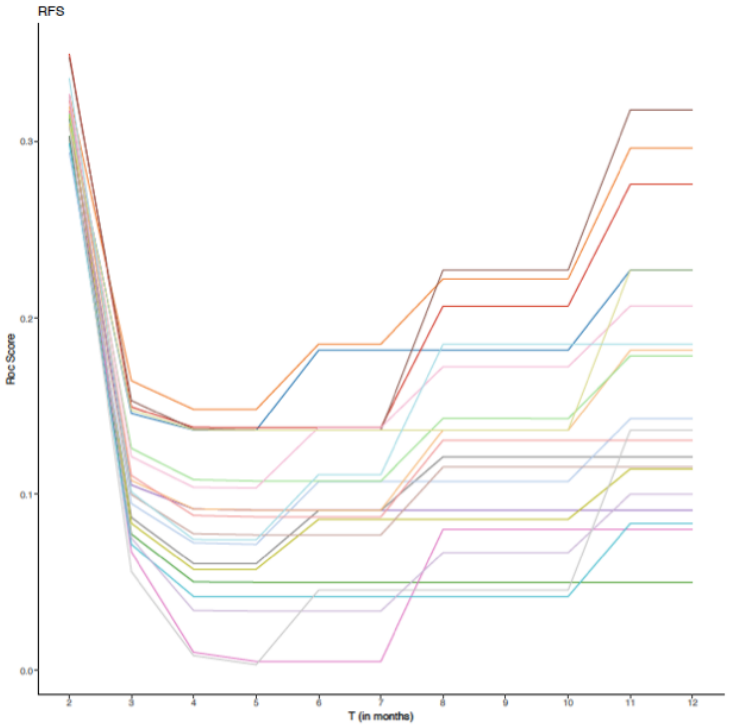

Figure 1** : Score ROC de l'algorithme d'identification des récidives en fonction de la période d'interruption T (en mois) pour toute rechute (décès inclus, hors second cancer) (RFS : Relapse Free Survival). 20 échantillons indépendants avec répétition possible ont été générés (méthode de bootstrapping) à partir d'une cohorte de 4 134 patientes de la base institutionnelle de l'Institut Curie appariée aux données du SNDS.* ***L'algorithme le plus performant est obtenu pour une période d'interruption de 5 mois.

=> Cet algorithme a démontré d'excellentes performances pour détecter les rechutes, qui sont présentées dans le tableau ci-dessous (*n = 4134*) :

	Sensibilité	Spécificité
Toute rechute (décès inclus, hors second cancer)	0.88	0.99
Rechute métastatique (décès inclus)	0.94	1

***Tableau 1** : Sensibilité et spécificité de l'algorithme d'identification des récidives de cancer du sein à partir des données du SNDS. L'algorithme est évalué à partir des données annotées de la base institutionnelle de l'Institut Curie appariée au SNDS (n= 4134). La période d'interruption retenue est de 5 mois.*

- Nous avons également étudié la **concordance entre les comédications déclarées dans le dossier médical et les médicaments délivrés en pharmacie.** A partir d'un échantillon de la base institutionnelle de l'Institut Curie apparié au SNDS (*n=981*), nous avons observé que la concordance variait avec la classe thérapeutique des molécules. Si une concordance satisfaisante était retrouvée pour la majorité des molécules, on observe cependant une **sous-déclaration significative de certaines comédications dans les dossiers médicaux des patients** (anti-acides, supplémentations vitamino-calciques..). Les données du **SNDS** permettent par ailleurs de **détecter des situations supposées d'adhérence infra-optimale au traitement** (quantité délivrée inférieure à la quantité théorique prescrite).

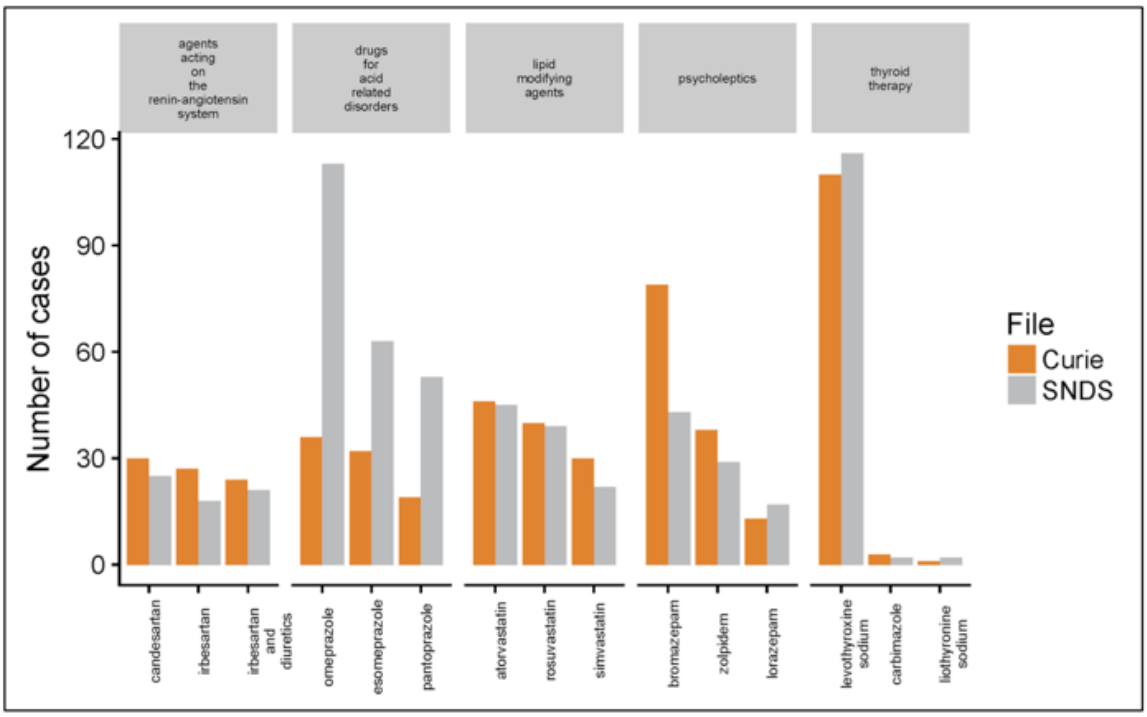

***Figure 2** : Comparaison entre la délivrance de médicaments en pharmacie (SNDS) et la déclaration de comédications dans le dossier médical (données Curie) pour 981 patientes appariées. Les molécules sont regroupées par classe ATC.*

- Enfin, nous avons étudié la concordance entre le **sous type histologique** de cancer du sein (*luminal, HER2+, triple négatif (TNBC))* et leur approximation par les traitements reçus par les patientes.

  Pour classer les sous types de cancer du sein à partir des données du SNDS, nous avons défini 4 postulats :

  - 1 patiente ayant reçu du trastuzumab est porteuse d'un cancer du sein ***HER2*-positif (*HER2+*) ;**
  - 1 patiente ayant reçu de l'hormonothérapie mais pas de trastuzumab est porteuse d'un cancer du sein **luminal ;**
  - 1 patiente ayant reçu de la chimiothérapie sans autre traitement systémique associé est porteuse d'un cancer du sein **triple négatif (TNBC)** ;
  - Les patientes traitées par chirurgie +/- radiothérapie sans traitement systémique ont été classées comme sous type **non défini**.

  Nous avons ensuite comparé les résultats obtenus aux sous types histologiques "vrais" définis dans la base de l'institut Curie (*n=1412*). La concordance entre les sous types "approximés" par les traitements dans le SNDS et les sous types "vrais" est résumée dans le tableau 2.

		Sous-types histologiques "vrais"			
		TNBC	*HER2+*	*Luminal*	*Total*
Sous-types approximés par les traitements reçus à partir des données du SNDS	*TNBC*	**40**	2	1	43
	HER2+	0	**294**	4	298
	Luminal	2	27	**880**	907
	Non défini	50	9	155	164
	Total	92	332	1040	*1412*

***Tableau 2** : Concordance obtenue entre les sous-types histologiques "vrais" annotés dans la base institutionnelle de l'Institut Curie et les sous-types approximés par les traitements reçus à partir des données du SNDS pour une cohorte de 1412 patientes de l'Institut Curie appariée aux SNDS. L'algorithme permet d'identifier le sous-type dans 86% des cas rencontrés.*

Cette approche valide le fait d'utiliser les données du SNDS pour approximer le sous type de cancer du sein avec d'excellentes mesures de **spécificité** : 93% pour les triples négatifs, 99% pour les HER2+ et 97% pour les luminaux. En d'autres termes, les sous-types approximés sont corrects lorsqu'ils sont identifiés par l'algorithme. Les cas où le sous-type approximé est **non défini** seront retirés de l'étude stratifiée par sous-groupe de cancer du sein**.**

> ***=> En conclusion, contrairement à la majorité des bases institutionnelles, le SNDS répertorie les soins dispensés quel que soit l'établissement ou les acteurs du parcours de soin (médecin traitant, hôpitaux, centres d'imagerie, pharmaciens…) assurant la prise en charge. Les données, actualisées en continu, sont disponibles à des fins de recherche jusqu'au 31 Décembre 2017. Il constitue donc la ressource idéale agrégeant l'ensemble des données d'un patient pour évaluer son devenir oncologique à long terme.***

=> Concernant les annotations clinico biologiques essentielles à l'étude des comédications et leur impact sur le pronostic du cancer du sein :

- *Les **récidives de cancer du sein** peuvent être identifiées par l'intermédiaire de l'application d'un algorithme simple, et ce malgré l'absence d'indicateur structuré de survenue d'une rechute.*
- *Le SNDS apparaît comme une source d'information plus fiable que le dossier médical patient pour identifier la prise effective de **comédications non anti cancéreuses** au cours d'une pathologie cancéreuse.*
- *Enfin, les traitements spécifiques au cancer permettent d'identifier de manière fiable **le sous type histologique de cancer du sein** dans la grande majorité des cas.*

Le SNDS offre donc des perspectives majeures dans le champ de l'étude des comédications-comorbidités et leur influence sur le pronostic du cancer du sein.

2. OBJECTIFS ET FINALITES

2.1. Objectif principal

L'étude COMBICANCER-SEIN vise à sélectionner des médicaments « candidats » **non anti-cancéreux** identifiés à partir de l'analyse du Système Nationale de Données de Santé (SNDS) pour leur potentiel **thérapeutique** anti-tumoral ou leur interaction **délétère** en combinaison avec les traitements de cancer du sein. L'objectif principal est donc d'analyser les interactions entre comédications / comorbidités et pronostic sur la cohorte quasi-exhaustive de patientes traitées pour un cancer du sein en France.

Dans un deuxième temps, et en dehors du paramètre de l'étude COMBICANCER-SEIN, les interactions suggérées pourront être testées expérimentalement par criblage in vitro sur modèles murins xénogreffés avec des tumeurs primaires issues de patients.

Les interactions délétères validées expérimentalement feront l'objet **d'alertes de pharmacovigilance**. Les molécules ou combinaisons de molécules identifiées par l'analyse statistique et la validation expérimentale comme apportant un **bénéfice** sur le pronostic pourraient être testées dans le cadre du **repositionnement de médicaments.** Des essais cliniques de repositionnement de ces molécules non anti-cancéreuses pourraient être envisagés en monothérapie ou en combinaison avec des traitements anti-cancéreux correspondants aux standards de traitement.

2.2. Justification de l'intérêt public

- Avec 400 000 nouveaux cas par an en France, le cancer est un réel enjeu de **santé publique.** Il est la première cause de mortalité (163 602 décès par an / 567 078). Le **coût des traitements** s'alourdit de manière très rapide (14.7 milliards remboursement CNAM en 2016), et il représente la pathologie la plus coûteuse par patient (coût patient avec cancer : 11 500 euros /an *versus* patient avec diabète : 2 500 euros / an) [17].

- L'incidence du cancer du sein augmente avec l'âge, de même que l'incidence de nombreuses autres maladies chroniques, telles que le diabète, l'hypertension et les maladies cardiovasculaires. Avec le vieillissement de la population, le nombre de patients atteints de cancer, souffrant de **comorbidités** et ayant recours à des **comédications** augmente de manière inexorable. A contrario, ces populations avec comorbidités sont **largement sous-représentées dans les essais cliniques**, sélectionnant volontiers des patients non polypathologiques. L'utilisation de données de vie réelle pour effectuer des recherches sur cette population paraît une alternative pertinente.

- Le **repositionnement de médicaments** déjà développés prescrits pour des indications thérapeutiques non anti-cancéreuses constitue une voie de recherche potentiellement efficace dans le traitement de cancer en comparaison au développement de nouvelles molécules. En effet, le taux d'attrition est élevé dans le développement de nouveaux médicaments anticancéreux (moins de 5% des médicaments entrant dans les essais de phase obtiennent finalement une autorisation de mise sur le marché).

- Certains médicaments non anticancéreux peuvent **potentialiser les traitements oncologiques** (radiothérapies, chimiothérapies, thérapies ciblées, hormonothérapie, immunothérapies) du cancer du sein. Ces molécules seraient des compléments idéaux aux traitements classiques car elles sont **peu onéreuses** en comparaison aux traitements oncologiques innovants. Par ailleurs, elles ont déjà été développées, commercialisées, et sont peu pourvoyeuses d'effets secondaires.

- A l'inverse, l'utilisation de certaines molécules au cours des traitements oncologiques pourraient en diminuer l'efficacité et entraver le pronostic des patientes à long terme. L'identification de molécules ou combinaisons de molécules aux effets **délétères** permettrait d'émettre des alertes de pharmacovigilance et de protéger les patients de prescriptions médicamenteuses dangereuses en cas de cancer.

- Le concept de recherche d'interférence entre les comédications et le cancer du sein est facilement généralisable à **d'autres types de cancer.** Au vu (i) du nombre conséquent de nouveaux cas de cancer dans le monde et (ii) du taux élevé de décès attribuable à cette maladie, la quantification de l'impact des comédications/ comorbidités sur le pronostic est un enjeu de santé publique.

3. METHODOLOGIE

Le projet **COMBICANCER-SEIN** comporte 3 étapes, qui s'échelonnent sur une durée de 2 ans :

1. **Construction de la cohorte SNDS-SEIN,** ensemble des femmes traitées pour cancer du sein incident entre 2007 et 2017 (*n= 450 000*).

2. Annotation clinico-biologique de la **cohorte SNDS-SEIN** à l'aide d'algorithmes validés sur des cohortes indépendantes.

3. Analyses statistiques de l'interaction entre les comorbidités/comédications et le pronostic sur la **cohorte SNDS-SEIN (*n= 450 000*).**

3.1. Données requises

3.1.1. DESCRIPTION DE LA COHORTE PROJET – COHORTE SNDS-SEIN

- **Description de la base de données source (SNDS) :**

Base de données source : Système National des Données de Santé (SNDS) : La CNAM consolide et pseudonymise (avec les fonctions FOIN1 et FOIN2) les données suivantes :

- Données de l'Assurance Maladie (consommations de soins en ville et en établissement remontées dans le SNIIRAM depuis 2006).
- Données hospitalières du PMSI issues de l'ATIH et depuis 2006.
- Données sur les causes médicales de décès du CépiDC-Inserm, de la période 2013-2015 à la date d'aujourd'hui.

Légalité : Article L. 1461-1 du code de la santé publique [18].

Périmètre : Ensemble des personnes ayant eu recours au système de soin français ou étant décédées sur le territoire.

Format des données : données structurées.

- **Description de la cohorte SNDS-SEIN : données de patientes atteintes de cancer du sein de l'ensemble du SNDS (*n=450 000*) :**

Critères d'inclusion : ensemble des patientes diagnostiquées et traitées pour un cancer du sein incident (in situ ou invasif) entre 2007 et 2017. Ne sont conservées que les patientes de plus de 18 ans au moment du diagnostic et étant sous le régime général (RG) ou Section Locale Mutualiste (SLM). On estime la cohorte à environ 45,000 patientes par an, soit **450,000 patientes** au total.

Critères d'exclusion : Sont exclues les patientes métastatiques d'emblée ainsi que celles mineures au moment du diagnostic. Le cancer du sein est dit métastatique d'emblée si des métastases sont identifiées dans les 6 premiers mois après le diagnostic. Les cancers du sein non incidents ainsi que les patientes avec un cancer concomitant ne sont pas concernées par l'étude.

Profondeur d'histoire des données extraites : du 01/01/2006 au 31/12/2017. L'analyse débutant 1 an avant la date de diagnostic pour l'ensemble des patientes ayant un cancer du sein entre 2007 et 2017, les données doivent être extraites à partir de 2006. De plus, contrairement à la majorité des cancers, le cancer du sein est une maladie dont l'histoire naturelle continue à évoluer après les 5 premières années [19]. Il existe notamment une possibilité de rechutes tardives. Un suivi aussi long et exhaustif que possible est donc nécessaire.

Durée de conservation : Les données seront conservées 5 ans : 2 ans pour la réalisation de l'étude, 3 ans pour la publication des résultats, puis elles seront supprimées définitivement.

Appariement : pas d'appariement prévu.

Données sensibles : Aucune donnée raciale, ethnique, religieuse, génétique ou biométrique ne sera étudiée. Les dates exactes de naissance ne sont pas requises pour l'analyse : l'âge est suffisant. Les dates exactes des soins et des décès sont indispensables pour l'étude des récidives et l'extraction des comédications. La commune de décès du bénéficiaire n'est pas étudiée. Le département de résidence est suffisant pour l'étude.

Méthodologie de ciblage de la cohorte : L'ensemble des personnes hospitalisées pour un diagnostic de cancer entre 2007 et 2017 ou bénéficiant d'une ALD de cancer du sein entre ces dates est identifiée par la CNAM. La cohorte est ensuite réduite aux femmes majeures bénéficiaires du RG ou SLM ayant reçu un traitement pour cancer du sein à l'aide des codes diagnostiques CIM-10 de cancer du sein in situ ou invasif (hors tumeurs à évolution imprévisible ou inconnue). Il est possible d'exclure par la suite les personnes ayant eu une ALD pour cancer du sein antérieure à 2007, et les personnes traitées pour cancer du sein en 2006, afin d'approcher au mieux les cas incidents. Le cancer du sein est dit métastatique d'emblée si des codes CIM10 de métastase d'organe (C78-C79) sont retrouvées dans les 6 premiers mois après le diagnostic. Les femmes avec un autre cancer primitif concomitant (du diagnostic de cancer du sein à 2017) sont exclues afin d'attribuer au mieux les traitements reçus au cancer du sein incident.

3.1.2. DESCRIPTION DES VARIABLES CLINICO-BIOLOGIQUES EXTRAITES OU INFEREES DU SNDS

Les variables clinico-biologiques n'existant pas en tant que telles dans les données du SNDS, elles doivent être extraites ou inférées par le biais d'algorithmes développés sur des cohortes indépendantes appariées au SNDS. Les indicateurs **simples** i.e. extraits directement du SNDS sont à distinguer des indicateurs **complexes** construits par le biais d'algorithmes.

- **<u>Indicateurs simples /directs :</u>**

Les **caractéristiques démographiques** des patientes (âge, département de résidence) sont directement extraites.

Les **dates de décès** sont structurées et exhaustives pour le Régime Général (RG) à partir de 2007. Les **causes de décès** sont disponibles pour de la période 2013-2015 uniquement.

Les **traitements reçus** (chirurgie, hormonothérapie, radiothérapie, chimiothérapie, thérapie ciblée) sont identifiés en utilisant les codes listés dans le tableau suivant :

	PMSI (MCO/SSR/RIM-P/HAD)			VILLES	
	Actes		Liste en sus	Actes	Pharmacies
	CCAM	CIM-10	UCD	CCAM	CIP
Chirurgie	Mastectomie partielle : QEFA004, QEFA001, QEFA017, QEFA008, QEFA016, QEFA018 Mastectomie totale : QEFA007, QEFA019, QEFA020, QEFA005, QEFA010, QEFA003, QEFA012, QEFA013, QEFA015 Curage ganglionnaire : QEFA001, QEFA008, QEFA020, QEFA005, QEFA010, QEFA003, FCFA029, FCFA018, FCQX004, FCQX006, FCQX010, FCQX012, FCQX011 Exerese du mamelon: QEFA009				
Radiothérapie	Actes d'irradiation externe ou interne (voir annexe)			Actes d'irradiation du chapitre 19 supplément de facturation (voir annexe)	
Chimiothérapie		Diagnostic principal :Z511 Diagnostic relié : non rempli ou D05 ou C50	Liste de molécules		Liste de molécules
Thérapie ciblée			Liste de molécules		
Hormonothérapie					Liste de molécules

L'identification des **comorbidités** s'appuie sur les diagnostics (codes CIM-10) principaux, associés et reliés mentionnés dans les séjours d'hospitalisation des patientes ou en motif d'affections de longue durée (ALDs). Certaines pathologies seront aussi identifiées par la délivrance d'actes ou de médicaments spécifiques à leur traitement. Les scores de Charlson et les données de la cartographie développée par la CNAMTS [16] (13 grandes catégories non exclusives de pathologies correspondant à 56 groupes non exclusifs) disponible à partir de 2012 seront aussi extraits.

Les **comédications** seront déduites des médicaments délivrés en officine de ville, en rétrocession et sur la liste en sus, délivrés des 6 mois précédant la date de diagnostic et jusqu'au 31 Décembre 2017. Les informations disponibles pour chaque délivrance seront extraites : code CIP ou UCD, classe ATC, quantité, date. Sont considérées comme comédications les traitements chroniques prescrits pour des pathologies chroniques. Les traitements intercurrents ainsi que les médicaments régulièrement prescrits en parallèle des chimiothérapies (anti-vomitifs, facteurs de stimulation des granulocytes, stéroïdes) n'entrent pas dans la définition. Concrètement, sont conservés les molécules délivrées en pharmacie au moins 2 fois en 6 mois et pouvant être utilisées au cours de traitements chroniques hors traitement du cancer du sein incident.

Les **toxicités** seront extraites à partir d'acte, de diagnostic ou de médications selon la méthodologie suivante :

Aplasie/ Neutropénies fébriles	Code CIM-10 D60-61-62-69-R502
Effets indésirables anti-tumoraux	Code CIM-10 Y431-433
Antibiotique	Code ATC J01 (en ville)
G-CSF	Code ATC Facteurs de croissance L03AA (en ville)
EPO	Code ATC B03X (en ville et liste en sus)
Transfusion	Code CIM-10 Z513 / Code CCAM FELF011
Diarrhées profuses	Code CIM-10 K59
Déshydratation	Code CIM-10 E86
Anomalie du métabolisme du calcium	Code CIM-10 E835
Altération état général	Code CIM-10 R53+0
Douleurs	Code CIM-10 R07-R10-R52

Ne seront conservées que les toxicités apparaissant à moins de 3 semaines d'un traitement par chimiothérapie.

- **<u>Indicateurs complexes/indirects :</u>**

Le **sous-type de cancer** est inféré des traitements reçus selon l'algorithme suivant, en cours de validation sur une cohorte de 9500 patientes de l'Institut Curie appariée au SNDS :

- Les femmes recevant du trastuzumab sont classées *HER2+.*
- Les femmes recevant une hormonothérapie sont classées récepteurs hormonaux positifs (*RH+*). En particulier, celles recevant de l'hormonothérapie mais pas de trastuzumab sont classées luminal.
- Les femmes ne recevant ni d'hormonothérapie ni de trastuzumab mais une chimiothérapie sont classées triple négatif (*TNBC*).
- Les femmes restantes sont classées comme non définies.

Les **rechutes hors décès** sont imputées par la reprise des traitements de cancer du sein (chirurgie, chimiothérapie, radiothérapie) après une période d'interruption des traitements d'au moins 5 mois (hors hormonothérapie) conformément à l'étude préliminaire menée par notre équipe dont les résultats sont décrits plus haut. Un algorithme d'identification des rechutes (hors décès) et dates de rechutes utilisant des réseaux de neurones récurrents (RNN pour Recurent Neural Network) est aussi en cours d'implémentation et pourrait être utilisé s'il conduit à des meilleures performances que la méthode ad hoc décrite ci-dessus.

3.1.3. DESCRIPTION DES VARIABLES NECESSAIRES POUR L'ANALYSE

Les variables explicatives à analyser seront :

- ***Démographiques*** : Âge, département de résidence.
- ***Anatomopathologiques*** : Sous-type inféré de cancer du sein (*HER2+, TNBC, luminal*).
- ***Thérapeutiques*** : Chirurgie, chimiothérapie, hormonothérapie, thérapie ciblée, radiothérapie (date, type et centre de traitement).
- ***Comédications*** : Les comédications seront analysées sur la base de la classification internationale ATC (Anatomical Therapeutic Chemical) développée par l'OMS. Les dates de délivrance en pharmacie seront aussi extraites.
- ***Comorbidités*** : Les comorbidités présentes au moment du diagnostic de cancer du sein ainsi que celles diagnostiquées secondairement. Elles seront organisées selon la classification internationale ICD-10 (International Classification of Diseases). L'indice de comorbidité de Charlson et les données de la cartographie développée par la CNAM à partir de 2012 seront aussi intégrés aux analyses. Les comorbidités suivantes seront principalement analysées: infarctus du myocarde, insuffisance cardiaque, atteinte vasculaire périphérique, paralysie d'origine vasculaire (hémiplégie, hémiparésie, paraplégie), insuffisance pulmonaire, démence, diabète, insuffisance rénale chronique, pathologie hépatique, ulcère gastrique, arthrite, maladie des tissus conjonctifs, antécédents de cancer, hypertension, SIDA, obésité, dépression, anxiété, pathologie thyroïdienne, hypercholestérolémie, dyslipidémie.

Les indicateurs pronostiques à analyser seront :

- ***Les effets secondaires des traitements*** (toxicité). L'analyse portera essentiellement sur les éléments suivants : nausées vomissements, diarrhées entérocolites, réactions allergiques, toxicités hématologiques. Les hospitalisations, les pathologies intercurrentes associées à ces toxicités et les traitements mis en place seront documentés et étudiés. Ne seront considérées que les toxicités apparaissant sous la forme de code diagnostic d'une hospitalisation à moins de 3 semaines d'un traitement par chimiothérapie. En particulier, les effets secondaires survenant hors hospitalisation ne sont pas identifiables.
- ***L'évolution cancérologique***: Les indicateurs habituels suivants seront analysés : récidive locorégionale, récidive métastatique, décès.

Les variables explicatives et indicateurs pronostiques seront étudiés lors de **l'analyse statistique** (étape 3).

3.2. Méthodes, traitements et analyses des données

Le projet **COMBICANCER-SEIN** comporte 3 étapes, qui s'échelonnent sur une durée de 2 ans.

Étape n°1 : Construction de la cohorte SNDS-SEIN

A partir des données du SNDS, nous générerons une base de données SNDS des patientes majeures, de sexe féminin et du RG ayant eu un premier diagnostic de cancer du sein incident entre 2007 et 2017, non métastatiques d'emblée et sans cancer concomitant : il s'agit de la **cohorte SNDS-SEIN.** On estime la taille de la cohorte à environ 450 000 patientes.

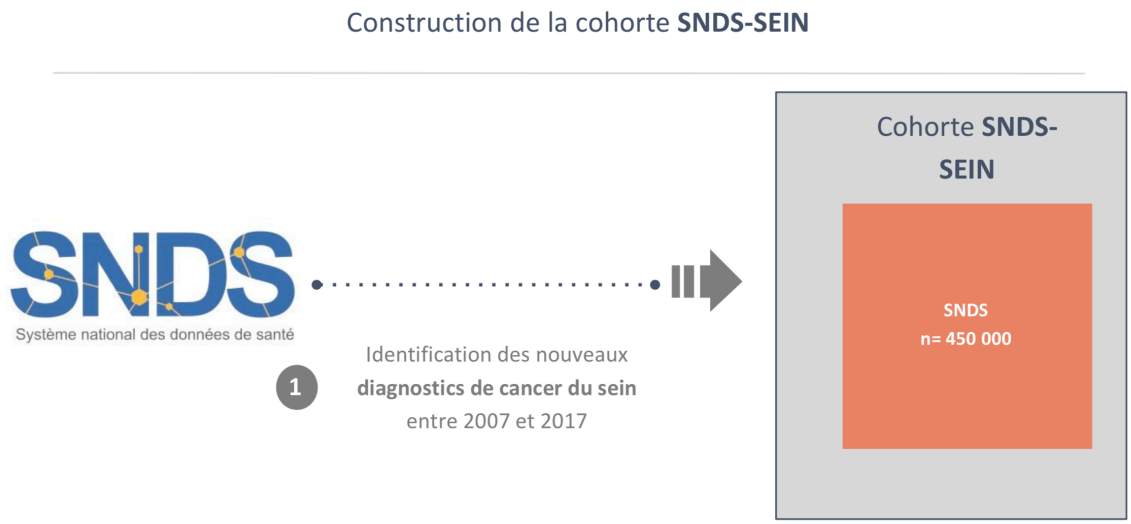

***Figure 3** : Construction de la cohorte SNDS-SEIN - étape 1*

Étape n°2 : Annotation clinico-biologique de la cohorte SNDS-SEIN

Au cours de cette étape, les données de la cohorte SNDS-SEIN (*n= 450 000*) sont **annotées** au moyen des indicateurs clinico-biologiques directs et indirects validés en amont de l'étude sur des cohortes indépendantes appariées au SNDS (publication en cours).

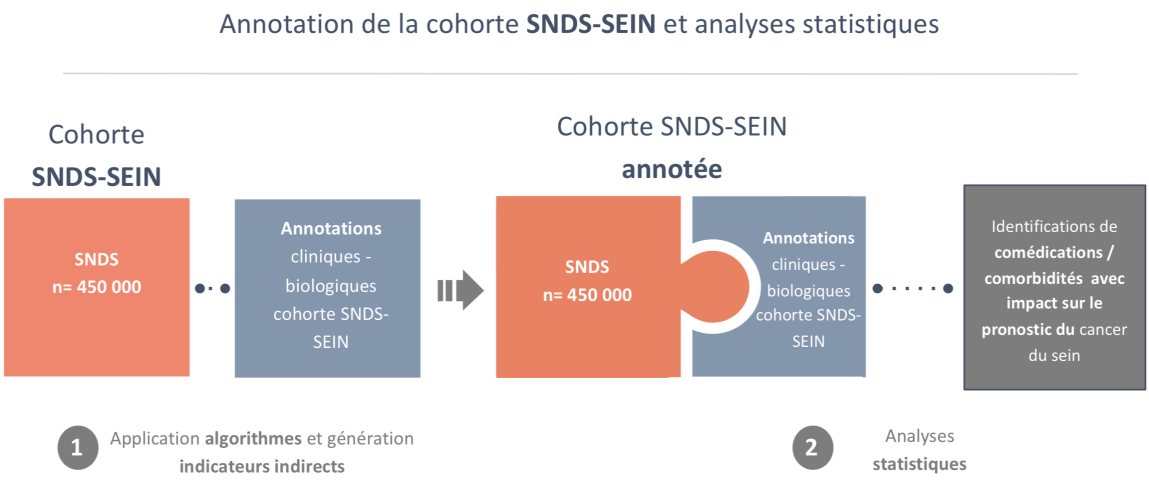

Figure 4 *: Annotation de la cohorte SNDS-SEIN (étape 2) puis analyses statistiques sur la cohorte SNDS-SEIN annotée (étape 3)*

Étape n°3 : Analyses statistiques sur la cohorte SNDS-SEIN

Le plan d'analyse suivant est considéré à partir de **la cohorte SNDS-SEIN annotée** :

1. Analyse descriptive des données

La fréquence, la répartition et le nombre de comédications par patiente sont évalués par tranche d'âge et en fonction du département de résidence, tout comme les taux de récidives et le pourcentage de décès. La concordance entre les comorbidités et la prise de comédications sera étudiée. La répartition des sous-types inférés sera comparée aux chiffres présentés dans la littérature dédiée [20].

2. Analyse de l'évolution carcinologique en fonction de l'existence de comédications et/ou de comorbidités

Les associations entre le pronostic et les caractéristiques patientes/tumeurs/traitements seront analysées par à un **modèle de Cox**. Les comédications seront testées molécules par molécules puis par grande classe anatomique ATC. Pour chaque classe, l'association à la survie sans récidive (SSR) sera testée. En cas de significativité de la p-value pour l'association inférieure à 0.1, les différentes classes thérapeutiques seront testées pour leur association à la SSR. Les analyses univariées sur la SSR seront suivies d'une analyse multivariée. Les analyses seront stratifiées selon le sous type de cancer du sein (luminal / HER2+ /triple négatif/ non défini). L'analyse statistique (modèle de Cox) souffre d'un biais de sélection induit par le fait que les patientes n'ont pas toutes la même probabilité de se faire délivrer une comédication. Nous utiliserons des méthodes de pondération par la probabilité inverse de prendre une comédication (Inverse Probability Weighting, IPW) [21] ou par standardisation pour atténuer ce biais. De même, des estimateurs doublement robustes (TMLE) [22] seront employés afin de pallier une éventuelle mauvaise spécification du modèle de pondération (IPW) ou du modèle d'analyse. L'utilisation de graphes dirigées acycliques (voir la figure 5) sera aussi considérée pour la prise en compte des facteurs confondants.

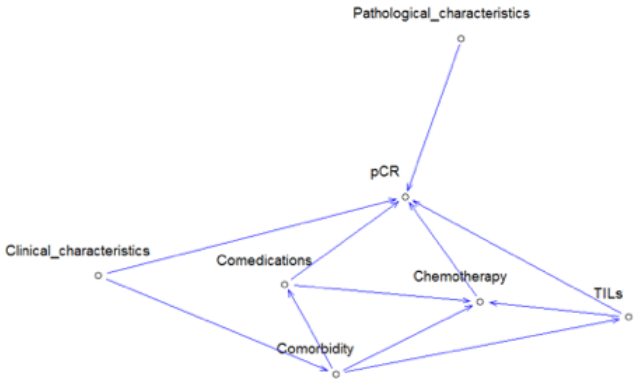

***Figure 5** : Exemple de DAG (Directed Acyclic Graph) caractérisant les relations entre les données clinico pathologiques et de traitement, les comédications, les comorbidités, l'infiltration immunitaire et la réponse histologique au traitement.*

3. Analyse des toxicités en fonction de l'existence de comédications et/ou de comorbidités

L'analyse menée pour l'évolution carcinologique est répétée pour l'analyse de l'interaction entre les comorbidités/comédications et l'apparition de **toxicités** à moins de 3 semaines d'un traitement par chimiothérapie. L'analyse sera stratifiée par type de toxicités et par sous-type de cancer. Les comédications seront testées molécules par molécules puis par grande classe ATC.

Enfin, l'usage de méthodes **d'apprentissage non supervisé** (clustering, réductions de dimension) permettra d'extraire des motifs caractéristiques des profils et parcours patients, créant ainsi des groupes de patients distincts (ou « *clusters*»). Les différences dans le nombre et le type de comorbidités, la prise de comédications, le pronostic du cancer du sein et l'apparition de toxicités pour chaque *cluster* sera étudié. Cela permettra notamment d'identifier des groupes de comédications et d'approfondir les études multivariées présentées ci-dessus en prenant en compte les interactions probables entre les molécules.

3.3. Circulation des données et appariement

L'extraction de la **cohorte SNDS-SEIN** (étape 1), à savoir les données du SNDS relatives aux femmes de plus de 18 ans ayant eu un cancer du sein incident non métastatique d'emblée entre 2007 et 2017 et sans cancer concomitant, est effectué par la CNAM pour répondre au besoin de l'étude.

L'extraction du SNDS est ensuite transmise au Health Data Hub via un réseau dédié, permettant un transfert privé (hors Internet) et sécurisé des données. Les données sont stockées et traitées dans un espace projet sécurisé du Health Data Hub, accessibles aux personnes habilitées seulement. Elles seront conservées pendant 5 ans : 2 ans pour la réalisation de l'étude, 3 ans pour la publication des résultats, puis elles seront supprimées définitivement.

3. SECURITE ET CONFIDENTIALITE DES DONNEES

4.1. Information des patients et protection de leur droit

4.1.1. INFORMATION INDIVIDUELLE DES PATIENTS

S'agissant d'une étude mobilisant uniquement les données du SNDS, aucune information individuelle ne sera possible. Les personnes concernées sont informées de l'existence du SNDS et de la réutilisation possible de leurs données dans les conditions prévues à l'article 111 du décret n° 2019-536 du 29 mai 2019 pris pour l'application de la loi Informatique & Libertés. Le Health Data Hub assurera une information collective sur le projet et ses résultats.

4.1.2. RESPECT DU DROIT DES PERSONNES

Les données de la cohorte SNDS-SEIN seront transmises consolidées et pseudonymisées par la CNAM. Aucune ré-identification des patients n'est prévu par le protocole.

S'agissant de données du SNDS, les droits d'accès, de rectification et d'opposition s'exercent conformément à l'article R.1461-9 du code de la santé publique :

- "Les droits d'accès, de rectification et d'opposition s'exercent, dans les conditions définies aux articles 92 à 95 du décret n° 2005-1309 du 20 octobre 2005 modifié, auprès du directeur de l'organisme gestionnaire du régime d'assurance maladie obligatoire auquel la personne est rattachée."
- "Le droit d'opposition prévu aux premier et troisième alinéas de l'article 56 de la loi n° 78-17 du 6 janvier 1978 porte sur l'utilisation des données dans les traitements mentionnés au 1° du I de l'article L. 1461-3 du code de la santé publique. Il ne s'applique pas aux traitements prévus au 2° du I du même article."

4.2. Support des données et sécurité

La plateforme de données de santé est en cours de constitution et homologation. La plateforme du Health Data Hub doit respecter le référentiel de sécurité applicable au Système National des Données de Santé (SNDS) décrit dans l'arrêté du 22 mars 2017, sur les aspects à la fois organisationnels et techniques.

La démarche de sécurité pour réaliser l'homologation de la plateforme du Health Data Hub en lien avec le HFDS & l'ANSSI est en cours de réalisation. Elle couvre notamment les axes suivants :

- **Une authentification multi-facteur forte** reposant sur une infrastructure à clés publiques maîtrisée par le Ministère des Solidarités et de la Santé ;
- **Un accès à distance** à la plateforme se reposant sur l'utilisation d'une infrastructure de bureaux virtuels, garantissant la sécurité des sessions de travail ;

- **Une traçabilité** complète de l'ensemble des actions réalisées sur le Hub retranscrite dans un interface ergonomique pour les besoins d'audit ;
- **L'import et l'export sécurisé et contrôlé** de résultats, de codes et de visuels en adéquation avec les habilitations reçues.

Les données requises pour l'étude seront hébergées et traitées sur la plateforme du Health Data Hub.

En accord avec la CNIL, tous ces points seront détaillés et complétés une fois l'analyse de risques (contexte, événements redoutés et scénarios de menace, étude des risques, étude des mesures de sécurité) et l'homologation de la plateforme réalisées à la fin du premier semestre 2019 soit avant le dépôt du dossier auprès de la CNIL.

5. CALENDRIER PREVISIONNEL

5.1. Calendrier prévisionnel

		Durée	Durée cumulée
Obtention des autorisations d'accès et de traitement		4 mois	4 mois
Etape n°1	Création de la cohorte SNDS-SEIN	3 mois	7 mois
Etape n°2	Annotation clinico-biologique de la cohorte SNDS-SEIN	6 mois	13 mois
Etape n°3	Analyses statistiques	9 mois	22 mois
Documentation de la base de données utilisée pour le projet		2 mois	24 mois
Publication et valorisation des résultats		36 mois	60 mois

5.2. Valorisation des résultats

Les résultats de ces travaux feront l'objet de publications dans des revues scientifiques internationales à comité de lecture et de communications dans les congrès appropriés.

6. BIBLIOGRAPHIE

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394–424 (2018).
2. Patnaik, J. L., Byers, T., Diguiseppi, C., Denberg, T. D. & Dabelea, D. The influence of comorbidities on

overall survival among older women diagnosed with breast cancer. *J. Natl. Cancer Inst.* **103**, 1101–1111 (2011).

3. Powe, D. G. *et al.* Beta-blocker drug therapy reduces secondary cancer formation in breast cancer and improves cancer specific survival. *Oncotarget* **1**, 628–638 (2010).
4. Zhao, Y. *et al.* Association between NSAIDs use and breast cancer risk: a systematic review and meta-analysis. *Breast Cancer Res. Treat.* **117**, 141–150 (2009).
5. Ahern, T. P. *et al.* Statin prescriptions and breast cancer recurrence risk: a Danish nationwide prospective cohort study. *J. Natl. Cancer Inst.* **103**, 1461–1468 (2011).
6. Haukka, J., Niskanen, L. & Auvinen, A. Risk of Cause-Specific Death in Individuals with Cancer-Modifying Role Diabetes, Statins and Metformin. *Int. J. Cancer* **141**, 2437–2449 (2017).
7. AACR Cancer Prevention Summit. https://www.aacr.org/Research/Research/PAGES/AACR-CANCER-PREVENTION-SUMMIT.ASPX.
8. Cancer today. http://gco.iarc.fr/today/home.
9. Bergström, A., Pisani, P., Tenet, V., Wolk, A. & Adami, H. O. Overweight as an avoidable cause of cancer in Europe. *Int. J. Cancer* **91**, 421–430 (2001).
10. Cui, Y. *et al.* Body mass and stage of breast cancer at diagnosis. *Int. J. Cancer* **98**, 279–283 (2002).
11. Louwman, W. J. *et al.* Less extensive treatment and inferior prognosis for breast cancer patient with comorbidity: a population-based study. *Eur. J. Cancer* **41**, 779–785 (2005).
12. Fabien Reyal, A.-S. H.-P. Comedications influence immune infiltration and pathological response to neoadjuvant chemotherapy in breast cancer.
13. Hamy, A.-S. *et al.* Celecoxib With Neoadjuvant Chemotherapy for Breast Cancer Might Worsen Outcomes Differentially by COX-2 Expression and ER Status: Exploratory Analysis of the REMAGUS02 Trial. *JCO* **37**, 624–635 (2019).
14. Hernán, M. A. & Robins, J. M. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* **60**, 578–586 (2006).
15. Bousquet, P. J. *et al.* Cancer care and public health policy evaluations in France: Usefulness of the national cancer cohort. *PLOS ONE* **13**, e0206448 (2018).

16. Caisse Nationale d'Assurance Maladie (CNAM, Méthode générale de la cartographie des pathologies, version G5 (années 2012 à 2016) Mise à jour : 8 mars 2018.
17. *LOI n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé - Article 193*. *2016-41* (2016).
18. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* **40**, 373–383 (1987).
19. Bousquet, P.-J. *et al.* [Using cancer case identification algorithms in medico-administrative databases: Literature review and first results from the REDSIAM Tumors group based on breast, colon, and lung cancer]. *Rev Epidemiol Sante Publique* **65 Suppl 4**, S236–S242 (2017).
20. Molina, J., Rotnitzky, A., Sued, M. & Robins, J. M. Multiple robustness in factorized likelihood models. *Biometrika* **104**, 561–581 (2017).